



Innhald

Føreord frå dei nye redaktørane	3
Innledning til <i>Norsk lingvistisk tidsskrifts</i> temahefte om andrespråkstilegnelse	5
MAUD WESTENDORP OG BJÖRN LUNDQUIST Tverrspråklig innflytelse fra L1 i tilegnelsen av argumentplassering i L2 norsk og svensk	13
ANDERS AGEBJÖRN Development of the form and meaning of definiteness in Russian-speaking learners of Swedish	41
EEVA-LIISA NYQVIST Komplexa svenska nominalfraser hos finskspråkiga språkbadselever	69
ISABEL NADINE JENSEN OG MARIT WESTERGAARD Flaskehalshypotesen: Syntaks og morfologi blant norske andrespråksinnlærere av engelsk	97
KRISTIN KILLIE 'I are in New York': om tilegnelsen av samsvarsbøyning i engelsk blant innlærere med norskspråklig bakgrunn	123
ANNA-MARIE KJØDE OLSEN Adjektivlæring hos vaksne innlærarar av norsk – ein empirisk studie	149
ANN-KRISTIN HELLAND GUJORD Syntaktisk kompleksitet på tvers av rammeverksnivå	177

Bokmelding

SVERRE STAUSLAND JOHNSEN

Ummæle av *Språket som vart norsk* 207



Føreord frå dei nye redaktørane

Kjære lesar

Ved årsskiftet overtok vi redaktøransvaret for Norsk Lingvistisk Tidsskrift, etter oppnemning av Det Norske Videnskabs-Akademi. Vi er takksame for tilliten vi har fått, og vi gler oss til å vere med på å løfte fram eit mangfald av faglege emne innanfor norsk lingvistikk. Vi ynskjer velkommen artiklar som omhandlar tematikk innanfor alle dei språkvitskaplege disiplinane. I dette arbeidet vil vi ha nytte og glede av samarbeid med det nye redaksjonsrådet.

I føreordet til nr. 20-2 takkar Ernst Håkon Jahr av det førre redaksjonsrådet, som då hadde fullført femårsperioden sin, og gjort ein solid innsats for NLT. Det nye redaksjonsrådet, som vi presenterer her, er samansett av fagpersonar frå eit breitt spekter av landets universitet og høgskular som har fokus på lingvistiske emne. Nokre utanlandske representantar har vi òg med. I det nye redaksjonsrådet sit Gisle Andersen (NHH), Kaja Borthen (NTNU), Kristin Melum Eide (NTNU), Þórhallur Eypórsson (Háskóli Íslands), Hilde Hasselgård (UiO), Ernst Håkon Jahr (UiA), Marit Julien (Lunds universitet), Tor Anders Åfarli (NTNU), Åse Mette Johansen (UiT), Tor Arne Haugen (HVO), David Håkansson (Uppsala universitet), Tore Nasset (UiT), Randi Neteland (UiB), Åshild Søfteland (HiØ) og Signe Laake (OsloMet). Vi er glade for at dei vil stille kompetansen sin til rådvelde for NLT, og ser fram til eit godt og fruktbart samarbeid dei komande fem åra.

I tidsskriftet ynskjer vi som vanleg å trykke vitskapleg solide tekstar om norsk lingvistikk. Den viktigaste tekststypen i NLT vil framleis vere vitskaplege artiklar. Fagfellevurdering er ei særskild viktig side ved vitskapleg publisering, og NLT vil halde på ordninga med at kvar artikkel blir lesen av to fagfellar i ein dobbeltblinda prosedyre. Vi er svært takksame for positive svar når vi kontaktar moglege fagfellar. I tillegg til artiklar vil vi også publisere bokmeldingar. Desse er viktige for å anerkjenne arbeidet som ligg i å skrive vitskaplege monografiar, antologiar og lærebøker. Vi ser også føre oss at vi stundom vil trykke disputasinlegg.

Redaktørane planlegg ei fordeling av frie allmennummer og temanummer med gjesteredaktørar, slik at det alltid skal vere råd å få plass til enkeltstående

4 Forord fra de nye redaktørene

artiklar i tidsskriftet. NLT vil bli redigert på nynorsk, i den opphavlege tradisjonen til NLT, som blei tatt opp igjen av Janne Bondi Johannessen frå 2016.

NLT ynskjer å arbeide for at tidsskriftet skal vere ein attraktiv publiseringskanal i formidlinga av fagtekstar om norsk språk, på norsk. Vi har også kollegaer utanfor Noregs grenser, og ein kan skrive på engelsk om emnet er norsk språk. I tillegg vil det nokre gonger vere fint å ha med bokmeldingar frå internasjonale anmeldarar på engelsk, slik at bøker som har med norsk språk å gjere, får bli godt kjent.

Ved å halde på norsk som hovudspråk i NLT vil vi bidra til å løfte norsk som fagspråk og vidareutvikle norsk fagterminologi. Berre gjennom å skrive fagstoff på norsk kan vi syte for at norsk som fagspråk og norsk fagterminologi i våre disiplinlar blir skjøtt og vidareutvikla til å dekke alle fagkategoriar og teoretiske retningar. Når ein kan skrive på førstespråket, opnar det også for ei anna form for fagskriving, med større nyanserikdom og auka presisjon. Norsk som publiseringsspråk gjer dessutan tekstane lettare tilgjengeleg for delar av lesargruppa.

Det å nytte sosiale medium er ein god måte å halde kontakt med lesarane på mellom kvar utgjeving av NLT. Vi har ambisjonar om å oppretthalde og vidareutvikle Facebooksida til NLT. Der vil vi legge ut informasjon om nye nummer, invitasjonar til å skrive artiklar eller bokmeldingar, og elles oppmode til ein levande samtale om faglege spørsmål.

Dette nummeret av NLT inneheld ei samling artiklar om ulike aspekt ved andrespråklæring, gjesteredigert av Ann-Kristin Helland Gujord og Guro Busterud. De kan lese meir om innhaldet med presentasjon av arbeidet og artiklane i føreordet som fylgjer etter dette. I tillegg inneheld nummeret ei bokmelding av *Språket som vart norsk* av Ivar Berg, skriven av Sverre Stausland Johnsen. Denne bokmeldinga blei bestilt av Janne Bondi Johannessen. Vi er glade for å kunne ha ho med her, og for å fullføre arbeidet med dei tekstane som Janne såg føre seg.

Vi ber om at nye manus blir sendt til Elin Gunleifsen (elin.gunleifsen@uia.no), og at de følgjer forfattarinstruksen på heimesida til NLT.

Elin Gunleifsen, Gro-Renée Rambø og Magnhild Selås



Innledning til *Norsk lingvistisk tidsskrifts* temahefte om andrespråkstilegnelse

Guro Busterud¹ og Ann-Kristin Helland Gujord²

Universitetet i Oslo¹ og Universitetet i Bergen²

Andrespråkstermen er en overordnet term som betegner tilegnelse av ikke-primære språk i ulike kontekster. Dette temaheftet handler om andrespråkstilegnelse i *Norden*. Felles for artiklene er at de presenterer studier av grammatiske aspekter ved andrespråkstilegnelse. Målspråk og språklæringskontekst er derimot gjenstand for variasjon. I noen artikler er de nordiske språkene norsk og/eller svensk målspråk, i andre artikler inngår norsk som innlærernes førstespråk. I to av studiene er norsk målspråket (Olsen og Gujord), to studier har svensk som målspråk (Agebjörn og Nyqvist), én studie ser på tilegnelse av både norsk og svensk (Westendorp & Lundquist), og i to studier er engelsk målspråk (Jensen & Westergaard og Killie). Temaheftet går slik utenfor rammen av fagfeltet *norsk* som andrespråk.

Temaheftet inkluderer ikke bare studier med forskjellige førstespråk (L1) og andrespråk (L2), men spenner også over bredden av andrespråklæringskontekster vi finner i Norden i dag. Noen av studiene undersøker tilegnelse av andrespråk i en kontekst der andrespråket er i allmenn bruk i samfunnet rundt språkinnlæreren. Men de nordiske språkene tilegnes også utenfor Norden i en fremmedspråkskontekst, som for eksempel informantene i Agebjörns studie, som lærer svensk i Russland, og deltakerne i to av eksperimentene til Westendorp & Lundquist, som lærer henholdsvis norsk og svensk i Nederland. En annen type L2-innlærere finner vi i Nyqvists studie av finsktalende barn som lærer svensk gjennom deltakelse i språkbadsprogram. Informantene i Killies studie og i studien til Jensen & Westergaard, norske elever som lærer engelsk på skolen, representerer en type andrespråklæring som ikke lett lar seg plassere i den tradisjonelle forståelsen av forskjellen på andrespråkstilegnelse i en andrespråks- og fremmedspråkskontekst. Norske barn lærer engelsk tidlig (fra 6 års

alder), og norsk ungdom har høy kompetanse i engelsk og bruker engelsk i domener der engelsk ikke er påkrevet for å kunne kommunisere (jf. Rindal 2020). Dette utfordrer tradisjonelle definisjoner av fremmedspråk og andrespråk. Ifølge Rindal (2020: 31) er engelsk verken et fremmedspråk eller et andrespråk i Norge, men befinner seg et sted midt imellom.

Selv om artiklene i dette temaheftet ikke er ensartet verken med hensyn til målspåket som undersøkes eller med hensyn til konteksten andrespråket læres i, representerer alle artiklene studier av andrespråkstilengelse, eller ikke-primær språktilegnelse, i en nordisk kontekst, der nordiske språk inngår i innlærerens språklige repertoar. Nordiske språk spiller en essensiell rolle i alle disse studiene. I studiene der norsk er innlærernes førstespråk, får vi innsikt i hvordan strukturelle aspekter ved norsk som L1 påvirker tilegnelsen av engelsk som andrespråk. I studiene der norsk og/eller svensk er målspåket, får vi utvidet vår kunnskap om ikke-primær tilegnelse av disse språkene. Norsk og svensk har mange likheter strukturelt sett, og i andrespråkssammenheng er derfor studier av tilegnelse av disse språkene som L2, gjensidig interessante. Artikkelen til Westendorp & Lunquist viser at selv om norsk og svensk har mange likheter, finnes det også mikrovariasjon mellom språkene, noe som gjør sammenligningen av disse språkene som L2, enda mer interessant.

Artiklene i temaheftet

Plassering av argumentene subjekt og direkte objekt relativt til negasjon i norsk, er gjenstand for variasjon. I ikke-subjektsinitiale deklarativer setninger kan subjektet stå både til høyre og venstre for negasjonen: *I går spilte {han} ikke {?han} fotball. I går spilte {Per} ikke {Per} fotball.* Også det direkte objektet kan ha denne vekslingen: *Hun likte {den} ikke {?den}. Hun likte {*huset} ikke {huset}.* Disse konstruksjonene, subjektveksling og objektsveksling, har vært gjenstand for flere studier (f.eks. Anderssen, Bentzen, Busterud, Dahl, Lundquist & Westergaard 2018), og interaksjonen mellom form, funksjon og informasjonsstruktur påvirker argumentplasseringen. I artikkelen *Tverrspråklig innflytelse fra L1 i tilegnelsen av argumentplassering i L2 norsk og svensk* undersøker Maud Westendorp og Björn Lundquist hvordan plassering av argumenter i L1 påvirker argumentplasseringen i andrespråkene norsk og svensk. Westendorp og Lundquist gjennomfører tre eksperimenter hvorav to ser på nederlandske L1-talere av henholdsvis norsk og svensk som L2, mens den siste studien ser på L2-innlærere av norsk og inkluderer en rekke ulike L1. Resultatene fra studiene tyder på at lingvistiske distinksjoner i L1 påvirker

produksjon og argumentplassering, spesielt plassering av objektet. Westendorp og Lundquist argumenterer også for at L2-innlærerne har utfordringer med å avlære plasseringsmønsteret som finnes i L1. Westendorp og Lundquist finner også at data fra grammatikalitetsvurderingstester og muntlige data fra en elisiteringstest stort sett gir sammenfallende mønster.

I artikkelen *Development of the form and meaning of definiteness in Russian-speaking learners of Swedish* presenterer Anders Agebjörn en studie av tilegnelse og utvikling av nomenfraser i ubestemt og bestemt form i svensk L2 hos russiske L1-talere. Agebjörn undersøker ubestemte modifiserte og umodifiserte nomenfraser i entall og flertall: *en katt, en vit katt, katten, den vita katten*, og bruker en eksperimentell elisiteringstest for å lokke fram muntlig produksjon av de relevante frasene. Ved å anvende denne metoden får Agebjörn undersøkt L2-innlærernes evne til å produsere morfemer som markerer (u)bestemthet, og om de produserer riktig form i riktig kontekst. Agebjörn tester innlærere på nybegynnernivå og innlærere som befinner seg på et mer avansert nivå, og kan dermed også si noe om utviklingen av bestemthet. Funnene fra studien viser at den komplekse morfosyntaktiske formen tilegnes gradvis, og Agebjörn argumenterer for at bestemte modifiserte nomenfraser tilegnes i en bestemt, implikasjonell rekkefølge: Bestemthetssuffikset på substantivet tilegnes før suffikset på adjektivet, som igjen tilegnes for den foranstilte artikkelen. Agebjörns studie viser også at det å tilegne seg en morfosyntaktisk form og det å assosiere en form med en abstrakt betydning, er to delvis ulike prosesser.

Eeva-Liisa Nyqvist presenterer en studie av andrespråkstilegnelse i en særskilt kontekst i artikkelen *Komplexa svenska nominalfraser hos finskspråkiga språkbads elever*. Hun undersøker den grammatiske kompetansen hos finske innlærere av svensk som deltar i språkbadsprogram i barnehage og senere i grunnskolen. Ved hjelp av en flervalgstest tester og sammenligner Nyqvist hvordan henholdsvis 77 og 86 elever på 6. og 9. trinn bedømmer grammatikaliteten i nominalfraser. Nominalfrasene er valgt ut for å representere to ulike typer kompleksitet, formell (dobbel bestemthet) og semantisk (possesivpronomen+adjektivattributt, genitivattributt med adjektivattributt, to definite pronomen). Formålet med studien er todelt. Felles for alle nominalfrasetyperne er at de inneholder elementer med lav saliens og at de er lavfrekvente i målspråket. Nyqvist vil finne ut hvilke av frasetyperne som oftest blir bedømt korrekt på individ- og gruppenivå, og om og eventuelt hvordan de to elevgruppene skiller seg fra hverandre i grammatikalitetsbedømmelsene. Mestring på individnivå er analysert ved hjelp av implikasjonsskala for også å undersøke

hvorvidt mestring av visse frasetyper forutsetter mestring av andre. Nyqvist operasjonaliserer mestring som 75 % korrekt bedømmelse. Nyqvist finner at de formelt komplekse nominalfrasene oftere blir korrekt bedømt enn de semantisk komplekse frasetyper, i begge elevgruppene, og både på individ- og gruppenivå. Hun finner også at frekvens spiller en rolle ved at de mest frekvente av frasetyperne oftest bedømmes korrekt. Et påfallende funn er at flere informanter ikke oppnår 75 % korrekthet i bedømmelsen av noen av de undersøkte frasetyperne, selv om de har deltatt i språkbadsprogram over en årrekke.

Isabel Nadine Jensen og Marit Westergaard presenterer i *Flaskehalshypotesen: Syntaks og morfologi blant norske andrespråksinnlærere av engelsk*, en studie som tester flaskehalshypotesen (*the bottleneck hypothesis*) til Slabakova (2008, 2013). Flaskehalshypotesen har til hensikt å forklare hva som er vanskelig og enkelt i andrespråkstilegnelse av grammatikk, og argumenterer for at elementer som tilhører den morfologiske modulen, representerer en større utfordring for L2-innlærere enn elementer som tilhører henholdsvis den syntaktiske og/eller semantiske modulen. Jensen og Westergaard tester denne hypotesen ved å se nærmere på tilegnelse av det syntaktiske fenomenet ordstilling (nærmere bestemt verbplassering i ikke-subjektsinitiale setninger og i setninger med setningsadverbial), og de morfologiske fenomenene subjekt-verbkongruens og preteritum hos norske L1-talere som lærer engelsk som L2. Gjennom to studier med til sammen 127 norske andrespråksinnlærere av engelsk, testes innlærernes evne til å identifisere ugrammatisk morfologi og syntaks ved hjelp av skriftlige akseptabilitetstester. Jensen og Westergaard finner delvis støtte for flaskehalshypotesens prediksjon om at morfologi er vanskeligere å tilegne seg enn syntaks. Deres funn tyder på at morfologi er mindre problematisk dersom det finnes likheter mellom første- og andrespråket, slik tilfellet er for markering av preteritum ved svake verb i norsk og engelsk. Samtidig viser dataene deres at ordstilling i setninger med setningsadverbialer er like vanskelig som subjekt-verb-kongruens, noe som er problematisk for flaskehalshypotesen fordi ordstilling er et syntaktisk fenomen. Jensen og Westergaard konkluderer med at tilhørighet til lingvistisk modul ikke alene kan forklare hva som er vanskelig eller lett å tilegne seg i et andrespråk, og peker på påvirkning fra førstespråket og variasjon i input som faktorer som også spiller inn.

Også Kristin Killie ser på tilegnelse av subjekt-verb-kongruens blant norske L1-talere som lærer seg engelsk som L2. Nærmere bestemt ser Killie på samsvarsbøyning ved *be* i engelsk i sin artikkel '*I are in New York*': om tilegnelsen av samsvarsbøyning i engelsk blant innlærere med norskspråklig bakgrunn.

Kille gjør bruk av skriftlige korpusdata fra *Corpus of Young Learner Language*, som består av tekster fra nasjonale prøver i engelsk. Innlærerne i studien er henholdsvis 12–13-åringer og 15–16-åringer. Killie undersøker korrekt og ukorrekt bruk av *be* og i presens og preteritum. Hun finner at en del 12–13-åringer overgeneraliserer formen *are* i presens, mens dette ikke forekommer i like stor grad blant 15–16-åringer. Killie argumenterer for at overgeneraliseringen av *are* trolig skyldes fonetisk og grafemisk likhet med den norske formen *er*. Begge aldersgruppene overgeneraliserer *was* i preteritum, og det argumenteres for at dette skyldes at formen har høy frekvens. Killie argumenterer også for at tilegnelsen av *be*-formene påvirkes av drilling, frekvens i input og output og ‘chunk-læring’. I likhet med Jensen og Westergaard analyserer Killie dataene sine opp mot flaskehalshypotesen (Slabakova 2019). Killie peker på at hennes funn viser at tilegnelse av samsvarbøyning ved *be* er en langsam prosess, og dataene støtter dermed hypotesen om at tilegnelse av morfologi er problematisk for L2-innlærere.

I artikkelen *Adjektivlæring hos voksne innlærere av norsk – ein empirisk studie* utforsker Anna-Marie Kjøde Olsen bruken av adjektiver i 300 skriftlige tekster hentet fra Norsk andrespråkskorpus (ASK), som er vurdert til å være på tre ulike ferdighetsnivå (A2–B2) etter *Det felles europeiske rammeverket for læring, undervisning og vurdering av språk* (Utdanningsdirektoratet 2011). Et viktig formål med studien er å skaffe frem et større empirisk grunnlag for bruk og tilegnelse av adjektiv i norsk innlærerspråk. I et ordforrådsperspektiv er også adjektiv i liten grad studert i skandinaviske innlærerdatabaser med unntak av noen få svenske (Axelsson 1994 og Kotsinas 1982) og norske studier (Svendsen 1996). I disse studiene ble blant annet betydningen av adjektivers frekvens i målspråket for tilegnelsen undersøkt. Frekvens er også en sentral variabel i Olsens studie, som hun forankrer i en bruksbasert tilnærming. I tillegg til frekvens undersøker Olsen hvilken rolle førstespråkspåvirkning (somal, tysk, vietnamesisk), utdanningsnivå og ferdighetsnivå har på bruken av adjektiv i tekstene. Olsen sammenligner også adjektivbruken i andrespråkstekstene med bruken av adjektiv i 100 tekster fra det norske kontrollkorpuset i ASK. Olsen finner at tokenfrekvensen av adjektiv er lavere i innlærertekstene enn i tekstene skrevet av L1-brukere, og også at den leksikalske variasjonen er større i kontrolltekstene. Hun finner også en sammenheng mellom tokenfrekvens og førstespråksbakgrunn ved at de tyskspråklige innlærerne bruker adjektiv signifikant oftere enn innlærerne med henholdsvis somali og vietnamesisk som førstespråk. Ferdighetsnivå og utdanningsnivå ser ikke ut til å ha nevneverdig betydning

for adjektivbruken. Et viktig funn er også den omfattende individuelle variasjonen mellom andrespråksinnlærerne, som Olsens studie synliggjør.

I artikkelen *Syntaktisk kompleksitet på tvers av rammeverksnivå* av Ann-Kristin Helland Gujord analyseres til sammen 120 tekster skrevet av voksne innlærere av norsk som andrespråk. Som Olsen, analyserer Gujord tekster fra ASK-korpuset (i tillegg til noen få tekster hentet fra arkivet til Norsk språktest, Folkeuniversitetet). Tekstene er vurdert til å være på fire ulike ferdighetsnivå i *Det felles europeiske rammeverket for læring, undervisning og vurdering av språk* (Utdanningsdirektoratet 2011), og hvert nivå er representert med 30 tekster hver (A2–C1). Formålet i denne studien er å undersøke sammenhengen mellom mål på syntaktisk kompleksitet og ferdighetsnivå, og mellom korrekthet og ferdighetsnivå. Studien er forankret i en forskningstradisjon, CAF-tradisjonen (Complexity, Accuracy, Fluency), som i liten grad er blitt utforsket i norsk andrespråksforskning, med unntak av tre tidligere norske studier (Moe 2013, Berggreen og Sørland 2016, Sørensen 2017). Et mål i denne studien er også å prøve ut en statistisk metode, ordinal logistisk regresjon, for å kunne utforske hvordan og i hvilken grad kompleksitetsmålene er relaterte til hverandre. Gujord finner at det er en signifikant positiv sammenheng mellom setningslengde og rammeverksnivå, fraseutbygging og rammeverksnivå og mellom korrekthet og rammeverksnivå, og at det er en signifikant negativ sammenheng mellom rammeverksnivå og sidestilt koordinasjon og underordning. Et viktig funn er også at modellen først og fremst estimerer sannsynligheten for plassering på nivå B1 og B2 med unntak av målet på korrekthet, som også skiller godt mellom de øverste nivåene (B2 og C1). Gujord har ikke på noen steg i prosessen vært involvert i vurderingen av sin egen artikkel.

Hver av artiklene i dette temaheftet er fagfellevurdert av to fagfeller. Vi vil takke alle fagfellene for deres innsiktsfulle og konstruktive innspill. Takk også til *Norsk lingvistisk tidsskrift* for muligheten til å lage et temahefte om andrespråkstilegnelse og for at vi fikk være gjesteredaktører for dette temaheftet.

Oslo og Bergen, mai 2021.

Guro Busterud og Ann-Kristin Helland Gujord

Litteratur

- Anderssen, Merete, Kristine Bentzen, Guro Busterud, Anne Dahl, Björn Lundquist & Marit Westergaard. 2018. The acquisition of word order in L2 Norwegian: The case of subject and object shift. *Nordic Journal of Linguistics* 41-3, 247–274.
- Axelsson, Monica. 1994. *Noun phrase development in Swedish as a second language: A study of adult learners acquiring definiteness and the semantics and morphology of adjectives*. Doktorgradsavhandling, Stockholm: Centrum för tvåspråkighet, Stockholms universitet.
- Kotsinas, Ulla-Britt 1982. MINS, Stockholms universitet. MINS. Stockholm.
- Rindal, Ulrikke. 2020. English in Norway: A language and a school subject in transition. I Brevik, Lisbeth, M. & Ulrikke Rindal (red.): *Teaching English in Norwegian classrooms: From research to practice*. Oslo: Universitetsforlaget, 23–42.
- Slabakova, Roumyana. 2008. *Meaning in the Second Language*. Berlin: Mouton de Gruyter.
- Slabakova, Roumyana. 2013. What is easy and what is hard to acquire in a second language: A generative perspective. I Mayo, García, Maria del Pilar, Maria Junkal Gutiérrez Mangado & Maria Martínez-Adrián (red.): *Contemporary approaches to second language acquisition*. Amsterdam/Philadelphia: John Benjamins, 5–28.
- Slabakova, Roumyana. 2019. The Bottleneck Hypothesis Updated. I Ionin, Tania & Matthew Rispoli (red.): *Three Streams of Generative Language Acquisition Research: Selected papers from the 7th Meeting of Generative Approaches to Language Acquisition – North America, University of Illinois at Urbana-Champaign*. Amsterdam: John Benjamins, 319–345.
- Svendsen, Bente Ailin. 1996. *Adjektiv i norsk som andrespråk: En frekvensbasert leksikalsk-semantisk tverrsnittstudie av adjektivtilegnelse hos voksne innlærere*. Hovedoppgåve, Institutt for nordistikk og litteraturvitenskap, Universitetet i Oslo.
- Utdanningsdirektoratet. 2011. *Det felles europeiske rammeverket for språk: Læring, undervisning, vurdering*. Oslo: Utdanningsdirektoratet.



Tverrspråklig innflytelse fra L1 i tilegnelsen av argumentplassering i L2 norsk og svensk

Maud Westendorp og Björn Lundquist

UiT Norges Arktiske Universitet

Plasseringen av subjekt og objekt i norsk og svensk avhenger av mange ulike faktorer. Det har tidligere blitt vist at andrespråkstalere i grammatikalitetsvurderinger ikke er sensitive til finkornete distinksjoner som er avgjørende for argumentplassering i norsk (Anderssen mfl. 2018). I denne artikkelen presenterer vi resultat fra tre eksperimenter som tester plassering av subjekt og objekt hos andrespråksinnlærere av norsk og svensk, med hensyn til både forståelse og produksjon. Vårt mål er å teste hvor mye argumentplasseringen i førstespråket (L1) påvirker plasseringsmønsteret i andrespråket (L2). Resultatene våre viser at L1 spiller en stor rolle: L2-innlærerne forventer at argumentplassering blir styrt av samme faktorer i andre- som i førstespråket. Videre finner vi, i likhet med Anderssen mfl. (2018), at plassering av nominalfrasesubjekt er særdeles vanskelig for L2-talere. Til slutt viser vi at mønster fra grammatikalitetsvurderinger og L2-produksjon stort sett stemmer overens.

Nøkkelord: andrespråkstilegnelse, argumentplassering, tverrspråklig innflytelse, L2 norsk, L2 svensk

1 Innledning*

I alle de nordiske språkene står subjektet i de fleste tilfeller i første posisjon i deklarativer setninger (60–80 %, Håkansson 1997: 50). Når et annet ledd står i

* Denne forskningen var støttet av doktorgradsprosjektet til M. Westendorp ved UiT Norges Arktiske Universitet og NFR-prosjektet *Micro-variation in Multilingual acquisition & attrition Situations* (MiMS, PI: Marit Westergaard, nr: 250857) som dekket utgifter forbundet med feltarbeidet. Vi vil takke våre deltakere ved Universitet i Amsterdam og UiT. Takk også til Matias Myhre, Charlotte Sant og Øystein A. Vangnes for språkvaske av artikkelen, samt NLT-fagfellene og redaktørene for mange nyttige kommentarer.

førsteposisjonen, havner subjektet etter verbet som følge av V2-regelen. I slike tilfeller, dvs. i ikke-subjekt initiale deklarativer, kan subjektet stå i ulike posisjoner: enten rett etter verbet, eller etter negasjon eller et annet setningsadverb. Pronomen havner alltid før negasjon, bortsett fra når de er sterkt kontrastive (se (1a)). Nominalfrasesubjekt kan derimot stå både før og etter negasjon, selv uten en slik kontrast (1b) (Svenonius 2002).

1. a. I går spilte **{han}** ikke **{?han}** fotball.
- b. I går spilte **{Per}** ikke **{Per}** fotball.

I norsk står objektet alltid etter hoved verbet, med mindre det står i første posisjon som følge av topikalisering. Setningsadverb og post-verbale subjekt kan gjøre at flere ledd ender opp mellom et finitt hovedverb og objektet. Hvor objektet plasseres, avhenger hovedsakelig av hvordan det er realisert: Pronominale objekt plasseres helst før setningsadverb (2a), mens NP-objekt står etter slike adverb eller negasjon (2b) (se Bentzen (2014) for en oversikt av litteraturen samt dialektvariasjon).

2. a. Hun likte **den** ikke **{?den}**.
- b. Hun likte **{*huset}** ikke **{huset}**.

I denne artikkelen ser vi på L2-innlæring av både norsk og svensk, og det er derfor viktig å påpeke at det finnes en del mindre forskjeller mellom norsk og svensk i disse ordstillingsmønstrene. Begge språk har likevel til felles at plassering av argument i stor grad er styrt av argumentets form: Pronomen havner ofte før setningsadverb, mens nominalfraser (NP) ofte ender opp etter adverb.

Så vidt vi vet, er det bare én studie som har sett på tilegnelse av disse ordstillingsmønstrene hos andrespråkstalere av norsk. Anderssen, Bentzen, Busterud, Dahl, Lundquist & Westergaard (2018) testet tilegnelsen av både subjekt- og objekt plassering vis-à-vis negasjon i norsk som andrespråk (L2) hos voksne med ulik språkbakgrunn ved hjelp av grammatikalitetsvurderinger. Generelt ga L2-talerne høy skår til *subjekt – negasjon*-sekvenser og lav skår på setninger med ordstillingen *objekt – negasjon*, uavhengig av form på argumentet. Sammenlignet med L1-talere, virker altså L2-innlærerne å legge mindre vekt på skillet mellom pronominal argumenter og NP-argumenter, og større vekt på syntaktisk funksjon. I denne studien bygger vi videre på studien til Anderssen mfl. (2018), og har som mål å forsøke å løse forskjellige begrensninger i studien, samt å undersøke effekten av argumentplassering i L1 på plasserings-

mønster i L2. Vi utførte tre eksperiment hvor vi brukte både grammatikalitetsvurderinger og elisert produksjon fra L2-talere. I tillegg så vi på to ulike L2-grupper: én gruppe med nederlandsk som førstespråk, og én gruppe med mer heterogen språkbakgrunn. Videre undersøkte vi effekten av målspråket: I den nederlandske gruppen sammenlignet vi L2-innlærere av norsk med L2-innlærere av svensk. Dette gjorde vi for å anslå hvor sensitive L2-innlærere var overfor subtile egenskaper i målspråket, eller om (og i hvilken grad) de fulgte mer generelle L2-strategier i språktilegnelsen. Vi går gjennom forskningsspørsmålene og den eksakte oppbygningen av studien i del 4. Først skal vi diskutere faktorer som styrer argumentplassering i norsk og svensk sammenlignet med andre språk i del 2. Del 3 ser på L2-tilegnelse av plasseringsmønstre i norsk og svensk. Etter en gjennomgang av oppbygningen av studien presenterer vi våre tre eksperiment (del 5–7). For hvert eksperiment beskriver vi først metoden og diskuterer deretter resultatene. Implikasjonene av disse resultatene blir diskutert til slutt.

2 Faktorer som styrer subjekt- og objektplassering

Som vi viste i (1), plasseres subjektspronomen i norsk nesten alltid før negasjon, mens nominalfrasesubjekt kan stå både før og etter negasjon. Ettersom pronomen utgjør rundt 95 % av alle subjekt (i tale fra voksne i «Tromsø acquisition corpus», Westergaard 2011: 309), forekommer ordstillingen *negasjon/adverb – subjekt* ganske sjelden. I den nordiske ordstillingsdatabasen ('Nordic Word order Database', NWD) (Lundquist, Larsson, Westendorp, Tengesdal & Nøklestad 2019), finner vi også at subjektspronomen konsekvent står foran adverb i både norsk og svensk. NP-subjekt står derimot ofte etter negasjon (70 % i norsk, 50 % i svensk i NWD). Når det gjelder subjektplassering, er det norske og svenske mønsteret ikke vanlig blant verdens språk. For det første realiseres subjektspronomen i de fleste av verdens språk som affiks på verbet (Dryer 2013). Dette gjelder for eksempel mange romanske (spansk, italiensk) og slaviske språk (tsjekkisk, serbisk). Subjektspronomen i norsk står derimot for seg selv, og plasseres alltid før negasjon. Samtidig skiller norsk og svensk seg ut i og med at NP-subjekt kan plasseres både før og etter negasjon uten at dette har noe tydelig effekt på setningens tolkning (Svenonius 2002). Dette kalles gjerne subjektveksling eller *Subject Shift* (SS) (se (1a) og (3a)). Også dette mønsteret er typologisk uvanlig: Om det finnes mer enn én subjektposisjon, er den nedre posisjonen ofte forbeholdt element som er nye i diskursen eller muligens er sterkt kontrastive (Diesing 1992). Dette gjelder for eksempel subjektplassering

i nederlandsk, hvor egennavn og bestemte nominalfraser bare kan stå etter adverbet i en kontrastiv kontekst. Dette kommer frem når vi sammenligner det norske eksemplet i (3a) med den nederlandske setningen i (3b): På nederlandsk kan egennavnet *Gerrit* bare stå etter negasjon om det er tydelig i konteksten at *Gerrit* kontrasteres med en annen NP («*ikke Gerrit men Anne*»).

3. a. I går laget {**Geir**} ikke {**Geir**} middag.
 b. Gisteren maakte {**Gerrit**} niet {^{??}**Gerrit**} het avondeten.
 i.går laget Gerrit ikke Gerrit DEF middag
 'I går laget ikke Gerrit middag.'

Norsk og svensk skiller seg i NP-subjektplassering også fra andre nordgermanske språk som dansk og islandsk, som har et mer fast mønster: I disse språkene står nemlig NP-subjekt som regel før negasjon (se Holmberg (1993), Svenonius (2002) for diskusjon; og Lundquist mfl. (2019) for produksjonsdata som underbygger dette).

Når vi ser på objektplassering, finnes det mange språk (f.eks. tysk, japansk, hindi og nederlandsk) hvor det derimot er mulig å flytte objektet uten å endre setningens tolkning. Denne flyttingen kalles ofte *scrambling*. Nedenfor fokuserer vi på *scrambling* i nederlandsk siden en stor del av våre L2-deltakere kommer fra Nederland. *Scrambling* er som regel betinget av informasjonsstrukturelle prinsipp: Bare gitte eller allerede introduserte objekt kan *scrambles*. Ikke bare pronomen, men også NP og PP-objekt *scrambler*, som i det nederlandske eksemplet i (4) (se Broekhuis & den Dikken 2012: 1061–79 for en diskusjon av nederlandsk *scrambling*, og Thráinsson 2001 for en diskusjon om forskjeller mellom *scrambling* og *Object Shift*).

4. a. Anne heeft {**de pasta**} waarschijnlijk {**de pasta**} gegeten. [nederlandsk]
 Anne har DEF pasta sannsynligvis DEF pasta spist
 'Anne har sannsynligvis spist pastaen.'
 b. Anne heeft {**op mijn email**} niet {**op mijn email**} gereageerd.
 Anne har på min e-post ikke på min e-post reagert
 'Anne har ikke reagert på eposten min.'

Selv om *scrambling* gjør at mange fraser kan forekomme til venstre for negasjon og andre adverbialer i nederlandsk, opprettholdes riktignok ordstillingen *subjekt – objekt* alltid. *Scrambling* er som nevnt ikke mulig med alle typer

direkte objekt. Objektspronomen står sjeldent etter negasjon i nederlandsk (5a), dette er bare mulig med spesielt trykk og en kontrastiv lesning (Bouma & De Hoop 2008). I motsetning til bestemte nominalfraser *scrambler* ubestemte nominalfraser nesten aldri (5b). Denne effekten av form og bestemthet kan også tolkes som en effekt av informasjonsstruktur (pronomen tilsvarer gitt informasjon, mens ubestemte NPer er nesten utelukkende nye i diskursen).

5. a. Anne heeft {**dat**} niet {^{??}**dat**} gezegd. [nederlandsk]
 Anne har det ikke det sagt
 ‘Anne har ikke sagt det.’
- b. Anne heeft {^{??}**een boek**} niet {**een boek**} gelezen.
 Anne har en bok ikke en bok lest
 ‘Anne har ikke lest en bok.’

De nordiske språkene har ikke *scrambling*, men såkalt objektsveksling (*Object Shift* (OS)). Objektsveksling påvirker bare pronominale objekt (og i islandsk også NP-objekt), se (6a–b).¹ I tillegg er OS betinget av at det finitte hovedverbet har flyttet til V2-posisjonen, og OS er derfor ikke mulig i (6c), siden et hjelpeverb står i V2-posisjonen (*Holmberg’s Generalization* (1986)).

6. a. Per så {**den**} ikke {[?]**den**}.²
 b. Per så {***ballen**} ikke {**ballen**}.
 c. Per har {***ballen/*den**} ikke sett {**ballen/den**}.

Svensk har samme plasseringsmønster for objekt som norsk, med den forskjell at objektsveksling av pronomen alltid er valgfritt (Josefsson 2003). Bentzen mfl. (2013) viser forskjellen mellom språkene i korpusstudiene sine: OS med pronomen (med nominell referent, som i (6a)) skjer i norsk i 87 % av tilfellene mens bare 64 % av disse objektene veksler i svensk. I NWD (Lundquist mfl. 2019) finner vi lignende tall: Objektsveksling skjer i omtrent 90 % av tilfellene

-
1. Det skal bemerkes at det er et geografisk skille i distribusjonen av OS i norsk; OS er nesten kategorisk i de fleste dialektene, men i Trøndelag og deler av Vestlandet blir også ordstillingen der pronominale objektet forblir i sin opprinnelige posisjon godtatt (Østbø Munch 2013: 138f.).
2. I de fleste tilfellene er det ikke OS når et pronominalt objekt refererer til et ikke-nominalt element (f.eks. en hel setning/VP) som i «Per synes ikke **det**.» (Bentzen & Anderssen 2019). Bentzen, Anderssen & Waldmann (2013) viser at «**det**» som refererer til VP/frase står i sin opprinnelige posisjon etter adverbialt i 95 % av tilfellene i norsk og 99 % av tilfellene i svensk.

i norsk og 60 % i svensk. En annen forskjell mellom språkene er at svensk, men ikke norsk, i tillegg har såkalt lang objektsveksling som gjør at lette objektspronomen kan plasseres til venstre for et NP-subjekt (se (7)). Denne typen objektsveksling er begrenset til korte objekt, som *sig* og *mig*, og er bare mulig når subjektet er en leksikalsk nominalfrase (NP eller egennavn).

7. a. I går hjalp {*meg} læreren {meg} med leksene. [norsk]
 b. Igår hjälpte {mig} läraren {mig} med läxan. [svensk]

3 Argumentplassering hos andrespråksinnlærere

3.1 Tidligere forskning på argumentplassering i L2/Ln norsk

Anderssen mfl. (2018) testet 59 voksne andrespråkstalere av norsk angående deres subjekt- og objektplasseringspreferanser ved hjelp av grammatikalitetsvurderinger, og de sammenlignet denne gruppen med en kontrollgruppe av førstespråkstalere. L2-innlærerne kom fra forskjellige deler av verden og gikk på norskkurs på tre ulike nivåer (A2–C1 på Europarådets nivåskala (CEFR)). Norskferdighetene deres ble i tillegg testet med en såkalt luketest bestående av 20 setninger som manglet ord som deltakerne skulle fylle inn. Testens mål er å gi innsikt i deltakernes generelle leksikalske og morfosyntaktiske kunnskaper om norsk.

Resultatene fra grammatikalitetsvurderingstestene viste at de aller fleste L2-innlærerne foretrakk å ha subjekt før, og objekt etter setningsadverb. Dette gjorde de nesten uavhengig av formen på argumentet (dvs. pronomen eller NP). L2-innlærere veksler altså ikke på bakgrunn av lingvistisk kategori (NP/pronomen) eller informasjonsstruktur (kjent/ny informasjon), men det er heller grammatisk funksjon (subjekt/objekt) som er hovedfaktoren i variasjonen (Anderssen mfl. 2018: 260–263). L2-innlærerne med høyest nivå i norsk (dvs. de med høyest skår på luketesten) ble mer målspråkslike i objektplasseringen, der det så ut som at de hadde lært å skille mellom leksikalske nominalfraser og pronominal objekt (2018: 264f). Plasseringen av pronominal subjekt var også mer målspråkslik blant disse talerne (større preferanse for pre-adverbial plassering). Derimot var deres plassering av NP-subjekt overraskende lengre unna målspråksgrammatikken (dvs. disse deltakerne hadde preferanse for ordstillingen *NP-subjekt – negasjon*).

Studien til Anderssen mfl. har en del begrensninger, noe som først og fremst gjør det vanskelig å undersøke innflytelsen av førstespråket i tilegnelsesprosessen. For det første hadde deltakerne i studien nokså ulik språkbakgrunn (27 ulike førstespråk). Anderssen mfl. bemerker faktisk at resultatene viste tendenser til at deltakernes L1 kan utgjøre en betydelig faktor, men siden studien ikke var designet med dette for øye, kunne ikke deltakernes L1 bli inkludert i den statistiske modellen (2018: 272). Videre testet ikke Anderssen mfl. bestemte NP-objekt. Mange av setningene i studien inneholdt derimot ubestemte NP-objekt. Et slikt eksempel fra materialet er gjengitt i (8a) (kontekstsetning i kursiv) med en nederlandsk versjon i (8b). Som vi diskuterte i del 2, og som også (8b) illustrerer, står ubestemte NP-objekt sjelden før setningsadverbialet i nederlandsk og andre språk som har objekt-*scrambling*. Det er mulig å tenke seg at det for talere med slik L1-bakgrunn (dvs. nederlandsk, eller et annet språk med objekt-*scrambling*) derfor kan ha virket som mange NP-objekt i testmaterialet til Anderssen mfl. ikke kunne veksles.³

- 8 a. *Det var salg på kjøpesenteret.* Mona kjøpte ikke klær der i går.
 8 b. Mona kocht daar {*kleren} geen {kleren} gisteren. [nederlandsk]
 Mona kjøpte der klær ingen klær i går

Dette kan være en mulig forklaring for at disse talerne hadde mer norsk-lignende vurdering av objektplassering. Det er derfor vanskelig å si om lav skår i disse kontekstene er et resultat av innflytelse fra L1 eller om deltakerne faktisk har tilegnet seg betingelsene for objektsveksling. I studien ble heller ikke forholdet mellom pronominal objekt og NP-subjekt (lang objektsveksling) testet. Dette er spesielt interessant for deltakere med klitiske pronomen i førstespråkene sine. Til slutt påpeker Anderssen mfl. (2018: 269) i studien at en produksjonsstudie av L2-innlærere kan være gunstig i fremtidige undersøkelser. Selv om det er sannsynlig at strukturer som får høy grammatikalitetsvurdering i deres studie også er strukturer som talere vil produsere i et produksjonseksperiment, er dette noe som må undersøkes nærmere.

3.2 Utfordringer i L2-tilegnelse av argumentplassering

I del 2 diskuterte vi at det er typologisk uvanlige trekk i norsk og svensk som gjør argumentplasseringen spesiell: Plassering av NP-subjekt vis-à-vis negasjon

3. 12 av 59 deltakere i Anderssen mfl. hadde nederlandsk eller tysk som førstespråk (2018: 256).

(i ikke-subjektsinitiale deklarativer) er ikke åpenbart styrt av informasjonsstruktur og semantikk, og objektplassering er styrt av forskjellen mellom ulike realiseringer (pronomen vs. NP), ikke informasjonsstruktur. Dette, i tillegg til betydelig mikrovariasjon⁴ i plasseringsmønsteret, gjør at argumentplassering er vanskelig å tilegne seg, spesielt for andrespråksinnlærere. En annen utfordring er at det ikke finnes mange kontekster som synliggjør plasseringsreglene. Subjektsveksling er for det første avhengig av at subjektet ikke står initialt (dette er tilfellet i 20–40 % av setninger (Håkansson 1997)). For det andre kan subjektets eksakte posisjon kun fastsettes hvis det finnes et setningsadverb i setningen. Det er naturlig å tenke seg at konstruksjoner med ikke-initiale subjekt som i tillegg står sammen med setningsadverb er sjeldne. I tillegg er det mulig med vekslet NP-subjekt (ca. 20–30 %), så det finnes bare en liten andel setninger som kan tydeliggjøre for andrespråksinnlærere at NP-subjekt i norsk kan stå *etter* setningsadverb (f.eks. (1b) *J går spilte ikke Per fotball.*). Videre er objektsveksling avhengig av at det finite hoved verbet står i V2-posisjonen, hvilket kraftig minsker forekomsten av kontekster for objektsveksling. Tengesdal og Lundquist (2019) viser at så lite som 25 % av alle hovedverb står i V2-posisjonen i norsk talemål. Det er altså få kontekster som kan signalisere til innlæreren hva som er mønsteret i norsk.

Hvordan tilegner L2-talere seg så argumentplassering i andrespråket? For en del syntaktiske fenomen, og i mange forskjellige språk, har det blitt vist at andrespråksinnlærere enten overgeneraliserer flytting (se f.eks. Clahsen & Muysken (1986) om verbflytting i L2-tyske underordnede setninger) eller ikke flytter i det hele tatt (f.eks. (manglende) V2-flytting i norsk som andrespråk (Eide 2015)). Den siste strategien blir også vist av Unsworth (2005) for L2-innlærere av nederlandsk som tilegner seg *scrambling*-reglene. Både de unge og voksne L2-talere gjennomgår nemlig et stadium der de ikke flytter objekt som egentlig skal *scramble* i nederlandsk. Unsworth argumenterer for at dette skyldes innflytelse fra L1 engelsk, hvor adverb aldri kan plasseres mellom verbet og objektet (2005: 248f). Generelt kan begrepet *tverrspråklig innflytelse* («crosslinguistic influence», Odlin 2013) brukes omtrent synonymt med andre begreper, spesielt *transfer* og *interferens*, for innflytelsen fra ett språk på et annet, vanligvis i andrespråkstilegnelse. Slik tverrspråklig innflytelse fra L1 kan ha stor effekt i tilegnelsesprosessen. Noen studier indikerer at ikke bare førstespråket, men også andre tidligere tilegnede språk kan ha innflytelse i

4. Vi bruker begrepet mikrovariasjon for å henvise til prosesser som er avhengige av fine distinksjoner i syntaks eller informasjonsstruktur som involverer undergrupper av elementer, f.eks. forskjellige typer ledd (noen, men ikke alle NPer).

tilegnelse av et tredje (eller fjerde, femte osv.) språk (f.eks. ved objekt plassering i tysk som L3 (Falk & Bardel 2011)).

Ut ifra teorien om tverrspråklig innflytelse kan det tenkes at L2-innlærere har ulike tilnærminger når de lærer seg norsk argument plassering, avhengig av hvilke språk de behersker fra før. For nederlendere som lærer norsk eller svensk som andrespråk forventer vi at de legger stor vekt på semantikk og informasjonsstruktur, dvs. faktorer som er viktige i nederlandsk argument plassering. En slik innlærer som antar at variasjonen er betinget av informasjonsstruktur (+ semantisk rekkevidde) og som har tilegnet seg lokale ordstillingsbegrensninger (subjekt før objekt, og VO-ordstilling), vil faktisk få det meste riktig i L2 norsk/svensk. Vi kan da forvente at nederlandske førstespråktalere plasserer både subjekt- og objektspronomen før adverb (på samme måte som i nederlandsk), men at også en del NP-objekt blir plassert her siden disse objektene ofte flytter over adverb i nederlandsk som resultat av *scrambling*. Den største utfordringen for nederlandske L2-innlærere av norsk og svensk vil være at NP-subjekt i norsk og svensk kan plasseres både foran og etter adverb uten å være kontrastive. L2-innlærere med L1 med verbklitikon vil få en annen type utfordring. Tverrspråklig innflytelse fra slike språk vil gjøre at disse talerne plasserer subjekts- og objektspronomen rett etter verbet og før negasjon (samme plass som klitikon), og nominalfraser etter adverbet. Disse talerne vil i så fall ha pronomenene på rett sted. I tillegg kan vi forvente at disse talerne produserer lang objektsveksling (som i L1 svensk men ikke norsk, jf. (7b)). Innflytelse fra engelsk, enten som førstespråk eller som tidligere tilegnet L2-språk, vil føre til enda et annet mønster. Engelsk er et SVO-språk som verken har *scrambling* eller klitikon, og setningsadverb kan aldri plasseres mellom verbet og objektet. Å bruke engelsk plasseringsmønster i norsk/svensk vil derfor lede til en sterk preferanse for ordstillingen *negasjon – objekt og subjekt – negasjon* uavhengig av om argumentet er realisert som pronomen eller nominalfrase.

4 Vår studie

Vi har som hovedhensikt med vår studie å undersøke om, og på hvilken måte argument plasseringen (her: plassering av subjekt og direkte objekt) i L1 påvirker plasseringsmønster i L2 norsk og svensk. Videre trekker vi inn produktionsdata og sammenligner dette med grammatikalitetsvurderinger. Vi utfører tre ulike eksperimenter (se oversikt i tabell 1). To av eksperimentene har talere med nederlandsk som førstespråk, og i det tredje eksperimentet har deltakerne

forskjellige førstespråk. I tillegg bruker vi ulike metoder, og målspråket er forskjellig.⁵

Tabell 1. Oppbygning av studien og oversikt av eksperimentene.

Eksp.	Metode	L1	L2 ⁵	Syntaktiske fenomen
1	Grammatikalitetsvurderinger	nederlandsk	norsk/svensk	Subjekt- og objektplassering (NP/pronomen) og lang objektsveksling (pronomen)
2	Elisitert produksjon	nederlandsk	norsk/svensk	Subjekt- og objektplassering (NP/pronomen) og lang objektsveksling (pronomen)
3	Elisitert produksjon	9 ulike L1	norsk	Subjekts- (NP/pronomen) og objektplassering (pronomen)

I det første eksperimentet bruker vi testmateriale fra Anderssen mfl. (2018) for å teste en gruppe L2-innlærere av norsk- og svensk med nederlandsk som førstespråk på deres grammatikalitetsvurderinger av norsk/svensk argumentplassering. Eksperiment 2 og 3 er produksjonseksperiment og i disse eksperiment anvender vi metoden som også er brukt i utviklingen av Nordic Word order Database (Lundquist mfl. 2019). I det andre eksperimentet elisiterer vi talespråkmateriale fra en gruppe nederlandske studenter på både norsk og svensk. Det tredje eksperimentet bruker samme elisiteringstest, men her tester vi L2-talere av norsk med mer variert språkbakgrunn. Det er litt variasjon i hvilke syntaktiske fenomen vi ser på i eksperimentene, og eksperiment 3 skiller seg ut i og med at vi ikke tester plassering av NP-objekt eller lang objektsveksling. Metodene og deltakerne i eksperimentene blir diskutert i del 5–7 som diskuterer de tre eksperimentene.

Forskningsspørsmålene kan sammenfattes på følgende måte:

1. Klarer L2-innlærere å tilegne seg interaksjonen mellom form, funksjon og informasjonsstruktur i plassering av subjekt og direkte objekt i norsk og svensk?
2. Påvirker innlærernes L1 tilegnelsen av plassering av subjekt og direkte objekt i L2 norsk og svensk?

5. Vi fortsetter å bruke begrepet 'L2(-innlærer)' her, men alle deltakerne våre behersker engelsk som L2 (og noen også flere andre språk) og kan derfor også klassifiseres som 'Ln'-talere. To deltakere i eksperiment 1 hadde 2L1: nederlandsk og frisisk/limburgsk (begge disse språk har også *scrambling*).

3. Er det noen sammenheng mellom produksjon og grammatikalitetsvurderinger for plassering av subjekt og direkte objekt i norsk og svensk som L2?

Med vår tredelte studie forsøker vi å svare på disse forskningsspørsmålene, samtidig som vi prøver å løse forskjellige begrensninger i tidligere forskning. Først og fremst prøver vi å isolere effekten av førstespråket ved å sammenligne nederlandske andrespråksinnlærere med en gruppe andrespråksinnlærere med ulike L1 (eksp. 2 vs. 3). Vi undersøker også hvordan frekvensen av ulike grammatiske trekk (husk forskjellene mellom norsk og svensk objektplassering) påvirker L2-grammatikken ved å sammenligne L2-innlærere med ulike mål-språk (eksp. 1 og 2). Her har vi valgt å inkludere svensk i tillegg til norsk siden dette gir oss mulighet til å utforske om tendensene for «vanlig» og lang objektsveksling kan forklares ut ifra forskjeller i mål-språket, eller om det forklares best ut ifra innflytelse fra L1. Til slutt trekker vi inn produksjon i tillegg til grammatikalitetsvurderinger for å kunne få et bedre og mer helhetlig bilde av variasjonen og tilegnelsesprosessen (eksp. 2 og 3). En slik L2-produksjonsstudie av norsk subjekt- og objektplassering har ikke blitt gjort tidligere, og er et potensielt viktig tilskudd til litteraturen.

5 EKSPERIMENT 1: Grammatikalitetsvurderinger med nederlandske L2-innlærere

5.1 Deltakere og metode

33 studenter med nederlandsk som førstespråk i Amsterdam (Nederland) deltok i studien. Av disse studerte 15 svensk og 18 norsk på bachelornivå (gjennomsnittsalder: 21,4 år). Vi testet ikke språkferdighetene til deltakerne, men alle hadde studert mål-språket i minst ett semester, noe som skal gi dem A2/B1 nivå (etter Europarådets nivåskala (CEFR) (Studiekatalogen UvA)).

Vi satte sammen en grammatikalitetsvurderingsstudie med 70 setninger i pseudo-randomisert rekkefølge. Testsettet inkluderte alltid to versjoner av hver setning: én med flytting, og én uten flytting av subjekt/objekt. 12 setninger testet plassering av subjekt i forhold til negasjon, 20 testet plassering av direkte objekt i forhold til negasjon (se tabell 2). For å ha noe variasjon i testen inkluderte vi også 20 underordnede setninger med og uten subjektsveksling, og 10 setninger med et ikke-referensielt objektspronomen. Vurderingene av disse setningene er *ikke* tatt med i resultatene. Denne delen av materialet (62/70 set-

ninger) er overtatt fra Anderssen mfl. (2018) med tillatelse. I tillegg satte vi opp 8 nye setninger som testet lang objektsveksling (se (9)). Bemerk at lang objektsveksling av et refleksivt objekt alltid er ugrammatisk i norsk, og bare tillatt med NP-subjekt i svensk.

9. *I dag har Kalle et jobbintervju. Derfor barberer han seg nå.*

1 2 3 4 5 6

Deltakerne ble spurt om å gi en gradering fra 1 (helt uakseptabel) til 6 (helt akseptabel). Alle setningene ble innledet av en kontekstsetning (i kursiv) som deltakerne eksplisitt ble bedt om å ikke bedømme. Samme eksperiment ble brukt på norsk og svensk.

Tabell 2. Oppbygning eksperiment 1.

Fenomen	Form/realisering	Antall
Subjektplassering		
<i>Dessverre rakk {Nils} ikke {Nils} fram i tide.</i>	NP	6
<i>Derfor dro {han} ikke {han} dit i dag.</i>	pronomen	6
Objektsveksling		
<i>Hun bruker {sminke} ikke {sminke} i det hele tatt.</i>	NP	10
<i>Magnus rakk {den} ikke {den} i morges.</i>	referensielt pronomen	10
Lang objektsveksling		
<i>Derfor vasket {seg} løperen {seg} rett etter.</i>	NP	4
<i>Derfor skilte {seg} han {seg} fra henne.</i>	pronomen	4
Fillers		
<i>Hun visste at {Per} ikke {Per} ville rekke det.</i>	underordnet SS med NP	10
<i>Han trodde at {hun} ikke {hun} visste det.</i>	underordnet SS med pronomen	10
<i>Jeg tror {det} ikke {det}, dessverre.</i>	OS med ikke-ref. pronomen	10
	Total	70

Studien ble utført i undervisningstimene på universitetet, individuelt og på papir. Instruksen ble gitt på nederlandsk og inkluderte to norske/svenske eksempler for å illustrere oppgaven. Nedenfor sammenligner vi resultatene fra våre andre-språkssinnlærere av norsk med kontrollgruppen (voksne L1-talere) fra Anderssen mfl. (2018), og den svenske L2-gruppen med en gruppe på 17 svenske førstespråkstalere (gjennomsnittsalder 33,1 år) som ble rekruttert via personlige kontakter. Deltakelsen i undersøkelsen var frivillig for alle.

5.2 Resultat

Vi skal nå presentere resultatene for de tre ulike variablene (subjekt- og objekt-plassering, og lang objektsveksling) i eksperiment 1. En oversikt over resultatene for subjekt-plassering vises i tabell 3.

Tabell 3. *Subjekt-plassering, gjennomsnittsvurderinger (skala 1–6).*

	norsk L1		norsk L2		svensk L1		svensk L2	
	NP	Pron	NP	Pron	NP	Pron	NP	Pron
Subjekt – Negasjon <i>Derfor dro Petter/han ikke dit.</i>	3,2	5,9	5,1	5,0	5,4	5,7	5,4	5,5
Negasjon – Subjekt <i>Derfor dro ikke Petter/han dit.</i>	5,9	3,6	2,5	2,4	3,7	2,9	2,3	2,1
Preferanse Sub – Neg	-2,7	2,3	2,6	2,6	1,7	2,8	3,1	3,4

Vi finner det forventede mønsteret i L2-resultatene for subjekt-plassering: L2-talerne har en sterk preferanse for pronominale og NP-subjekt foran negasjon, og de ser ikke ut til å skille mellom pronomen og NP. Det er ingen stor forskjell mellom de to L2-gruppene, verken for NP eller pronominale subjekt. For pronominale subjekt ser L1- og L2-resultatene nokså like ut. Den største forskjellen sees med NP-subjekt, og spesielt i norsk: Der L1-talere har en sterk preferanse for ordstillingen *negasjon – subjekt*, viser L2-talere en nesten like stor preferanse for motsatt rekkefølge. I L1-svensk ser vi en preferanse for subjekt før negasjon med NP, men ikke så sterk som hos L2-gruppen. Kort sagt påvirker verken form (pronomen/NP) eller målspåket (norsk/svensk) i særlig grad preferansen for L2-subjekt-plassering: Alle subjekt foretrekkes direkte etter verbet, før negasjon.

Tabell 4. *Objekt-plassering, gjennomsnittsvurderinger (skala 1–6).*

	norsk L1		norsk L2		svensk L1		svensk L2	
	NP	Pron	NP	Pron	NP	Pron	NP	Pron
Objekt – Negasjon <i>Daniel så Linda/henne ikke der.</i>	1,3	5,4	3,1	4,9	1,3	4	3	4,5
Negasjon – Objekt <i>Daniel så ikke Linda/henne der.</i>	5,7	3,8	5,1	4,1	5,2	5,1	4,4	3,6
Preferanse Obj – Neg	-4,4	1,6	-2,0	0,8	-3,9	-1,1	-1,4	0,9

I tabell 4 viser vi resultatene for objekt-plassering i setninger med hovedverb i V2-stilling. I likhet med resultatene for subjekt-plassering, ser vi ikke noen stor forskjell mellom de to L2-gruppene: Begge L2-gruppene har en sterk preferanse for å ha NP-objekt etter negasjon, i likhet med begge L1-gruppene, og en svak

preferanse for pronominale objekt før negasjon. Legg merke til at det er en ganske stor forskjell mellom L1-gruppene i vurderingen av setninger med pronomen: Den norske L1-gruppen har preferanse for *objekt – negasjon* mens den svenske har preferanse for motsatt ordstilling. Likevel får begge ordstillingene relativt høy skår hos L2-innlærere av begge språk. Oppsummert ser vi her at form (pronomen eller nominalfrase) spiller en viss rolle for L2-talernes ordstillingspreferanse, mens målspåket ikke gjør det.

Nedenfor i tabell 5 gir vi resultatene for plassering av subjekt vis-à-vis objekt (såkalt lang objektsveksling). Her har vi ingen resultat fra norske L1-talere, men vi antar at det er strengt ugrammatisk å ha objekt til venstre for subjekt i norsk (jf. Faarlund, Lie & Vannebo 1997: 857, se også resultat i del 6.3). Vi tester bare refleksive objekt (*seg/sig*) her, siden disse ikke kan tolkes kontrastivt.

Tabell 5. Plassering av subjekt ift. (refleksivt) objekt, gjennomsnittsvurderinger (skala 1–6).

	norsk L1		norsk L2		svensk L1		svensk L2	
	NP	Pron	NP	Pron	NP	Pron	NP	Pron
Subjekt – Objekt <i>Derfor barberer Pål/han seg.</i>	NA	NA	5,3	5,7	5,2	5,7	5,5	5,7
Objekt – Subjekt <i>Derfor barberer seg Pål/han.</i>	NA	NA	2,4	3,4	5,4	3,6	2,5	2,6
Preferanse Sub – Obj	NA	NA	2,9	2,3	-0,2	2,1	3,0	3,1

Her ser vi heller ingen effekt av målspåket eller form: Begge L2-gruppene foretrekker sterkt å ha subjekt før et lett, refleksivt objekt, uavhengig av om subjektet er en nominalfrase eller et pronomen. Den norske L2-gruppen viser slik sett et L1-lignende mønster her, da lang objektsveksling ansees å være lite akseptabelt på norsk. Den svenske L1-gruppen aksepterer som forventet lette objekt både før og etter NP-subjekt, men har sterk preferanse for subjektspronomen før objekt.

5.3 Diskusjon eksperiment 1

Vi fant ingen store forskjeller mellom de to L2-gruppene i eksperiment 1, selv om målspåkene og de to L1-gruppene skilte seg ganske tydelig ut for tre av variablene: (i) plassering av NP-subjekt ift. negasjon, (ii) plassering av pronominale objekt ift. negasjon og (iii) plassering av NP-subjekt ift. lette pronominale objekt. I alle tilfellene finner vi at L2-gruppene har en preferanse for ordstillingen som er typisk i nederlandsk: Ikke-kontrastive subjekt plasseres

foran både negasjon og lette objekt, pronominale objekt ender opp før negasjon, og lette objektspronomen foretrekkes etter subjekt. Det skal imidlertid nevnes at preferansen for objektsveksling er relativt svak, selv om ordstillingen *pronominalt objekt – negasjon* er mer eller mindre obligatorisk i nederlandsk. L2-gruppene aksepterer heller ikke NP-objekt etterfulgt av negasjon i stor grad, selv om såkalt *scrambling* av NP-objekt over negasjon er mulig i deres L1. Som vi nevnte i del 3.1, kan imidlertid den lave akseptabiliteten av *(NP-)objekt – negasjon* forklares med det faktum at mange av objektene i testen er ubestemte, noe som sjelden fører til *scrambling*.

Våre resultat er forskjellige fra Anderssen mfl. (2018) når det gjelder objektsveksling. I deres studie foretrakk deltakerne å ha *alle* objekt etter negasjon. Våre deltakere, som ut ifra kursnivå er sammenlignbare med den minst erfarne gruppen hos Anderssen mfl. (A2/B1 nivå), er mer målspråkslike i sine preferanser for plassering av objektspronomen. Det er likevel vanskelig å si om den observerte forskjellen mellom NP og pronominale objekt viser at våre L2-talere har tilegnet seg at form påvirker plassering i større grad enn deltakerne i Anderssen mfl., eller om de bare følger det nederlandske *scrambling*-mønstret. Vi kommer tilbake til dette i diskusjonen av eksperiment 2.

6 EKSPERIMENT 2: Produksjonsstudie med nederlandske andrespråksinnlærere

I eksperiment 1 brukte vi grammatikalitetsvurderinger for å undersøke argumentplassering i norsk og svensk hos nederlandske L2-innlærere. Resultatene ga oss innsikt i informantenes oppfatning av forskjellige ordstillinger, men produksjonsdata vil gi oss et bedre og mer helhetlig bilde av variasjonen og tilegnelsesprosessen. Dette er målet med eksperiment 2.

6.1 Deltakere

Til dette eksperimentet rekrutterte vi 21 nederlandske L1-talere fra Universitetet i Amsterdam (UvA). Studien ble utført nesten ett og et halvt år etter eksperiment 1, der noen av de samme studentene deltok. Deltakerne utgjorde to grupper: 12 L2-innlærere av svensk, 9 L2-innlærere av norsk (se tab. 6). Alle hadde studert norsk/svensk ved UvA i minst 8 måneder, og 7 hadde fullført hele bachelorprogrammet i nordisk (3 norsk, 4 svensk).⁶ Før eksperimentet gjorde deltakerne

6. På dette punktet i studieåret forventes studentene å fungere som 'uavhengige brukere' på Europarådets nivåskala (reseptivt: nivå B2, produktivt nivå B1) (Studiegids UvA).

en språkferdighetstest (luketest, samme som Anderssen mfl. (2018)) hvor de fleste deltakerne fikk alt rett (sammenlagt gjennomsnitt 35/36). Dette resultatet kan tyde på at testen var for «lett» for våre deltakere. En utfordring er derfor at vi ikke klarer å skille deltakere på ulike nivå fra hver andre, vi kommer tilbake til dette i diskusjonen. Deltakelsen var frivillig, og informantene fikk ikke betaling.

Tabell 6. Deltakeroversikt eksperiment 2.

Målspråk	Antall	Deltok i eksp. 1	Alder	Gjennomsnittsalder	Luketest
L2 norsk	9	2	18–60	29,1	34,4
L2 svensk	12	5	20–33	24,4	35,5
Total	21	7	18–60	26,4	35,0

6.2 Metode og prosedyre

Denne undersøkelsen er en del av Nordic Word order Database-prosjektet (NWD) (Lundquist mfl. 2019) og bruker en justert versjon av forskningsmetoden i prosjektet. Materialet, opptak og dokumentasjon er tilgjengelig på nettet i *ordstillingsdatabasen*. Produksjonseksperimentet besto av 40 setningspar presentert på dataskjerm (se 10): Først vises bakgrunnssetningen, som deltakerne blir bedt om å lese høyt. Deretter blir triggeren, starten av en ny setning, vist på skjermen. Deltakerne skal fullføre setningen ved hjelp av ordene fra bakgrunnssetningen (forventet respons i hakeparenteser).

10. (Bakgrunn) Læreren ga meg en ny bok i går.
 (Trigger) I går ... [ga læreren meg en ny bok].

Eksperimentet besto av tre deler med ulike oppgaver (se oversikt og eksempler i tab. 7); setningene ble presentert i tilfeldig rekkefølge innenfor de ulike delene og hver del startet med fire øvelsessetninger.

Vi testet plassering av subjekt med både pronominale subjekt og NP-subjekt (enten ikke-modifiserte bestemte nominalfraser som *læreren* eller egennavn). Materialet for (lang) objektsveksling omfattet fire ulike typer objekt: lette refleksive objekt (*seg*), førstepersons pronomen (*meg*), tredjepersons pronomen (entall og flertall) og nominalfraser.

Eksperimentet ble satt opp i programvaren OpenSesame (Mathôt, Schreij & Theeuwes 2012). Ettersom eksperimentet er svært enkelt, forsto de fleste

Tabell 7. Oppbygning eksperiment 2.

Del	Oppgave	Fenomen	Antall setninger
1	1: Subjekt–verb inversjon <i>Hun tok ikke bussen til jobb i går. > I går ...</i>	subjektsveksling	10
		lang objektsveksling	15
2	2: Sammensatt til enkel verbform <i>Ranerne ble ikke arrestert av politiet i går. > Politiet arresterte ...</i>	subjektsveksling	5
		objektsveksling	5
3	3: Passiv til aktiv <i>Han ble ikke truffet av kulen. > Kulen traff ...</i>	pron. objektsveksling	6
		NP-objektsveksling	5
Total			46

deltakerne oppgaven med én gang og vi trengte ikke å gi spesifikke instruksjoner. Alle elementene i eksperimentet var bygd opp på følgende måte:

- i. Bakgrunnssetning på svart skjerm, hvit skrift (1000 ms);
- ii. Pipelyd (300 ms) hvoretter setningen blir rød for å vise at setningen skal leses høyt;
- iii. Etter et tastetrykk av eksperimentleder blir begynnelsen av den nye setningen presentert i hvit skrift under den første setningen (som forblir på skjermen)⁷;
- iv. Pipelyd (300 ms) hvoretter setningen blir rød, og deltakeren fullfører setningen.

Lydopptak ble gjort underveis med hjelp av en ekstern mikrofon. Opptakene ble automatisk segmentert med hjelp av tidsangivelsene samlet i OpenSesame. Vi la så til merknader i ELAN (Wittenburg, Brugman, Russel, Klassmann & Sloetjes 2006), for å notere hvilken ordstilling som ble produsert: f.eks. SN (*subjekt – negasjon*) eller NS (*negasjon – subjekt*).

6.3 Resultat

Vi presenterer resultatene for de tre variablene i samme rekkefølge som i eksperiment 1. I stedet for gjennomsnittsverdier for de to ordstillingene som testes, gir vi relativ fordeling uttrykt i prosent. Hver tabell inneholder en tredje rad som heter «annet». Den omfatter setninger der deltakerne gjorde en eller annen feil. (I de fleste tilfellene skyldtes det at de glemte et ord.) Resultatene for L1

7. Det er selvfølgelig mulig å gjenta bakgrunnssetningen slik den står på skjermen i stedet for å danne en ny setning. I så fall får man en ugrammatisk ikke-V2 setning (**I går ... læreren ga meg en ny bok.*). Ingen av våre deltakere produserte noe slikt.

kommer fra NWD (Lundquist mfl. 2019) og er basert på det samme testmaterialet. Resultatene for subjektplassing presenteres i tabell 8.

Tabell 8. Subjektplassing, produksjon. Relativ fordeling i prosent.

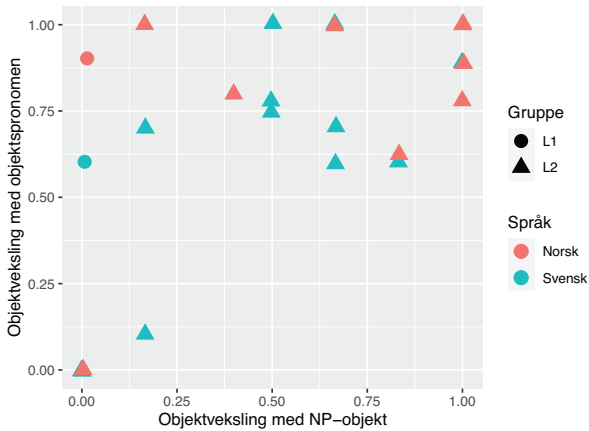
	norsk L1 %		norsk L2 %		svensk L1 %		svensk L2 %	
	NP	Pron	NP	Pron	NP	Pron	NP	Pron
Sub – Neg <i>I går tok Kalle/han ikke bussen.</i>	23,0	96,8	81,7	87,2	51,5	97,4	95,4	98,9
Neg – Sub <i>I går tok ikke Kalle/han bussen.</i>	71,9	0,1	16,1	8,3	47,8	0,5	3,8	0
Annet	5,1	3,1	2,2	4,4	3,7	2,1	0,8	0,1

Som i eksperiment 1 ser vi en sterk preferanse for rekkefølgen *subjekt – negasjon* i L2-gruppene, uavhengig av formen på subjektet. Vi ser igjen en forskjell mellom L1-gruppene i plasseringen av NP-subjekt: De norske har en preferanse for ordstillingen *negasjon – subjekt*, mens begge sekvensene er produsert like ofte hos de svenske L1-talene. Denne forskjellen ser også ut til å gjenspeiles i resultatene fra de to L2-gruppene: gruppen som lærer seg norsk som L2 produserer mer *neg – sub* enn de som lærer svensk. I det norske L2-materialet finner vi 16,1 % *neg – sub* mot 3,8 % i det svenske. Forskjellen drives fram av en mindre gruppe informanter: bare fire av ni innlærere av norsk produserer denne ordstillingen (to av disse bruker *neg – NPsub* konsekvent). I motsetning til førstespråkstalere skiller disse ikke mellom subjekt av ulik form: tre av deltakerne produserer denne ordstillingen både med NP- og med pronominale subjekt. Også hos de svenske deltakerne er det et fåtall informanter (N = 3) som står for alle forekomster av ordstillingen *neg – subjekt*. Til forskjell fra de norske L2-talene produserer disse deltakerne ordstillingen bare med NP-subjekt og aldri med pronomen.

Tabell 9. Objektplassing, produksjon. Relativ fordeling i prosent.

	norsk L1 %		norsk L2 %		svensk L1 %		svensk L2 %	
	NP	Pron	NP	Pron	NP	Pron	NP	Pron
Objekt – Neg <i>I går hjalp Marie/hun meg ikke.</i>	1,6	88,7	58,3	73,9	0	73,6	58,3	68,8
Neg – Objekt <i>I går hjalp Marie/hun ikke meg.</i>	94,7	5,1	38,9	16,4	85,9	21,1	41,7	28,6
Annet	3,7	6,2	2,8	9,7	16,1	5,3	0	2,6

I tabell 9 viser vi resultatet for objektplassering. Vi ser et forventet resultat hos L1-deltakerne: *NP – neg*-sekvensen blir nesten aldri produsert, og ordstillingen *pronomen – neg* er mer eller mindre obligatorisk i norsk (bare 5,1 % fravær av objektsveksling). De svenske resultatene viser større variasjon, men vi ser fortsatt en sterk preferanse for vekslende objekt. Vi ser at L2-gruppene også veksler de fleste pronominala objektene, men på gruppenivå litt under L1-gruppene. Vi finner den største forskjellen mellom L1 og L2-gruppene med NP-objekt: L2-gruppene plasserer i over 50 % av tilfellene NP-objekt før negasjon.



Figur 1. Objektplassering i L2 norsk og L2 svensk. Hver trekant representerer en deltaker. (Gjennomsnittsverdi for norsk og svensk L1 fra NWD (sirkler) til sammenligning.)

I figur 1 presenterer vi resultatene fra objektsveksling hos de individuelle L2-innlærerne av norsk og svensk (hvert symbol representerer én informant). Det er 3 informanter (to for norsk, en for svensk) som aldri produserer OS, verken med pronomen eller med nominalfraser. Samtidig er det også slik at alle deltakerne som produserer objektsveksling med pronomen, også produserer objektsveksling med NP minst én gang i løpet av eksperimentet. Dette står i sterk kontrast til L1-gruppene, som aldri produserer denne ordstillingen.

Til slutt viser vi resultatet for såkalt lang objektsveksling (LOS) i tabell 10, dvs. plassering av et lett objekt (*seg* eller *meg*) i forhold til subjekt. L1-mønsteret ser ut som forventet: Norske førstespråkstalere plasserer konsekvent alle typer subjekt foran objekt, mens de svenske oftere plasserer NP-subjekt etter det lette objektet. Deltakerne som studerte norsk, produserte bare sekvensen

subjekt – objekt slik som det er påkrevd i norsk (og nederlandsk). Når det gjelder svensk, som tillater lang objektsveksling, produserte fire av 12 L2-innlærere ordstillingen der det lette objektet kommer før subjektet. Det er verdt å bemerke at det er de samme deltakerne som behersket SS-reglene som ser ut til å ha tilegnet seg reglene for LOS.

Tabell 10. Plassering av subjekt i forhold til (lett) objekt i prosent, produksjon.

	norsk L1 %		norsk L2 %		svensk L1 %		svensk L2 %	
	NP	Pron	NP	Pron	NP	Pron	NP	Pron
Subjekt – Obj <i>I går ga læreren/hun meg en ny bok.</i>	95,6	96,1	97	100	36,6	96,5	80	96,7
Objekt – Sub <i>I går ga meg læreren/hun en ny bok.</i>	0,1	0,1	0	0	57,7	1,0	20	1,0
Annet	4,3	3,3	3	0	5,7	2,5	0	3,3

6.4 Diskusjon

Som i eksperiment 1 ser vi i eksperiment 2 igjen lite forskjell mellom de to nederlandske L2-gruppene med ulike målpråk. Selv i lang objektsveksling hvor (L1) norsk og svensk tydelig er forskjellige, er det bare et fåtalls L2-innlærere som skiller seg ut. Akkurat som vi så med grammatikalitetsvurderingene, der vekslede objektspronomen ble godtatt i større grad enn vekslede nominalfraser, produserer våre deltakere i eksperiment 2 mer objektsveksling med pronomen enn med NPer. Når vi sammenligner eksperimentene, ser vi på den andre siden en tendens til at deltakerne gjør mindre forskjell mellom NP- og objektspronomen i produksjon enn i sine grammatikalitetsvurderinger. Det er usikkert om dette er en effekt av testtypen (vurderinger vs. produksjon) eller av at materialet i testene er forskjellig. Siden resultatene for subjekt-plassering er nokså sammenlignbare i begge eksperimentene, er det stor sannsynlighet for at den høyere forekomsten av NP-objektveksling i produksjon er et resultat av bruk av bestemte nominalfraser og egennavn i produksjonseksperimentet. Slike fraser *scrambles* nemlig langt oftere i nederlandsk enn ubestemte nominalfraser. På individuelt nivå ser vi videre at deltakerne som aldri produserer *NP-subjekt – negasjon*, heller ikke gjør det med pronominalt subjekt (i motsetning til L1-talene), og at de som veksler pronominalt objekt, generelt også plasserer NP-objekt foran negasjon. Det ser dermed ut som om de nederlandske deltakerne som produserer objektsveksling her, appliserer et nederlandsk *scrambling*-mønster på norsk objekt-plassering.

7 EKSPERIMENT 3: Produksjon hos L2-innlærere med ulik språkbakgrunn

I eksperiment 2 finner vi at nederlandske L2-innlærere av norsk og svensk ofte plasserer objekt foran negasjon og setningsadverbial nesten uavhengig av form på objektet. I vårt første eksperiment finner vi derimot at nederlandske førstespråkstalere ofte foretrekker å ha objekt (i alle fall NPer) etter negasjon (og subjekt før negasjon). Vi beskrev dette mønstret som en effekt av både språkbakgrunn og materiale i de forskjellige eksperimentene. I tillegg ser vi i produksjonseksperimentet (eksp. 2) at alle som lærte norsk, og de fleste svenske L2-talere, ikke plasserer objekt før subjekt, dvs. at de ikke har lang objektsveksling (LOS). Dette kan tyde på at innlærerne rett og slett ikke har fått tilstrekkelig innputt av LOS, eller det kan skyldes innflytelse fra førstespråket. På den andre siden er det også mulig at tendensene vi ser hos våre nederlandske deltakere gjenspeiler en mer generell L2-strategi. Derfor sammenligner vi i det siste eksperimentet, resultatene fra den homogene nederlandske gruppen som tilegner seg L2 norsk (eksp. 2), med en gruppe L2-innlærere av norsk med forskjellig språkbakgrunn i det samme produksjonseksperimentet.

7.1 Deltakere og metode

Metode og materiale i eksperiment 3 er likt eksperiment 2, minus del 3 (objektsveksling med nominalfraser), som ikke er inkludert her. Vi utførte testen kun med innlærere av norsk som L2. 21 voksne deltakere (23–45 år) ble rekruttert ved UiT Norges arktiske universitet i Tromsø. Deltakerne hadde ulike førstespråk: engelsk, russisk, finsk, tysk, bulgarsk, romansk, tsjekkisk, spansk og kroatisk. Alle deltakerne hadde bodd i Norge i minst et halvt år, men de fleste hadde vært i Norge i mange år (i gjennomsnitt 7,9 år, maks 16 år). Alle bortsett fra én deltaker snakket og hørte norsk daglig. Resultat fra ferdighetstesten (gjennomsnittsskår 34/36), som vi også brukte i det andre eksperimentet, viste at informantene var på lignende nivå med de i eksperiment 2. Det er ingen korrelasjon mellom antall år deltakerne hadde vært bosatt i Norge og deres skår på ferdighetstesten. Alle fikk en liten gave for deltakelsen i eksperimentet.

7.2 Resultat og diskusjon

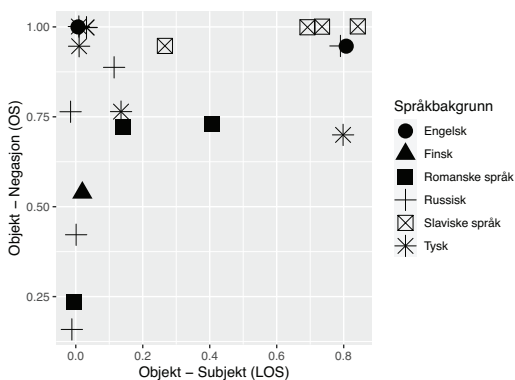
Tabell 11 viser resultatene fra eksperiment 3 (i LOS kolonnene refererer «vekslet» til flytting av objekt over subjekt). Som i eksperiment 2 sammenligner vi L2-resultatene med data fra L1 fra NWD (Lundquist mfl. 2019). Når det gjelder subjekt plassering, ser vi det nå kjente L2-mønsteret som vi har sett tidligere: L2-innlærerne foretrekker å plassere alle subjekt før negasjon. Et fåtall

talere produserer *neg – NP-subjekt* uten å produsere *neg – subjektspronomen* og har dermed målspråklignende subjekt plassering. Dette gjelder fire talere med ulik språkbakgrunn (russisk, finsk, spansk) som oppgir at minst 50 % av deres daglige produksjon er på norsk. Det er ikke noe i resultatene som tyder på at preferansene for subjekt plassering er L1-spesifikke.

Tabell 11. Plassering av argument i forhold til negasjon i eksperiment 3. Relativ fordeling i prosent.

	L1 subj. %		L2 subj. %		L1 obj. %		L2 obj. %		L1 LOS %		L2 LOS %	
	NP	Pron	NP	Pron	NP	Pron	NP	Pron	NP	Pron	NP	Pron
Vekslet subj./obj.	23,0	96,8	73,0	88,9	1,6	88,7	NA	67,3	0,1	0,1	27,8	1,2
Uvekslet subj./obj.	71,9	0,1	18,4	2,5	94,7	5,1	NA	20,9	95,6	96,1	65,2	91,7
Annet	5,1	3,1	8,6	8,6	3,7	6,2	NA	11,8	4,3	3,3	7,0	7,1

Vi ser nærmere på resultatet for objektsveksling i figur 2. Her settes OS med pronomen (vertikalt) opp mot lang OS (horisontalt) for de individuelle deltakerne. I figuren ser vi at språkbakgrunn påvirker objektsveksling i norsk i stor grad. Informanter med førstespråk som skiller mellom klitikon og pronomen, f.eks. italiensk og romansk (romanske språk) eller tsjekkisk/kroatisk (slaviske språk) havner ofte til høyre: Disse informantene produserer ofte ordstillingen (*lett*) *objekt – subjekt* i tillegg til vanlig objektsveksling.⁸ Infor-



Figur 2. Objektsveksling (OS) og lang objektsveksling (LOS) med pronomen i L2 norsk hos talere med ulik språkbakgrunn.

8. I eksperimentet er det både setninger med reflexivt *seg* og med førstepersons pronomenet *meg* i setningsparene som tester lang objektsveksling (se tabell 7/del 6.2).

manter med førstespråk uten en slik forskjell mellom pronomen og klitikon (finsk, russisk, tysk eller engelsk) plasserer derimot sjeldent objekt før subjekt. Kun én deltaker med engelsk som førstespråk produserte lang objektsveksling, men denne taleren var også en flytende taler av tsjekkisk som L2. I eksperiment 2 så vi at nederlandske L2-innlærere av norsk aldri plasserte objektspronomen (*meg/seg*) foran subjektet (0 % LOS) og i gjennomsnitt vekslet omtrent 85 % av alle objektspronomen. Mange av deltakerne her er klart forskjellige fra de nederlandske deltakerne til tross for at gruppene har sammenlignbar skår på ferdighetstesten (35/36 i eksp. 2 vs. 33,9/36 i eksp. 3). Dette kan tyde på at det er ulike L2-strategier som overføres fra L1 når disse informantene begynner å lære seg norsk: nederlandskspråklige ser ut til å ta utgangspunkt i sine *scrambling*-regler, og slaviske eller romanske språkbrukere baserer seg på sine språkspesifikke regler for plassering av klitikon når de plasserer lette objekt før subjekt. Bare en liten minoritet av våre informanter har tilegnet seg at form er en viktig faktor i argumentplasseringen i norsk (øverst til venstre i figur 2) og behersker argumentplasseringen.

8 Diskusjon

I denne studien har vi gjennomført tre eksperiment som testet plassering av subjekt og direkte objekt i L2 norsk og svensk, og har vist at L2-innlærere har mange vanskeligheter med å tilegne seg reglene for argumentplasseringen. Det største problemet for L2-talerne er plassering av post-verbale NP-subjekt. Her ser vi at L2-talerne tydelig skiller seg fra L1-talerne, uavhengig av metode (vurdering eller produksjon), språkbakgrunn og målspråk (svensk eller norsk). Vi finner at L2-talerne konsekvent plasserer både pronominal- og NP-subjekt til venstre for negasjon. Svært få deltakere produserer ordstillingen *neg – NPsubjekt* i det hele tatt, og de få som gjør det, plasserer også som regel pronominal subjekt etter negasjon. L1-talere plasserer derimot NP-subjekt vanligvis etter negasjon, ikke bare når subjektet er kontrastivt eller utgjør ny informasjon, men også når det er ikke-kontrastivt og utgjør gammel/gitt informasjon. Vi har argumentert for at det norsk-svenske mønsteret er typologisk uvanlig, og at det er lite evidens i innputt som kan hjelpe innlæreren med å lære NP-subjekt-plasseringsreglene (dvs. det er få setninger med adverb, hvor subjektet realiseres som NP, og hvor det ikke står setningsinitialt). Gitt at ikke-kontrastive subjekt nesten alltid blir realisert som pronomen, og derfor plasseres foran setningsadverb, er det en god sjanse for at L2-taleren også konkluderer med at gitte/ikke-kontrastive NP-subjekt er obligatorisk plassert der. Vi finner et inter-

essant mønster hos L2-innlærere som begynner å produsere *neg – subjekt-sekvenser* med gitte/ikke-kontrastive nominalfraser. Når L2-innlærere begynner å plassere NP-subjekt etter negasjon, så ser de ut til å bruke dette mønsteret til en viss grad også med pronomen. Dette funnet forsterker én av konklusjonene fra Anderssen mfl. (2018), nemlig at L2-talere ikke legger så stor vekt på form på subjekt: De går generelt ut ifra at pronomen og NP har samme distribusjon og at disse blir påvirket av informasjonsstrukturelle faktorer på samme måte. Dette er også en rimelig antagelse gitt at både NP og pronominal subjekt vanligvis forekommer initialt i setningen eller rett etter det finitte verbet.

Når det gjelder objektplassering finner vi flere mønster i de samlede resultatene som vi beskrev med henvisning til innflytelse fra L2-innlærernes førstespråk. I motsetning til det vi ser med subjektplassering, virker disse mønstrene nemlig mer L1-spesifikke. Anderssen mfl. (2018) fant at L2-talere generelt har en preferanse for objekt etter negasjon, men våre nederlandskspråklige deltakere har en annen preferanse: De plasserer objektspronomen i høy grad før negasjon, og gjør dette også med NP-objekt (eksp. 2). I motsetning til de nederlandske deltakerne, produserer L1-talere av språk med enklitiske objektspronomen (eksp. 3) ofte ordstillingen *objekt – subjekt* i tillegg til vanlig objektsveksling.

Dette fører oss til et svar på det andre forskningsspørsmålet vårt, som omhandler effekten av tversspråklig innflytelse fra L1: Lingvistiske distinksjoner overført fra L1 har en betydelig effekt på både vurdering og produksjon av argument- og særlig objektplassering i norsk og svensk. Det virker som om de nederlandskspråklige informantene tolker nordisk objektsveksling som *scrambling*, og at L1-talere av språk med verbklitikon analyserer objektsveksling som en type verbklitisering på den måten som vi gjorde rede for i bakgrunnsdelen. Et annet argument som støtter innflytelse fra L1 finner vi i lang objektsveksling: Uansett målspråket (norsk eller svensk) er det lite som skiller de to gruppene med nederlandske talere, verken i vurdering (eksp. 1) eller i produksjon (eksp. 2). Selv om norsk og svensk er tydelig forskjellige når det kommer til dette fenomenet, er L2-innlærernes ordstillingsmønster nokså like. Vi kan ikke utelukke at deltakerne våre, og ordstillingsmønstrene vi finner, er påvirket av andre, tidligere tilegnede L2-språk. Alle våre deltakere behersker for eksempel engelsk. Vi har derimot argumentert for at tversspråklig innflytelse fra engelsk vil lede til preferanse for objekt *etter* negasjon, noe vi ikke ser i våre resultat.

Det siste forskningsspørsmålet vårt, angående forskjellen mellom produksjon og grammatikalitetsvurderinger, kan besvares ved å sammenligne eksperiment 1 og 2. Vi ser at konstruksjoner og ordstillinger som får høy skår i grammatikalitetsvurderingen i eksperiment 1, også er ordstillinger som produ-

seres i eksperiment 2. Dette gjelder spesielt subjektplassering. Derimot stemmer ikke vurderingene og produksjonen like godt overens når vi ser på objektsvekslingen: Deltakerne gir lav skår til vekslende NP-objekt i eksperiment 1, men i produksjon plasserer de fleste objekt før negasjon uavhengig av form. Vi har argumentert for dette som en effekt av materiale (bruk av bestemte nominalfraser og egennavn i produksjonseksperimentet).

Tidligere forskning har argumentert for at voksne L2-innlærere legger større vekt på frekvens enn unge L1-innlærere og at de generaliserer ut fra den mest frekvente posisjonen uten å ta hensyn til form (Anderssen mfl. 2018). Med en slik frekvens-basert strategi havner alle subjekt vanligvis foran setningsadverbial, og alle objekt etter setningsadverbial i andrespråket. I eksperiment 1, hvor vi brukte samme testmetode som Anderssen mfl. (2018), skilte våre nederlandske deltakere seg derimot fra deltakerne i studien til Anderssen mfl. ved at de var mer målspråklige i plasseringen av objektspronomen. Både ovenfor og i diskusjonen av våre resultat tidligere, har vi derfor argumentert at det frekvensbaserte mønsteret (*subjekt – neg* og *neg – objekt*) er en noe forenklet beskrivelse av L2-mønsteret som ikke tar hensyn til innflytelsen av argumentplasseringsregler fra førstespråket.

Anderssen mfl. (2018) skilte mellom ulike grupper deltakere basert på deres nivå i norsk, og antydte at deltakerne med høyest ferdighetsnivå hadde enda sterkere preferanser for den atypiske *subjekt – neg*-ordstillingen enn de mindre dyktige deltakerne. I studien vår er det vanskelig å skille mellom deltakere med ulike nivå i målspråket, siden deltakerne har en nokså lignende (og høy) skår på ferdighetstesten. Sannsynligvis skyldes dette at ferdighetstesten vi brukte, er for kort og enkel. Det er imidlertid naturlig å anta at informantene i eksperiment 3 har hatt mye mer norsk innputt, og har dermed høyere nivå i norsk, ettersom de bor i Norge. Det samme gjelder noen av de nederlandske deltakerne i eksperiment 2 som har bodd i Norge eller Sverige i lengre perioder. I motsetning til mønsteret i Anderssen mfl. (2018) finner vi i våre resultat fra eksp. 2 og 3 et fåtall deltakere (disse brukte L2-språket på daglig basis i lang tid) som har et nesten målspråklignende plasseringsmønster av NP og pronominale subjekt.

9 Konklusjon

I denne artikkelen har vi presentert en studie om tilegnelse av subjekt- og objekt-plassering i norsk og svensk hos L2-innlærere med forskjellige førstespråk: én gruppe med nederlandsk som L1 og én mer heterogen gruppe med forskjel-

lige L1-språk. I tillegg var den nederlandske gruppen delt i to basert på om mål-språket til deltakerne var norsk eller svensk. Vi utførte tre eksperimenter som testet argumentplassering i andrespråket med grammatikalitetsvurderinger og i produksjon. Våre samlede resultater viser at L2-innlærere særlig har vansker med å lære riktig plassering av NP-subjekt, og mer generelt med å avlære plasseringsmønstre fra sine førstespråk. Vi finner tydelige forskjeller i L2-ordstillsingspreferanser mellom de nederlandske innlærerne og den heterogene innlærerguppen samt innenfor den siste gruppen. Vi har tolket disse forskjellene som resultat av tverrspråklig innflytelse fra førstespråket.

Referanser

- Anderssen, Merete, Kristine Bentzen, Guro Busterud, Anne Dahl, Björn Lundquist & Marit Westergaard. 2018. The acquisition of word order in L2 Norwegian: The case of subject and object shift. *Nordic Journal of Linguistics* 41-3, 247–274.
- Bentzen, Kristine. 2014. Object Shift. *Nordic Atlas of Language Structures (NALS) Journal* 1, 332–343.
- Bentzen, Kristine & Merete Anderssen. 2019. The form and position of pronominal objects with non-nominal antecedents in Scandinavian and German. *Journal of Comparative Germanic Linguistics* 22, 169–188.
- Bentzen, Kristine, Merete Anderssen & Christian Waldmann. 2013. Object Shift in spoken Mainland Scandinavian: A corpus study of Danish, Norwegian, and Swedish. *Nordic Journal of Linguistics* 36-2, 115–151.
- Bouma, Gerlof & Helen de Hoop. 2008. Unscrambled pronouns in Dutch. *Linguistic Inquiry* 39-4, 669–677.
- Broekhuis, Hans & Marcel den Dikken. 2012. *Syntax of Dutch: Nouns and Noun Phrases Volume 2*. Amsterdam: Amsterdam University Press.
- Clahsen, Harald & Pieter Muysken. 1986. The availability of Universal Grammar to adult and child learners – a study of the acquisition of German word order. *Second Language Research* 2-2, 93–119.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.
- Dryer, Matthew S. 2013. Expression of Pronominal Subjects. I Dryer, Matthew S. & Martin Haspelmath (red.): *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/101>
- Eide, Kristin Melum. 2015. Tilegnelse av verbale kategorier. I Eide, Kristin M. (red.): *Norsk andrespråkssyntaks*. Oslo: Novus, 135–196.

- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Falk, Ylva & Camilla Bardel. 2011. Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research* 27-1, 59–82.
- Håkansson, Gisela. 1997. Barnets väg till svensk syntax. I Söderbergh, Ragnhild (red.): *Från joller till läsning och skrivning*. Malmö: Gleerups, 47–60.
- Holmberg, Anders. 1986. *Word order and syntactic features in the Scandinavian languages and English*. Stockholm: Stockholm University dissertation.
- Holmberg, Anders. 1993. Two subject positions in Mainland Scandinavian. *Working Papers in Scandinavian Syntax* 52, 29–41.
- Josefsson, Gunlög. 2003. Four myths of Object Shift ... and the truth. I Delsing, Lars-Olof, Cecilia Falk, Gunlög Josefsson & Halldór Sigurðsson (red.): *Grammar in Focus, volume II: Festschrift for Christer Platzack, November 18, 2003*. Lund: Department of Scandinavian Languages, Lund University, 199–207.
- Lundquist, Björn, Ida Larsson, Maud Westendorp, Eirik Tengesdal & Anders Nøklestad. 2019. Nordic Word order Database: Motivations, methods, material and infrastructure. *Nordic Atlas of Language Structures (NALS) Journal* 4-1, 1–33. [Database tilgjengelig online via <https://tekstlab.uio.no/nwd/>.]
- Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44-2, 314–324.
- Odlin, Terence. 2013. Crosslinguistic influence in second language acquisition. I Chapelle, Carol A. (red.): *Encyclopedia of Applied Linguistics*. Oxford: Blackwell, 1–6.
- Svenonius, Peter. 2002. Subject positions and the placement of adverbials. I Svenonius, Peter (red.): *Subjects, Expletives and the EPP*. Oxford: Oxford University Press, 201–242.
- Tengesdal, Eirik & Björn Lundquist. 2019. *Partikkelplassering og partikkelak-sent i tre norske dialektområde i laupet av dei siste 80 åra*. Presentasjon på LIA sluttseminar, Trondheim.
- Thráinsson, Höskuldur. 2001. Object shift and scrambling. I Baltin, Mark & Chris Collins (red.): *Contemporary Syntactic Theory*. Malden: Blackwell Publishing, 148–202.
- Unsworth, Sharon. 2005. *Child L2, Adult L2, Child L1: Differences and Similarities. A Study on the Acquisition of Direct Object Scrambling in Dutch*. Ph.d.-avhandling, Utrecht University.

- Westergaard, Marit. 2011. Subject positions and information structure: The effect of frequency on acquisition and change. *Studia Linguistica* 65-3, 299–332.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. I *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 1556–1559.
- Østbø Munch, Christine. 2013. *North Germanic Negation. A Microcomparative Perspective*. Doctoral thesis, University of Tromsø.

Summary

The placement of subjects and objects in Norwegian and Swedish is dependent on many different factors. It has been shown previously with grammaticality judgements that second language learners are not sensitive to fine-grained distinctions, such as information structure and form, which are crucial in Norwegian argument placement (Anderssen et al. 2018). In this article, we present results from three experiments that test the placement of subjects and objects in second language learners of Norwegian and Swedish, both in comprehension and production. Our goal is to test how much the argument placement rules of the first language (L1) affect the placement pattern in the second language (L2). Our results show that L1 plays a major role: L2ers expect argument placement to be governed by the same factors in L2 as in L1. Furthermore, we find, like Anderssen et al. (2018), that placement of NP-subjects is particularly difficult for L2 learners. Finally, we show that patterns from grammaticality judgements and L2 production largely overlap.

Maud Westendorp
Doktorgradsstipendiat lingvistikk
CASTL/Institutt for språk og kultur
UiT Norges Arktiske Universitet
Postboks 6050 Langnes
9037 Tromsø
maud.westendorp@uit.no

Björn Lundquist
Førsteamanuensis lingvistikk
CASTL/Institutt for språk og kultur
UiT Norges Arktiske Universitet
Postboks 6050 Langnes
9037 Tromsø
bjorn.lundquist@uit.no



Development of the form and meaning of definiteness in Russian-speaking learners of Swedish

Anders Agebjörn

Lund University and University of Gothenburg

The study investigates the development of definiteness in two groups of Russian-speaking learners of Swedish, one beginner group and one more advanced group. While Russian does not have articles, Swedish expresses definiteness through a complex noun-phrase (NP) structure. Using an oral elicitation task, the study examines the learners' ability to produce morphemes that encode (in)definiteness, their ability to accurately choose between indefinite and definite forms, and the relationship between these two abilities. Findings include that the complex NP structure emerged gradually while there was no evident development with regard to meaning. Initially, however, learners who used the morphemes more also tended to overuse them, while later in development those who produced many morphemes were also more likely to use them accurately. The paper thus demonstrates that the acquisition of a morphosyntactic form and the association of this form with its meaning are two separate processes to some extent.

Keywords: second language acquisition; articles; noun phrase; morphosyntax; semantics

1 Introduction¹

Articles and definiteness have received much interest in the field of Second Language Acquisition. Numerous studies, typically dealing with second-language (L2) English, have reported that L2 learners with an article-less first language (L1) persistently omit and substitute articles (e.g. Huebner 1985; Jarvis 2002; Ionin 2003; Trenkic 2000). However, disentangling the morphosyntactic and semantic/pragmatic aspects of definiteness in learner data poses a methodological challenge since, as pointed out by Kupisch (2006: 168), “the absence of a form may be due to the absence of the associated function, and vice versa”. The present longitudinal and cross-sectional study investigates the development of definiteness in L2 Swedish by learners who are L1 speakers of Russian. While Russian does not have articles, Swedish expresses definiteness using several morphemes in a complex noun-phrase (NP) structure, as shown in (1).² Hence learners of Swedish must figure out both what indefinite and definite NPs look like (form) and when they are used (meaning). The morphosyntactic complexity of definite NPs opens up the possibility for learners to express definiteness before a target-like structure is acquired, and the study attempts to exploit this fact in order to track the development of the *form* and *meaning* of definiteness separately.

-
1. Acknowledgements: The paper has benefited from comments by Åsa Wengelin, Susan Sayehli, Tanja Kupisch, two anonymous reviewers and the NLT editors. Johan Segerbäck reviewed my English and Anika Agebjörn proofread. All remaining faults are my own. Anna Ransheim and Kahterina Shpakovskaya translated to and from Russian; Camilla Håkansson volunteered as a second transcriber/coder. Native speakers of Swedish were recruited with help from Sebastian Fannon, Melanie Lilja, Matilda Lindgren, Övind Linnerud and Frida Splendido; learners with help from Mariya Sakovets, Nina Shpakouskaya and Nastassia Maiskaya and others at the Centre for Swedish Studies in Minsk, Belarus. Data collection in Minsk was enabled by travel grants from Stipendiefonden Viktor Rydbergs minne (DS2016-0766) and the Swedish Institute (22713/2017). Jonas Pehrandel and Kyskåpspoesi © provided wonderful gifts for the participating learners. Many thanks to all of you, and not least to all the participants.
 2. Throughout the paper, N and A stand for Noun and Adjective, INDEF for the indefinite article, and DEF for any of the three definite morphemes: the nominal suffix, the adjectival suffix and the left-edge article.

- (1) a. en katt
 INDEF cat
 ‘a cat’
- b. en (vit) katt
 INDEF (white)cat
 ‘a (white) cat’
- c. katt-en
 cat-DEF
 ‘the cat’
- d. den vit-a katt-en
 DEF white-DEF cat-DEF
 ‘the white cat’

Both the form and the meaning of definiteness have been found to pose a challenge to learners of L2 Swedish (e.g. Axelsson 1994; Kolaczek 2018; Nyqvist 2013, 2015, 2018). However, previous research is inconclusive with regard to which of the grammatical morphemes in (1) are more difficult to acquire. Further, no study has investigated whether, and if so how, the two tasks involved – developing the form and associating it with its meaning – are related to each other. In other words, nobody has addressed the question whether, say, learners who are sensitive to the meaning of definiteness are more likely to express this meaning. The aim of the paper is to contribute to the general understanding of L2 acquisition of grammatical form and meaning by describing the development of a complex morphosyntactic structure and the association between this structure and its abstract meaning, and by exploring the relationship between these two processes.

2 Background

The present study is descriptive and explorative; patterns observed are discussed primarily in terms of their input frequency and difficulty. Ellis (2002: 143) describes L2 acquisition as “the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them”. According to DeKeyser (2005), a grammatical phenomenon can be difficult because of its form, its meaning and the form–meaning mapping. Difficulty of form depends on “the number of choices involved in picking all the right morphemes and allomorphs [...] and putting them in the right place” (pp. 5–6) while difficulty of meaning depends on novelty and abstractness – DeKeyser mentions

articles as an example of forms that express “highly abstract notions” (p. 5). Difficulty of form–meaning mapping depends on the transparency of this mapping, which can be obscured by redundancy, optionality and opacity, that is, different forms expressing the same meaning and different meanings being expressed by a single form. Input frequency as well as these different types of difficulty might explain why definiteness poses a challenge to learners of Swedish whose L1 lacks articles.

2.1 *Definiteness and the learning task*

The present study distinguishes between definiteness as a *category of meaning* and as a *grammatical category* (cf. Lyons 1999). Semantically, a definite NP signals that the referent is unique within a *pragmatic set* which is shared by, or mutually manifest to, the speaker and the hearer (Hawkins 1991: 409) and so is identifiable to the hearer.³ The pragmatic set can be established anaphorically or deictically: the speaker can refer to *the* book if a unique, identifiable book was mentioned in previous discourse or is present in the real-world context. The pragmatic set can be narrow, as when a speaker refers to *the* kitchen of the present apartment, or wide, as when a speaker refers to *the* moon of the present planet. Further, the referent can be either directly or indirectly identifiable within the set. Indirect identifiability requires the hearer to make inferences based on knowledge about the world. For example, when the speaker and the hearer share a set that includes a book, the speaker can refer to *the* author on the assumption that the hearer knows that a book normally has a unique author.

Some languages, such as Swedish (Teleman et al. 1999), obligatorily express whether an NP has indefinite or definite reference, typically using articles. In these languages, bare nouns are often ungrammatical. As shown in (1a–b) above, Swedish has an indefinite article at the left edge of the NP. This article is used with singular count nouns. It has two allomorphs (*en* and *ett*); the choice between them depends on gender (common/neuter). Further, as shown in (1c), Swedish has a definite “article” suffixed to the noun. This nominal suffix has at least three allomorphs (*-(e)n*, *-(e)t* and *-na*); the choice depends on gender, number (singular/plural) and the noun’s declension class. In addition, in adjectivally pre-modified NPs, definiteness is expressed not only by this suffix but also by an adjectival-agreement marker and a left-edge article. This structure, shown in (1d), is often referred to as *double definiteness* (Julien 2005: 26). The adjectival suffix has two allomorphs, one default (*-a*) and one used optionally

3. Teleman et al.’s (1999: 156) corresponding term is *frame of identification*.

(and rarely) with singular male referents (-*e*). The default allomorph is homonymous with the plural agreement marker used in both indefinite and definite NPs, but in singular-head NPs the -*a* inflection unambiguously encodes definiteness (Julien 2005: 45–47). Finally, the definite left-edge article has three allomorphs (*den*, *det* and *de*), the choice depending on gender and number. They are all homonymous with personal pronouns ('it', 'they') and demonstratives ('that', 'those').

There are several exceptions to the system exemplified in (1) (Julien 2005; Teleman et al. 1999). For example, the nominal suffix is not used in NPs including possessives and some definite quantifiers (e.g. *min katt(*-en)* 'my cat'; *samma katt(*-en)* 'the same cat'); it is omitted in certain expressions (e.g. *har inte den blekaste aning(*-en)* 'haven't the slightest idea') and is optional if the NP includes a restrictive relative clause and the definite left-edge article (e.g. *den katt(-en) som du såg* 'the cat that you saw'). Further, the definite left-edge article is not used in proper names (e.g. *(*det) Vita huset* 'the White House') and is optional with inherently definite modifiers such as ordinal numbers and superlatives (e.g. *(den) första dagen* 'the first day'). In fact, to some extent it is optional also with ordinary adjectives if the referent is known to the hearer and/or is present in the immediate contexts, that is, if the adjective enables identification of the referent.

Like most Slavic languages, Russian does not have articles. Hence bare nouns are grammatical and can be interpreted as either indefinite or definite, depending on a range of linguistic and contextual factors (Sussex & Cubberley 2006). However, importantly, this does not mean that definiteness as a category of meaning is absent in speakers of Russian, who have been found systematically to conceive of NPs as indefinite or definite even though their grammar does not encode this meaning with dedicated morphology (e.g. Brun 2001; Lyons 1999; Slabakova 2008; Trenkic 2000).

To sum up, Russian-speaking learners of Swedish thus have to acquire a morphosyntactic structure characterised by complexity, optionality and redundancy, and associate it with the abstract meaning of definiteness, which is not obligatorily expressed in their L1. This is no easy task. Indeed, as pointed out by DeKeyser (2005: 8), in cases where abstractness of meaning coincides with novelty and redundancy of form and with opacity of form–meaning mapping, "the learning problem is particularly severe".

2.2 Previous research

On a general level, research into articles and definiteness in L2 English has suggested that learners whose L1 does not have articles are prone both to omit articles and to ascribe non-target meanings to them (resulting in article-substitution errors). However, both omission and substitution rates tend to decrease with time and increasing proficiency (e.g. Ionin 2003; Trenkic 2000, 2009). When it comes more specifically to Swedish and Norwegian (which is structurally similar to Swedish), Table 1 provides an overview of relevant studies, indicating for each study what languages were involved, whether those languages have articles and what types of data were analysed.

L1 Norwegian and Swedish children have been reported to use a rather adult-like NP structure early in development, around the age of two years (Anderssen 2007; Bohnacker 1997). According to Bohnacker (1997), this cannot be explained in terms of unanalysed chunks or prosodic templates. Further, the definite nominal suffix is acquired before the indefinite article and the definite left-edge article. Indeed, children acquiring Scandinavian languages are younger when they start using this suffix than their peers acquiring other Germanic languages (German and English), with left-edge-articles, are when they start using definite articles (Kupisch et al. 2009). Interestingly, Anderssen and Bentzen (2013) reported that a Norwegian–English bilingual child used the definite left-edge article more than monolingual children did, suggesting cross-linguistic influence from English. When it comes to the meaning of articles, there are indications that children’s earliest definite forms encode specificity – that is, whether the speaker intends to refer to a specific referent – rather than definiteness (Svartholm 1978; cf. Karmiloff-Smith 1981).

With regard to the form of definiteness in L2 Swedish, learners unsurprisingly tend to make more formal errors when producing more complex NPs (Axelsson 1994; Nyqvist 2013). In fact, Nyqvist (2013) reported the rate of formal errors to *increase* with time in L1 Finnish learners of L2 Swedish, probably owing to their production of increasingly more complex NPs. However, Nyqvist (2018) also argued that the frequency of NP types and the complexity of the form–meaning mapping may be more important factors than the complexity of the NP structure *per se*. When it comes to the indefinite article and the definite nominal suffix, which are used in both non-modified and modified NPs, development seems to be partly determined by the learner’s L1: the suffix is less challenging for learners with highly inflectional L1s, such as Russian and Finnish, than for learners with less inflectional L1s, such as English and Spanish (Axelsson 1994; Latomaa 1992; Nordanger 2017). When it comes to double

definiteness, research is inconclusive with regard to whether the left-edge article is more likely to be omitted than the nominal suffix or vice versa: in Axelsson (1994), L1 Finnish, L1 Polish and L1 Spanish learners of Swedish all omitted the suffix more frequently than the left-edge article, whereas the L1 Finnish learners of Swedish in Lahtinen (1993) showed the opposite pattern. Nordanger (2017) found that L1 English learners of Norwegian were more likely to produce the definite left-edge article while L1 Russian learners preferred the nominal suffix, which again is indicative of cross-linguistic influence: the presence of a definite left-edge article in the L1 appears to trigger the use of the Scandinavian definite left-edge article (cf. Anderssen & Bentzen 2013). Finally, existing research on definiteness in L2 Norwegian and Swedish has little to say about adjectival agreement. Jin (2007), investigating sensitivity to agreement errors in end-state L1 Chinese, L1 English and L1 Italian/Spanish learners of Norwegian, found that the Chinese-speaking learners (lacking articles in their L1) were less sensitive to adjectival definiteness agreement than the others (having articles in their L1). By contrast, Lahtinen's (1993) Finnish-speaking learners of Swedish produced the adjectival suffix more consistently than the definite left-edge article.

With regard to the meaning of definiteness, often operationalised as article choice, it has been found that learners of L2 Norwegian and Swedish who have articles in their L1 generally substitute articles less frequently than learners with an article-less L1 (Eriksson & Wijk-Andersson 1988; Nordanger 2017), which is in line with what numerous studies have shown for L2 English (e.g. Jarvis 2002). Nordanger (2017) reported both English- and Russian-speaking learners of Norwegian to overuse definite forms initially, but the overuse was more persistent in the Russian group. Moreover, in the Russian group, the overuse of definite forms was triggered by specificity, just like in L1 children (see above) and in Ionin's (2003) Russian-speaking learners of English. According to Ionin (2003), L2 learners misinterpret articles as encoding specificity rather than definiteness, causing them to overuse definite articles in indefinite specific contexts (cf. Huebner 1985). Finally, Kołaczek (2018) and Nyqvist (2013), investigating Polish- and Finnish-speaking learners of Swedish, found that NPs with direct anaphoric reference were more frequently marked as definite than NPs with indirect anaphoric reference. These two studies also reported a development towards the target norm over time with regard to article choice.

In Nyqvist (2013), morphosyntactic errors were more frequent than article-choice errors. This may not seem particularly surprising, given the 50% chance of making the right choice between indefinitely and definitely marked NPs, but

Table 1. Previous research into L1 and L2 acquisition of definiteness and NP structure in Norwegian and Swedish [+Art] and [-Art] indicate presence and absence, respectively, of articles in participants' L1s

Study	Languages investigated	Participants	Method
Andersen (2007)	L1 Norwegian	Two L1 acquirers.	Oral production. Longitudinal.
Andersen & Bentzen (2013)	L1 English-Norwegian	One simultaneous bilingual child.	Oral production. Longitudinal.
Axelsson (1994)	L2 Swedish	L1 Finnish [-Art] ($n=20$), L1 Polish [-Art] ($n=20$) and L1 Spanish [+Art] ($n=20$) adult immigrants in Sweden.	Oral interviews. Longitudinal.
Bohnacker (1997)	L1 Swedish	One L1 acquirer.	Oral production. Longitudinal.
Eriksson & Wijk-Andersson (1988)	L2 Swedish	L1 German [+Art] and L1 Polish [-Art] exchange students in Sweden, and L1 Swedish controls ($n=10$ /group).	Written essays. Longitudinal.
Jin (2007)	L2 Norwegian	L1 Chinese [-Art] ($n=6$), L1 English [+Art] ($n=6$) and L1 Spanish/Italian [+Art] ($n=5$) adult immigrants in Norway. L1 controls ($n=14$).	Production and perception. Experimental.
Kolaczek (2018)	L2 Swedish	L1 Polish [-Art] university students ($n=16$) in Poland.	Written essays. Longitudinal.
Kupisch et al. (2009)	L1 English; L1 German; L1 Norwegian; L1 Swedish	L1 acquirers of English [+Art], German [+Art], Norwegian [+Art] and Swedish [-Art].	Oral production. Longitudinal and cross-sectional.
Lahtinen (1993)	L2 Swedish	L1 Finnish [-Art] students ($n=342$) in Finland.	Written essays. Cross-sectional.
Nordanger (2017)	L2 Norwegian	L1 English [+Art] ($n=4$) and L1 Russian [-Art] ($n=7$) immigrants in Norway.	Written story retelling. Longitudinal.
Nyqvist (2013)	L2 Swedish	L1 Finnish [-Art] adolescents ($n=67$) in Finland.	Written essays, oral picture-description task, grammar test. Longitudinal.
Nyqvist (2015)	L2 Swedish	L1 Finnish [-Art] university students in Finland ($n=159$).	Longitudinal. Written essays.
Nyqvist (2018)	L2 Swedish	L1 Finnish [-Art] adolescents in Finland ($n=130$).	Grammaticality judgement.
Svartholm (1978)	L1 Swedish	One L1 acquirer.	Oral production. Longitudinal.

the opposite pattern was actually found in somewhat older L1 Finnish learners of Swedish (Nygqvist 2015). However, previous research has not investigated whether the development of the forms and that of the form–meaning association are related to each other. In sum, whether L2 learners have articles in their L1 affects their acquisition of the form and meaning of definiteness. For L2 Norwegian and Swedish, a further relevant factor may be whether the L1 has a rich inflectional morphology. In addition, the complexity of the NP structure also appears to play a role.

2.3 The present study

The present study investigates the developing use of the four structures exemplified in (1) above in native speakers of Russian studying Swedish. It poses three research questions:

- What does the development look like with respect to the morphosyntactic structure through which (in)definiteness is expressed?
- What does the development of the form–meaning association look like?
- What is the relationship between these two processes?

3 Method

3.1 Participants

L1 Russian learners of L2 Swedish were recruited in Minsk, Belarus. First, to track the early development of definiteness, a group of beginners was followed longitudinally during their first year of Swedish study. They attended Swedish lessons twice a week (3–4 hours a week) in two groups at the Belarusian State University and one group at the Centre for Swedish Studies. For this study, they were tested at the beginning of their first term after having received 14–24 hours of Swedish instruction (data point 1), after four months of study (data point 2) and after another three months of study (data point 3). Second, to investigate the potential long-term development of definiteness, another group of learners who had studied Swedish for at least two years and still used the language regularly was tested on one occasion; this group is referred to as the advanced group, meaning only that they were more advanced than the beginners. Finally, to confirm the validity of the oral-production task described below, a group of native speakers of Swedish, recruited from universities and upper-secondary schools in Sweden, were tested on one occasion.

Three beginners were excluded from the analysis for not participating throughout the study, and two beginners and one advanced learner were excluded because they partly misunderstood the elicitation task in such a way that their data could not contribute to answering the research questions. Summary information about the participants finally included, broken down by group, is given in Table 2: number of participants, their age, age at onset of acquisition and years of exposure.

Table 2. Number of participants and their age, age at onset and time of exposure⁴

Group	<i>n</i>	Age				Age at onset				Years of exposure			
		<i>M</i>	<i>Md.</i>	<i>SD</i>	Range	<i>M</i>	<i>Md.</i>	<i>SD</i>	Range	<i>M</i>	<i>Md.</i>	<i>SD</i>	Range
Beg. 1	21	18.8	17	4.8	17–39	18.8	17	4.8	17–39	0.0	0	0.0	0–0
Adv.	22	25.5	23	5.6	19–37	20.8	19	4.8	17–33	4.6	4	2.6	2–13
L1	26	17.5	17	1.5	16–22	0.0	0	0.0	0–0	17.5	17	1.5	16–22

According to a background questionnaire filled in by all participants, all members of both the beginner group and the advanced group were native speakers of Russian. Nine of the beginners and six of the advanced learners reported Belarusian as an additional L1, and one advanced learner reported Ukrainian as an additional L1. Russian, Belarusian and Ukrainian are similar grammatically; crucially for the present purposes, all three languages lack articles (Sussex & Cumberley 2006). Before learning Swedish, all had learned English (a language with articles) to varying degrees of proficiency; the average age of onset was 8 years for the beginner group (range=5–17) and 9 years for the advanced group (range=4–20). The beginners' self-reported English proficiency, assessed using the Russian version of the Global Scale of the Common European Framework of Reference (Council of Europe 2001), ranged from A2 to C1 with the mean located between B1 and B2, while that of the advanced learners' ranged from B2 to C2 with the mean at C1. The advanced learners' higher proficiency in English may be due to the fact that they were also generally older than the beginners and had consequently used English for a longer period of time. In addition, 12 of the beginners and 15 of the advanced learners had studied other languages with articles – Germanic and Romance languages – to varying degrees of proficiency: generally A1–A2 for the beginners and B1–B2 for the advanced learn-

4. Here and henceforth, *n*, *M*, *Md* and *SD* stand for number, mean, median and standard deviation; Beg. 1, Beg. 2, Beg. 3 stand for the beginner group at data points 1, 2 and 3; Adv. stands for the advanced group; and L1 stand for the reference group.

ers. The present study does not control for the learners' previous knowledge of article languages.

After data collection, the learners completed the grammar and vocabulary sub-tests of two standardised Swedish-proficiency tests: Swedex A2 (maximum score: 10) and Swedex B1 (maximum score: 40) (Folkuniversitetet 2019). As shown in Table 3, the advanced group outperformed the beginners on both tests (A2: $t=5.05$; $p<.001$; B1: $t=5.57$; $p<.001$).

Table 3. L2 Swedish proficiency

Group	Swedex A2				Swedex B1			
	<i>M</i>	<i>Md.</i>	<i>SD</i>	Range	<i>M</i>	<i>Md.</i>	<i>SD</i>	Range
Beg. 3	6.9	7	2.3	2–10	18.5	19	7.7	2–34
Adv.	9.5	10	0.7	8–10	31.5	34	7.6	13–40

3.2 Task performed by participants

An oral-production task, a game loosely inspired by Trenkic (2000) and Jaensch (2009), elicited adjectivally modified and non-modified NPs in indefinite and definite contexts. The researcher met the participants individually and their interaction was audio recorded. Between them was a board depicting a town with a church, a park, a restaurant, etc., as well as wooden building blocks depicting sets of identical people, animals and objects such as chairs and cars. The participant was given a map showing where on the board different blocks should be located and was instructed to explain to the researcher where to put the blocks. The procedure was repeated three times, with new blocks that had to be placed in relation to the ones that had already been placed on the board. In this way, the task forced the participant both to introduce non-identifiable referents and to refer to referents that were identifiable owing to their position on the board. For example, the participant might say, “En pojke står på stolen i parken” (‘A boy is standing on the chair in the park’).

For the beginners, who completed this task on three different occasions, a few wooden blocks were added at data points 2 and 3. The idea was to give them the impression that there was a developmental aspect to the task so that they would abstain from trying to prepare for it. The L1 Swedish participants were randomly assigned one of the three sets of blocks. Because new blocks were added, the new versions elicited a slightly larger number of NPs, but a comparison of the three L1 groups using ANOVAs showed that there were no differences between the three versions with respect to the eight variables de-

scribed in Section 3.4 below. The advanced learners were assigned the first set of blocks.

The vocabulary needed to solve the task consisted of words included in the first chapters of the textbooks that the beginners had already worked through by data point 1 (Levy Scherrer & Lindemalm 2007; Nyborg, Pettersson & Holm 2001) as well as a few Russian–Swedish–English cognates such as *bank* and *park*. The beginners were instructed in English, the advanced learners and the L1 participants in Swedish. The beginners also received an instruction sheet, written in Russian and English, which included a list of words needed to solve the task. In that list, the nouns appeared with the indefinite article, which probably influenced the beginners' production of the indefinite article; this should be kept in mind when interpreting the results.

3.3 Transcription and coding

First, the recordings of the participants solving the task were transcribed and all singular count NPs referring to the wooden blocks were excerpted. Immediate, verbatim repetitions of NPs were not excerpted, as they were deemed to say little about the participants' actual usage. The NPs excerpted were then coded for reference and morphosyntactic structure. With regard to reference, based on the L1 participants' behaviour, NPs referring to non-unique blocks that had not yet been placed on the board were coded as indefinite, while NPs referring to blocks that were already on the board or to blocks that had been mentioned right before were coded as definite. With regard to morphosyntactic structure, NPs were coded in terms of whether or not they included an adjective, the definite adjectival suffix, the indefinite article, the definite nominal suffix and the definite left-edge article. Gender was not considered; for example, both *en* and *ett* counted as an indefinite article regardless of whether the correct allomorph was chosen in a given context. The demonstrative *den här* 'this', which was used a few times, was counted as a definite left-edge article.

Second, an inter-coder reliability test was carried out. Six randomly chosen recordings (two of L1 speakers, two of advanced learners and two of beginners at data point 1) were transcribed and coded by a second Swedish-speaking researcher. For the excerpting of relevant NPs, they agreed in 96.5 % of cases, and for morphosyntactic structure, they agreed in 93.9 % of the cases (the second coder did not code the NPs for reference, as reference was defined by the design of the task). Where they did not agree, this was often due to the difficulty of hearing whether a common-gender noun ending in a vowel, such as *flicka* 'girl', is inflected (*flicka-n* 'the girl'). Afterwards, the first transcriber/coder

went through the recordings a second time and paid special attention to such cases.

Finally, before the analysis, all NPs with self-corrections (1.8 % of the data) were excluded. This was to avoid a discussion about which version of the NP should be included – the spontaneous one or the corrected one. Certain other NPs were also excluded: those containing relative clauses, inherently definite adjectives (see Section 2.1) and English words, as well as those without a noun (in all, another 3.7 % of the data). This was because omission of the morphemes investigated is grammatical and/or expected in such NPs; for example, learners who switched to English as a communicative strategy could of course not be expected to supply Swedish functional morphology.

The data analysed included a total of 5,075 NPs. However, since omission and substitution rates were calculated for each participant (at each data point) separately (as described in the next section), the number of NPs produced *per participant* is more relevant than the total number of included NPs. This information is summarised in Table 4, showing that the average number of NPs in indefinite contexts ranged from 20–27, the average number of NPs in definite contexts ranged from 17–24, and the average number of double-definiteness contexts (i.e., the number of adjectivally modified NPs in definite contexts, which is obviously a subset of all NPs in definite contexts) ranged from 8–10. As can also be seen, the lowest number of NPs produced by one single participant was 10 for indefinite contexts and 12 definite contexts, while the lowest number of adjectivally modified NPs in definite contexts was only 4, which is admittedly a rather low number for calculating a rate.

Table 4. *Obligatory contexts per participant by group/data point*

Group	NPs in indefinite contexts				NPs in definite contexts				Modified NPs in definite contexts			
	<i>M</i>	<i>Md.</i>	<i>SD</i>	Range	<i>M</i>	<i>Md.</i>	<i>SD</i>	Range	<i>M</i>	<i>Md.</i>	<i>SD</i>	Range
Beg. 1	24.3	23	2.8	19–31	18.6	17	3.6	15–30	10.5	10	3.1	4–16
Beg. 2	26.4	26	3.5	16–34	19.3	19	3.7	12–30	9.5	9	3.5	4–18
Beg. 3	26.6	27	2.5	20–31	22.6	22	3.7	17–34	9.4	9	1.7	6–13
Adv.	19.5	20	4.0	10–28	19.9	18	4.6	15–30	8.1	8	1.9	5–12
L1	26.3	26	2.9	21–32	24.3	24	6.0	16–43	10.1	10	2.6	5–17

3.4 *Measuring the form and the meaning of definiteness*

For each participant and data point, eight variables were calculated: five targeting the form of definiteness and three targeting its meaning. The first form variable, *general suppliance*, measures the extent to which a participant marked NPs as indefinite or definite. It was calculated by dividing the number of NPs

marked as indefinite or definite by the total number of NPs. In this context, NPs with the indefinite article and without any definite morphemes were considered indefinitely marked (e.g. *en katt*, *en vit katt*); NPs with at least one of the three definite morphemes and without the indefinite article were considered definitely marked, independently of accuracy (e.g. *den vit-a katt-en*, *?vit-a katt-en*, **vit katt-en*, **den katt*). The value of this variable potentially ranges from 0 (indicating that no NPs were indefinitely or definitely marked) to 1 (indicating that all NPs were so marked).

The other form variables (*indefinite article*, *definite nominal suffix*, *definite adjectival suffix* and *definite left-edge article*) are morpheme-specific suppliance variables which were calculated for each of the four morphemes investigated by dividing the number of instances of the particular morpheme by the number of instances of obligatory contexts for that morpheme. In other words, these variables measure the extent to which the morphemes were produced in relation to how often one would expect them to be produced, which enables comparison across the four morphemes. Note that whether or not the morphemes were actually used in obligatory contexts was not considered here, as these variables did not target the participants' knowledge of the meaning of definiteness. Hence, simply, the number of indefinite articles was divided by the number of NPs in indefinite contexts, the number of definite nominal suffixes was divided by the number of NPs in definite contexts, and the number of adjectival suffixes and the number of definite left-edge articles were divided by the number of adjec-tively modified NPs in definite contexts.

Regarding these four morpheme-specific variables, the highest possible value was set to 1, meaning that participants who produced a larger number of a certain morpheme than there were obligatory contexts for that morpheme did not score more than 1. This was because it would be complicated to handle variables for which the learners could deviate from the native speakers both “upwards” and “downwards”, where downwards could indicate that the learner had not acquired the form and upwards could indicate that the learner had not acquired its meaning. Again, the variables described here were intended to measure only the extent to which the morphemes were used (form), not the extent to which they were used in accurate contexts (meaning); *overuse* of morphemes is instead measured by the meaning variables, described right below. For this reason, the value of the morpheme-specific form variables potentially ranges from 0 (indicating that the morpheme was never used) to 1 (indicating that the morpheme was supplied *at least* as often as one would expect).

As mentioned above, in addition to the five form variables described, three meaning variables were also calculated. The first of them, *general NP choice*, measures the accuracy with which participants chose between indefinite or definite forms according to context. It was calculated by dividing the number of indefinitely or definitely marked NPs (as defined above) produced in pragmatically accurate contexts by the total number of indefinitely or definitely marked NPs. The second variable, *indefinite NP choice*, was calculated by dividing the number of indefinitely marked NPs produced in indefinite contexts by the total number of indefinitely marked NPs; the third variable, *definite NP choice*, was calculated in the corresponding way. Hence the value of the three meaning variables also potentially ranges from 0 (indicating that no indefinitely/definitely marked NPs were used in accurate contexts) to 1 (indicating that all indefinitely/definitely marked NPs were used in accurate contexts).

Finally, it should be mentioned that some beginners did not produce any indefinitely and/or definitely marked NPs at all, such that the meaning variables could not be calculated for them (as nothing can be divided by zero). Specifically, for this reason, the general-NP-choice and indefinite-NP-choice variables were impossible to calculate for 5 beginners (out of 21) at data point 1, while the definite-NP-choice variable was impossible to calculate for 16 beginners at data point 1, for 4 beginners at data point 2, and for 3 beginners at data point 3. In other words, as only 5 of 21 beginners produced any definite morphology at data point 1, the study does not reveal much about the learners' general knowledge of the meaning of definiteness at this early stage in development.

3.5 Statistical analysis

Since the groups were small and the variables were not normally distributed, non-parametric statistics were used. For each variable, tests were carried out with regard to the difference between the three groups (the L1 group and the advanced group were compared with the beginners at data point 3), the development between the three data points for the beginner group, and correlation with time of exposure for the advanced learners, who had studied Swedish for 2–13 years. In addition, comparison with regard to the production of the four morphemes was carried out for each group/data point. Finally, to explore the form–meaning relationship, correlations between the form variables and the meaning variables were tested. All analyses were carried out using the R language (R Core Team 2019; Kim 2015).

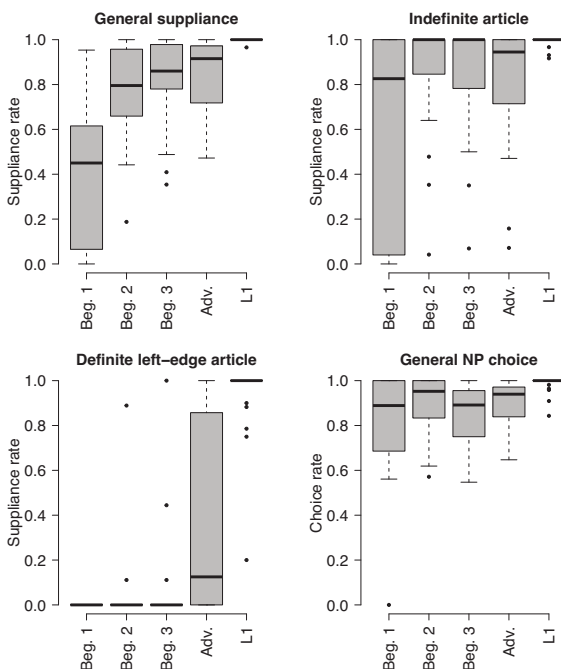
4 Results

Results for the eight variables are visualised in Figure 1. The form variables are investigated in Section 4.1, the meaning variables in Section 4.2 and the correlations between them in Section 4.3.

4.1 Development of form

To give a good idea about what the data actually looked like, Table 5 provides the number of NPs by group/data point and structural pattern. As expected, the L1 speakers primarily used the four structures given in (1) in the Introduction. Sometimes, however, they used the definite left-edge article as a demonstrative pronoun in non-modified NPs (e.g. *den katt-en*), and sometimes they omitted the left-edge article in definite modified NPs (e.g. *?vit-a katt-en*). By contrast, the two groups of learners combined the four morphemes in practically all pos-

Figure 1 (below and right): The eight variables by group/data point. Thick horizontal lines represent medians, boxes represent the middle 50% of participants, whiskers represent the range and dots represent outliers.



sible ways. However, it was rare for indefinite and definite morphemes to co-occur, indicating that the morphemes were not used randomly. Besides the structures in (1), the patterns dominant among the learners were bare NPs (e.g. **katt*, **vit katt*), inflected nouns with uninflected adjectives (e.g. **vit katt-en*) and the double-definiteness structure with the left-edge article omitted (e.g. ?*vit-a katt-en*).

Regarding the five form variables, the L1 controls obtained high values for all of them ($Md.=1$), suggesting that those variables were valid. Kruskal–Wallis tests found significant differences between the three groups for each form variable: general suppliance ($\chi^2(2)=36.0$; $p<.001$), the indefinite article ($\chi^2(2)=12.9$; $p<.01$), the definite nominal suffix ($\chi^2(2)=26.9$; $p<.001$), the definite adjectival suffix ($\chi^2(2)=29.0$; $p<.001$) and the definite left-edge article ($\chi^2(2)=41.9$; $p<.001$). Post-hoc Wilcoxon tests showed that the L1 group outperformed both learner groups for each variable ($Z=2.97$ – 5.98 ; $p<.01$ – $.001$). The advanced learners outperformed the beginners with respect to the definite nominal suffix ($Z=2.58$; $p<.05$) and the definite left-edge article ($Z=3.20$; $p<.01$) but not with respect to general suppliance, the indefinite article or the adjectival suffix. Fried-

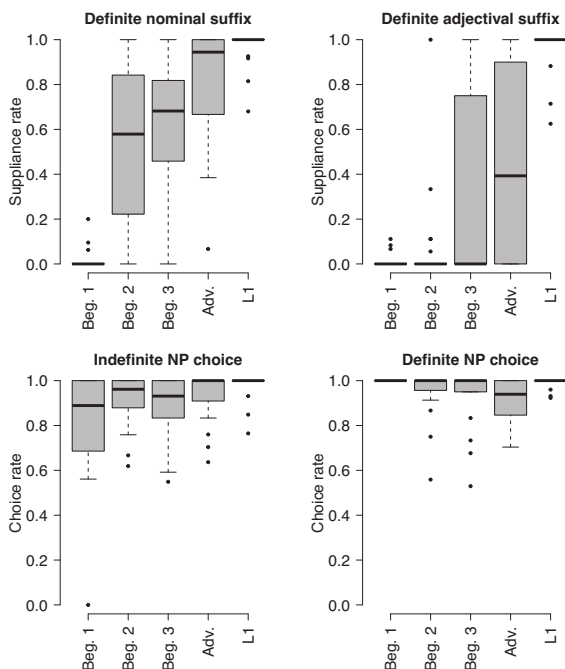


Table 5. Number of NPs by group and structural pattern. The four structures in bold correspond to the examples given in (1) in the Introduction.

Group	INDEF N	INDEF A N	N-DEF	DEF A-DEF N-DEF	DEF N-DEF	* N	*DEF N	* A N	* A N-DEF	?A-DEF N-DEF	* A-DEF N	*DEF A-DEF N	*INDEF A N-DEF	*INDEF A-DEF N	*INDEF N-DEF	*INDEF A-DEF N-DEF
Beg. 1	157	201	3	0	0	220	0	311	5	0	2	0	0	1	0	0
Beg. 2	253	252	131	5	1	76	0	137	92	2	0	3	4	4	0	1
Beg. 3	269	267	188	9	3	84	1	86	46	50	5	3	4	5	10	3
Adv.	230	134	192	56	19	54	1	71	62	29	1	0	7	3	0	0
L1	334	359	354	243	12	2	0	0	0	11	0	0	0	0	0	0

man tests found significant differences in the beginner group for general suppliance ($\chi^2(2)=23.4$; $p<.001$), the indefinite article ($\chi^2(2)=11.2$; $p<.01$), the nominal suffix ($\chi^2(2)=26.4$; $p<.001$) and the adjectival suffix ($\chi^2(2)=8.5$; $p<.05$). Post-hoc Wilcoxon tests showed that general suppliance as well as use of the indefinite article and the nominal suffix increased between data points 1 and 2 ($Z=2.91-3.91$; $p<.01-.001$), while use of the adjectival suffix increased between data points 2 and 3 ($Z=2.29$; $p<.05$). By contrast, use of the definite left-edge article did not increase significantly above zero during the beginners' first year of Swedish study.

The L1 participants produced the four morphemes equally often (relative to the number of obligatory contexts), as shown with Wilcoxon tests. The beginners used the indefinite article more than the definite morphemes at each data point ($Z=2.74-3.65$; $p<.05-.001$). At data point 1, they used the three definite morphemes equally often (i.e., almost never); at data points 2 and 3, they used the definite nominal suffix more frequently than the adjectival suffix and the definite left-edge article ($Z=3.38-3.54$; $p<.05-.001$); and at data point 3, they used the adjectival suffix more frequently than the definite left-edge article ($Z=2.69$; $p<.01$). The advanced learners used the indefinite article and the nominal suffix equally often, and they also used the adjectival suffix and the definite left-edge article equally often. However, the former two morphemes were supplied more consistently than the latter two ($Z=3.10-3.64$; $p<.01-.001$).

In sum, the two morphemes that are used in both modified and non-modified NPs (the indefinite article and the nominal suffix) were produced earlier in development and more frequently (relative to the number of obligatory contexts) than those that are used only in modified NPs (the adjectival suffix and the def-

inite left-edge article). Further, the adjectival suffix was used earlier than the definite left-edge article. This suggests an implicational order of acquisition for definite NPs: the nominal suffix is used earlier in development than the adjectival suffix, which is used before the left-edge article. The fact that the beginners used the indefinite article earlier and more consistently than the definite morphemes was probably because this article was presented to them in the instruction sheet. However, Spearman correlations between the five form variables and time of exposure in the advanced group, shown in Table 6, also suggest that the definite forms may actually develop more slowly than the indefinite article; only the use of definite morphemes correlated positively and significantly with years of exposure. (Recall that the advanced learners did not get the word list where nouns appeared with the indefinite article.)

Table 6. Correlations between the form variables and years of exposure in the advanced group[†]

	General suppliance	Indef. art.	Def. nom. suffix	Def. adj. suffix	Def. left-edge art.
Years of exposure	.26	-.11	.46*	.27	.53*

[†] Asterisks indicate that a correlation is significant at the $p < .05$ level.

4.2 Development of meaning

The L1 Swedish speakers' high values for the three meaning variables ($Md.=1$) suggest that the task was effective at eliciting NPs with unambiguously indefinite or definite reference. As for the learners, they performed well above chance level: their median values ranged from 0.89–1. Kruskal–Wallis tests found significant differences between the three groups for general NP choice ($\chi^2(2)=23.7$; $p < .001$), indefinite NP choice ($\chi^2(2)=13.1$; $p < .01$) and definite NP choice ($\chi^2(2)=13.9$; $p < .001$). Post-hoc Wilcoxon tests showed that the L1 participants outperformed both learner groups with respect to general NP choice and indefinite NP choice ($Z=2.54$ – 4.32 ; $p < .05$ – $.001$), but they outperformed only the advanced group with respect to definite NP choice ($Z=3.85$; $p < .001$). The advanced learners did not outperform the beginners or vice versa on any meaning variable, and there were no significant differences between the three data points in the beginner group. A similar lack of development is also reflected in Table 7, which shows that the Spearman correlations between the three NP choice variables and time of exposure in the advanced group were not significant.

Table 7. Correlations between the meaning variables and time of exposure in the advanced group

	General NP choice	Indefinite NP choice	Definite NP choice
Exposure	.12	.21	-.20

4.3 Relationship between the two developments

Table 8 displays the correlations between the meaning and the form variables. Column 2 shows the correlations between the general-NP-choice and the general-suppliance variables; it answers the question whether those who were generally good at encoding NPs as indefinite or definite were also more (or less) accurate in choosing between indefinite and definite forms according to the context. Column 3 shows the correlations between the indefinite-NP-choice and the indefinite-article variables, answering the question whether those who supplied many indefinite articles were also more (or less) likely to produce them in accurate contexts. Finally, column 4–6 show the correlations between the definite-NP-choice and the three definite-morpheme variables, answering the question whether those who supplied these definite morphemes frequently were also more (or less) likely to use them in accurate contexts.

For the L1 participants, there were no significant correlations. For the beginners at data point 1, when they produced almost exclusively indefinite articles, general compliance correlated negatively with general NP choice. In other words, the more they used the morphemes, the more they overused them. Later, at data point 3, when most beginners had begun to use the definite nominal suffix, also definite NP choice correlated negatively with the definite-nominal-suffix variable. By contrast, advanced learners who produced many indefinitely or definitely marked NPs were also more likely to produce them in accurate contexts, as shown by the positive correlation between general NP choice and general compliance for the advanced group. Hence it would seem that morphemes tend to be overused at an early stage of development.

Table 8. Correlations between form and meaning variables[†]

	General	Indef. article	Def. nominal suffix	Def. adjectival suffix	Def. left-edge article
L1	-.11	-.13	-.15	-.13	.20
Adv.	.45*	-.26	.08	.28	.22
Beg. 3	-.02	-.33	-.52*	-.22	.05
Beg. 2	-.04	-.37	-.42	.12	.23
Beg. 1	-.52*	-.39	—	—	—

[†] Asterisks indicate that a correlation is significant at the $p < .05$ level. For the advanced group, correlations are partial, controlled for years of exposure. For the beginners at data point 1, correlations could not be tested for the nominal suffix as it was always produced in accurate contexts, for the adjectival suffix as it was always produced in inaccurate contexts, or for the definite left-edge article as it was never produced.

5 Concluding discussion

The present study investigated the development of the form and the meaning of definiteness in native speakers of Russian studying Swedish in Minsk, Belarus. While the analysis yielded evidence suggesting a gradual development of the morphosyntactic structure through which (in)definiteness is encoded, there was no evidence of such a development when it comes to associating the form with its meaning. Even so, it was found that, at early stages of development, the ability to produce the relevant morphemes correlated negatively with the ability to choose the right morpheme in context, while at later stages, learners who frequently produced the morphemes were also more likely to use them accurately. This might be indicative of a gradual development with regard to meaning as well.

Regarding the *form* of definiteness, three observations need to be discussed. First, the beginners used the indefinite article more frequently and earlier in development than the definite nominal suffix. This is contrary to findings from L1 research (Anderssen 2007; Bohnacker 1997) and both Scandinavian (e.g. Nordanger 2017; Nyqvist 2013) and international L2 research (e.g. Goad & White 2004; Huebner 1985; Trenkic 2000) showing that indefinite articles are generally acquired later and omitted more frequently than definite ones (but see Leung 2005). As mentioned above, the early and frequent use of the indefinite article in the beginner group was probably due in part to the fact that this article was presented to them in the instruction sheet for their task. However, as the data from the advanced learners also suggested that definite morphemes might develop more slowly than the indefinite article, further research should be carried out to explore other possible explanations. In this context it might be hypothesised, for example, that the general complexity and redundancy of Swedish definite modified NPs might play a role. It might also be of interest

that the Swedish indefinite article bears a resemblance – semantically, structurally and phonologically – to the Russian numeral *odin* ‘one’ (Sussex & Cumberley 2006) and the English indefinite article *a/an*; as pointed out above, all learners in the study had previously learned English to varying degrees of proficiency.

Second, the indefinite article and the definite nominal suffix, which are used in both adjectivally modified and non-modified NPs, were used more frequently (relative to the number of obligatory contexts) and earlier in development than the definite adjectival suffix and the definite left-edge article, which are used only in adjectivally modified NPs. For the beginners, this is readily explained by the fact that, according to their teachers, the double-definiteness structure was not explicitly taught until the second term of Swedish studies, that is, between data points 2 and 3. However, for the advanced learners, who had studied Swedish for at least two years, the lower suppliance rates for the adjectival suffix and the definite left-edge article may instead be explained in terms of input frequency (cf. Ellis 2002): as modified NPs are less frequent than non-modified ones, the modified NP structure was apparently less entrenched in the learners.

Third, the beginners used the definite adjectival suffix earlier in development than the definite left-edge article, despite the fact that these morphemes, according to their teachers, were taught simultaneously as part of the same structure. This suggests that the present study has identified an implicational order of acquisition for definite modified NPs: the nominal suffix emerges before the adjectival suffix, which emerges before the left-edge article. To some extent, the same pattern can be discerned in L1 Swedish children (Bohnacker 1997), as well as in L1 Finnish learners of Swedish (Lahtinen’s 1993). Of all definitely marked, modified NPs produced by Lahtinen’s (1993: 185–186) L1 Finnish learners of Swedish, 21.0 % included the adjectival suffix but not the left-edge article while only 3.8 % included the left-edge article but not the adjectival suffix; this would seem to confirm the above-mentioned implicational order of acquisition. It could be speculated that this is due to input frequency and the nature of the morphosyntactic structure (cf. Ellis 2002; DeKeyser 2005): while the adjectival suffix is a relatively reliable marker of definiteness (at least in singular-head NPs), the left-edge article has several allomorphs, is often optional, and is never used together with other definite determiners such as possessives. In addition, cross-linguistic influence might play a role: Finnish and Russian both have adjectival agreement (although not of definiteness), which may possibly have boosted the development of adjectival agreement in Lahtinen’s L1 Finnish learners and in the L1 Russian learners in the present study.

A final note on the form of definiteness is also warranted. Contrary to the present study, other studies have reported learners of L2 Norwegian and Swedish to sometimes omit the nominal suffix in definite modified NPs, resulting in the “Danish” structure of *den vit-a katt* (DEF white-DEF cat) or the “English” structure of *den vit katt* (DEF white cat) (Axelsson 1994; Jin 2007; Nordanger 2017; cf. Anderssen & Bentzen 2013). As modified definite NPs are infrequent in input compared to non-modified ones, the definite left-edge article is unlikely acquired before the nominal suffix. Hence the reported omission of the nominal suffix in modified definite NPs may not be indicative of a syntactic deficit (even though, as pointed out above, the influence exerted by other languages previously learned by the participants, such as English, may play a role), but might rather be an effect of processing limitations that come into play at later stages of development, when the complexity of the interlanguage increases. This resonates well with Trenkic’s (2009) suggestion that the more material an NP is supposed to include, the more likely it is that some of that material will be left out owing to competition for attentional resources.

Regarding the *meaning* of definiteness, the learners did not start from scratch even though, as expected, they were outperformed by the L1 Swedish participants when it came to choosing between indefinite and definite forms. Indeed, as soon as they began to produce the relevant morphemes, their choice between indefinite and definite form was accurate far more often than chance would predict. Surprisingly, however, there was no obvious improvement over time. It should be acknowledged that the production task elicited NPs with a limited number of reference types: the referents were always singular, countable, concrete objects present in the immediate context; the indefinite NPs always referred to one member of a set of identical possible referents, while the definite NPs always referred to a referent that was directly identifiable because of its position on the board. Other studies investigating a much wider range of indefinite and definite NP types have indeed reported a development towards the target norm with regard to article choice (Kołodziejek 2018, Nyqvist 2013; Trenkic 2000). Further, the observation made in the present study to the effect that the general form variable correlated negatively with the general meaning variable in the beginner group but positively in the advanced group might indicate that there was in fact a development with regard to meaning: eventually, learners who actually produced the forms were more likely to use them in the proper contexts. Nevertheless, the lack of development seen for the form variables is striking.

As pointed out in the Introduction, disentangling the form of definiteness from its meaning in L2 data poses a methodological challenge. Given that the present study investigated the learners' knowledge of meaning through their production of forms, it cannot really tease apart the development of form from the association of this form with its meaning; in particular, this was made eminently clear by the fact that the meaning variables could not be calculated for the participants who never produced the relevant forms. It could be speculated that those learners who were keen to use the morphology did not pay much attention to meaning, while others who were more sensitive to meaning were also more restrictive when it came to producing the morphology. Indeed, studies using more sophisticated techniques, such as reaction-time tests, have found both that L2 learners may be sensitive to the meaning of definiteness despite non-target-like production (Trenkic et al. 2014) and that L2 learners may be insensitive to formal errors despite target-like production (Jin 2007).

Considering that the learners in the present study were studying Swedish as a foreign language, it could be that both the gradual, steady development of the forms and the lack of development with respect to the form–meaning association might be due to the instruction they received. The instruction provided was not investigated, but it has been noted in previous research that the meaning of definiteness tends not to be extensively discussed in L2 textbooks (Ionin 2003: 239–240; Kołaczek 2018: 138–159; Nyqvist 2013: 64–69; Trenkic 2000: 59–65). In particular, Kołaczek (2018) found that the L2 Swedish textbooks used by the learners in the present study (Levy Scherrer & Lindemalm 2007; Nyborg, Pettersson & Holm 2001) concentrated more on the form of definiteness than on its meaning. Moreover, the examples given in the textbooks to illustrate the meaning of definiteness were mainly NPs with direct anaphoric reference, that is, NPs referring back to a referent that has been mentioned previously. This is problematic given that demonstratives, which are found in all languages, can also typically be used in that context (Lyons 1999). In other words, the definite article in an NP with direct anaphoric reference could very well be interpreted as a demonstrative, that is, as encoding deixis rather than definiteness. Instead, to force learners to associate definite forms with the abstract meaning of definiteness – uniqueness within a shared set (Hawkins 1991) – they should be presented with a broader range of NPs with different types of definite reference. One obvious reason why textbooks tend to refrain from providing detailed descriptions of the meaning of definiteness is that this meaning is inherently abstract (cf. DeKeyser 2005). The form might be complex and redundant, but it can undeniably be concretised using diagrams and explicit rules. By contrast,

teachers discussing the meaning of definiteness probably have to resort to vague notions such as “intuition”. Hence there is probably room for improvement when it comes to teaching the meaning of definiteness (and teaching teachers to do so). Indeed, Trenkic (2000: 109) suggested that the notion of a mutually manifestness within a shared set can be helpful for concretising the meaning of definiteness in L2 classrooms.

Needless to say, the present study is definitely not the last word on definiteness in L2 Swedish. Even so, by longitudinally and cross-sectionally describing the development of a limited set of NP structures with a limited set of reference types, it has contributed to the understanding of L2 acquisition of grammatical form and meaning. One obvious conclusion to be drawn is that the development of a morphosyntactic structure and the association of this formal structure with its abstract meaning constitute two separate processes. The morphosyntactic development does not appear to be driven by the learners’ desire to communicate that meaning. This finding should have a bearing on L2 teaching practices as well as on Second Language Acquisition theory.

References

- Anderssen, Merete. 2007. The acquisition of compositional definiteness in Norwegian. *Nordlyd. Tromsø University Working Papers on Language & Linguistics* 34-3, 252–275.
- Anderssen, Merete & Kristine Bentzen. 2013. Cross-linguistic influence outside the syntax–pragmatics interface: A case study of the acquisition of definiteness. *Studia Linguistica* 67-1, 82–100.
- Axelsson, Monica. 1994. *Noun Phrase Development in Swedish as a Second Language. A Study of Adult Learners Acquiring Definiteness and the Semantics and Morphology of Adjectives*. PhD Dissertation. Stockholm University.
- Bohnacker, Ute. 1997. Determiner phrases and the debate on functional categories in early child language. *Language Acquisition* 6-1, 49–90.
- Brun, Dina. 2001. Information structure and the status of NP in Russian. *Theoretical Linguistics* 27-2/3, 109–136.
- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.
- DeKeyser, Robert M. 2005. What makes learning second language grammar difficult? A review of issues. *Language Learning* 55-S1, 1–25.

- Ellis, Nick C. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24-2, 143–188.
- Eriksson, Anette & Elsie Wijk-Andersson. 1988. *Swedish Nouns and Articles in German and Polish Students' Swedish Writing*. Uppsala: Forskning och Utbildning i Modern Svenska.
- Folkuniversitetet. 2019. *Swedex – Swedish Examinations*. <<https://www.folkuniversitetet.se/swedex/>>
- Goad, Heather & Lydia White. 2004. Ultimate attainment of L2 inflection: Effects of L1 prosodic structure. In Susan Foster-Cohen, Michael Sharwood Smith, Antonella Sorace & Mitsuhiro Ota (eds): *EUROSLA Yearbook 4*. Amsterdam: John Benjamins, 119–145.
- Hawkins, John A. 1991. On (in)definite articles: implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27(2), 405–442.
- Huebner, Thorn. 1985. System and variability in interlanguage syntax. *Language Learning* 35-2, 141–163.
- Ionin, Tania. 2003. *Article Semantics in Second Language Acquisition*. PhD Dissertation. Massachusetts Institute of Technology.
- Jaensch, Carol. 2009. Article choice and article omission in the L3 German of native speakers of Japanese with L2 English. In María del Pilar García Mayo & Roger Hawkins (eds.): *Second Language Acquisition of Articles: Empirical Findings and Theoretical Implications*. Amsterdam: John Benjamins, 233–263.
- Jarvis, Scott. 2002. Topic continuity in L2 English article use. *Studies in Second Language Acquisition* 24, 387–418.
- Jin, Fufen. 2007. *Second Language Acquisition and Processing of Norwegian DP Internal Agreement*. PhD Dissertation. Norwegian University of Science and Technology.
- Julien, Marit. 2005. *Nominal Phrases from a Scandinavian Perspective*. Amsterdam: John Benjamins.
- Karmiloff-Smith, Anette. 1981. *A Functional Approach to Child Language: A Study of Determiners and Reference*. Cambridge University Press.
- Kim, Seongho. 2015. ppcor: Partial and Semi-Partial (Part) Correlation. R package version 1.1. <<https://CRAN.R-project.org/package=ppcor>>
- Kołaczek, Natalia. 2018. *Bestämmdhet och indirekta anaforer i svenskan ur främmandespråksperspektiv: En studie av polska studenters svenska*. PhD Dissertation. Adam Mickiewicz University in Poznań.

- Kupisch, Tanja. 2006. The emergence of article forms and functions in the language acquisition of a German–Italian bilingual child. In Conxita Lleó (ed): *Interfaces in Multilingualism*. Amsterdam: John Benjamins, 139–177.
- Kupisch, Tanja, Merete Anderssen, Ute Bohnacker & Neal Snape. 2009. Article acquisition in English, German, Norwegian and Swedish. Ronald P. Leow, Héctor Campos & Donna Lardiere (eds.): *Little Words: Their History, Phonology, Syntax, Semantics, Pragmatics, and Acquisition*. Georgetown University Press, 223–236.
- Lahtinen, Sinikka. 1993. Om nominalfrasens struktur i finska gymnasisters inlärsarsvenska och deras behärskning av typerna 'den gula bilen' och 'en gul bil'. In Anne Golden & Anne Hvenekilde (eds.): *Rapport fra det andre forskersymposiet om Nordens språk som andrespråk i Oslo den 19.–20. Mars 1993*. University of Oslo, 181–190.
- Latomaa, Sirkku. 1992. Om transfers roll vid inläringen av svenskans bestämdhet. In Monica Axelsson & Åke Viberg (eds.): *Nordens språk som andraspråk*. Stockholm University, 304–317.
- Leung, Yan-Kit Ingrid. 2005. L2 vs. L3 initial state: A comparative study of the acquisition of French DPs by Vietnamese monolinguals and Cantonese–English bilinguals. *Bilingualism: Language and Cognition* 8-1, 39–61.
- Levy Scherrer, Paula & Karl Lindemalm. 2007. *Rivstart: svenska som främmande språk. A1 + A2 Textbok*. Stockholm: Natur och kultur.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge University Press.
- Nordanger, Marte. 2017. The Encoding of Definiteness in L2 Norwegian: A study of L1 Effects and Universals in Narratives Written by L1 Russian and L1 English Learners. PhD Dissertation. University of Bergen.
- Nyborg, Roger, Nils-Ove Pettersson & Britta Holm. 2001. *Svenska utifrån. Lärobok i svenska*. Stockholm: Svenska institutet.
- Nyqvist, Eeva-Liisa. 2018. Mastering complex Swedish NPs: A comparison of non-immersion pupils and immersion L1 Finnish pupils. *Journal of the European Second Language Association* 2-1, 14–23.
- Nyqvist, Eeva-Liisa. 2015. Bestämdhetsformer och artikelbruk hos vuxna svenksinlärare. *Nordand. Nordisk tidskrift för andrespråksforskning* 10-2, 75–100.
- Nyqvist, Eeva-Liisa. 2013. *Species och artikelbruk i finskspråkiga grundskoleelevers inlärsarsvenska. En longitudinell undersökning i årskurserna 7–9*. PhD Dissertation. Åbo Akademi University.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

- Slabakova, Roumyana. 2008. *Meaning in the Second Language*. Berlin: Walter de Gruyter.
- Sussex, Roland & Paul Cubberley. 2006. *The Slavic Languages*. Cambridge: Cambridge University Press.
- Svartholm, Kristina. 1978. *Svenskans artikelsystem. En genomgång av artikelbruket i vuxenspråket och en modell för analys av bruket i barnspråket*. PhD Dissertation. Stockholm University.
- Teleman, Ulf, Staffan Hellberg, Erik Andersson & Lisa Christensen. 1999. *Svenska akademiens grammatik. 3 Fraser*. Stockholm: Svenska Akademien.
- Trenkic, Danijela. 2000. *The Acquisition of English Articles by Serbian Speakers*. PhD Dissertation. University of Cambridge.
- Trenkic, Danijela. 2009. Accounting for patterns of article omissions and substitutions in second language production. In María del Pilar García Mayo & Roger Hawkins (eds.): *Second Language Acquisition of Articles: Empirical Findings and Theoretical Implications*. Amsterdam: John Benjamins, 115–143.
- Trenkic, Danijela, Jelena Mirkovic & Gerry Altmann. 2014. Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition* 17-2, 237–257.

Anders Agebjörn
Centre for Languages and Literature, Lund University
Box 201, S-221 00, Lund, Sweden
anders.agebjorn@nordlund.lu.se



Komplexa svenska nominalfraser hos finskspråkiga språkbadselever¹

Eeva-Liisa Nyqvist

Helsingfors universitet

Denna artikel undersöker hur finskspråkiga språkbadselever i årskurserna 6 och 9 (n=77 resp. n=86) bedömer grammatikaliteten av några av svenskans mest komplexa nominalfraser i ett test med nominalfraser med dubbel bestämdhet, nominalfraser med de demonstrativa pronomenen *den här* och *denna* samt nominalfraser med både ett possessivattribut eller ett genitivattribut och ett adjektivattribut. Dessa har i tidigare undersökningar konstaterats vara problematiska för L2-inlärare oberoende av förstaspråket. Frastyperna representerar två olika typer av komplexitet: *formell komplexitet* och *komplexitet i förhållandet mellan form och betydelse*.

De formellt komplexa frastyperna bedöms korrekt oftare än sådana där förhållandet mellan formen och betydelsen är komplext. Även frekvensen påverkar: de högfrekventa frastyperna bedöms oftare korrekt. Svårighetshierarkierna är likadana på grupp- och individnivå, och skillnaderna mellan årskurserna är vanligen statistiskt nonsignifikanta. Det är dock anmärkningsvärt att ett fåtal informanter inte uppfyller kriterierna för 75-procentig korrekthet i någon av de undersökta strukturerna i slutet av årskurserna 6 resp. 9.

Nyckelord: nominalfras, bestämdhet, komplexitet, svenska som andraspråk, språkbad

1 Bakgrund och syfte

Finskspråkiga barn har möjlighet att delta i *tidigt fullständigt språkbad* (i forts. *språkbad*) i svenska, där andraspråket (*L2*) tillägnas vid sidan om de dagliga aktiviteterna i dagvård och senare i grundskolan (Bergroth & Björklund 2013).

1. Turku Institute of Advanced Studies vid Åbo universitet har finansierat denna studie.

Tidigare har endast begränsat forskningsintresse riktats mot utvecklingen av den grammatiska kompetensen inom språkbåd: djupgående undersökningar om den grammatiska kompetensens utveckling i språkbåd och med ett stort antal informanter har inte varit vanligt i Finland (se dock Nyqvist 2018ab, 2020, Nyqvist & Lundkvist 2020). Syftet med denna pseudolongitudinella studie är att analysera hur 12- resp. 15-åriga finskspråkiga språkbådselever i årskurserna 6 resp. 9 (i forts. *SB6* resp. *SB9*²) bedömer grammatikaliteten hos vissa komplexa semantiskt definitiva nominalfrastyper i ett flervalstest.

Studien behandlar de receptiva kunskaperna och de formella aspekterna av bestämdhet. Testet koncentrerar sig på dubbel bestämdhet, possessiv- och genitivattribut med adjektivattribut samt de definitiva pronomenen *den här* och *denna* (se 2.1 nedan) pga. att dessa frastyper har uppvisat problem i uppsatser skrivna av språkbådselever (Nyqvist 2018ab), men beläggen har varit för få för analyser på individnivå. Därför är det motiverat att undersöka dessa frastyper med hjälp av ett eliciterat material som garanterar att informanterna producerar tillräckligt många belägg på dem. Studien utgör en del av ett större forskningsprojekt om språkbådselevs grammatiska kompetens. Materialet i projektet består av såväl spontana skrivuppgifter (på ca 200 ord) som grammatiktest som eliciterar mer sällsynta strukturer (såsom testet i denna studie). Jag syftar nedan på problematiken kring valet mellan obestämd och bestämd form med den särskilda termen *species*, som härstammar från Noreen (1904) och används bl.a. i *Svenska Akademiens grammatik* (i forts. *SAG*). Som norm i undersökningen fungerar *Svenska Akademiens grammatik*, som är den mest omfattande deskriptiva grammatiken om modern svenska.

Av de olika synvinklarna på grammatisk svårighet (Housen & Simoens 2016) är den språkliga (inherenta språkliga drag, se 2.1 nedan) och den individuella (i denna studie åldern/tiden som tillbringats i språkbåd) relevanta för denna studie. *SB6* och *SB9* jämförs med varandra för att undersöka om en längre tid i språkbåd gynnar tillägnet av dessa komplexa frastyper.³ De centrala forskningsfrågorna är:

1. Vilka av de undersökta frastyperna är de lättaste resp. svåraste för språkbådseleverna, dvs. vilka bedömer de oftast resp. mest sällan korrekt på grupp- resp. individnivå? Varför?

2. Det finns inga språkbådsgymnasier i Finland (Bergroth 2015).

3. Det bör även minnas att andelen undervisning på svenska i årskurserna 7–9 är lägre än i årskurs 6 (för andelen undervisning på de olika språken i språkbåd, se Bergroth & Björklund 2013).

2. Hurdana skillnader förekommer det mellan årskurserna 6 och 9? Vad beror de på?

Artikeln uppbyggnad är följande: Sektion 2 handlar om teoribakgrunden och tidigare forskning och sektion 3 om materialet och metoden. I sektion 4 analyseras materialet på grupp- och individnivå. Sektion 5 består av den sammanfattande diskussionen.

2 Teoretisk bakgrund

2.1 Synvinklar till grammatisk svårighet av definitiva svenska nominalfraser

DeKeyser (2005) sammanfattar tre perspektiv på grammatisk svårighet: *betydelsen, formen* och *förhållandet mellan dessa två*. Betydelsen hos ett grammatiskt morfem är svår för en L2-inlärare om den är abstrakt eller främmande för ens L1 eller för övriga tidigare inlärd språk (DeKeyser 2005). Morfologiskt species är relativt sällsynt i världens språk: under 10 % av världens språk har både obestämd och bestämd artikel (Haspelmath 1998). Därför är det rimligt att utgå från att en stor del av L2-inlärare av svenska har ett L1 utan morfologiskt species och att de upplever speciesmorfologins funktion som svårbegriplig.

Eftersom finskan saknar morfologiskt species (se t.ex. Vilkuna 1992), är speciesmorfologin i synnerhet i början av inläringen en notorisk svårighetskälla för finskspråkiga svenskinlärare. Speciesmorfologins semantik är svår att tolka i inputen (jfr DeKeyser 2005) och svårigheterna är närvarande vid varje substantiv som informanten producerar, eftersom varje substantiv bör kodas för species. Trots sin höga frekvens har bestämdhetsmarkeringen i svenskan låg *saliens*, dvs. artiklar och ändelser är svåra att lägga märke till i inputen (jfr N. Ellis 2016, Goldschneider & DeKeyser 2005). Finskspråkiga inlärare kan desutom uppleva speciesmorfologin som redundant och därför utelämnat speciesmarkörerna (jfr Jarvis & Odlin 2000, Odlin 2003), dvs. förstaspråkets påverkan torde göra speciesmorfologin ännu mer insalient för dem. Förstaspråkets hinderande effekt avtar dock när inläringen avancerar (t.ex. Bybee 2008, Jarvis 2002, Trenkic & Pongpaioj 2013). Den fristående bestämda artikeln är dock inte nödvändigtvis helt ny för finskspråkiga svenskinlärare, eftersom pronomet *se* ("den, det") förekommer i talad finska på ett sätt som påminner om be-

stämd artikel (Laury 1996).⁴ Detta kan förklara varför de finskspråkiga svenskinlärarna oftare utelämnar substantivets ändelse än den fristående artikeln i dubbel bestämdhet (se 2.2 nedan).

Enligt DeKeyser (2005) borde morfologins roll för betydelse betonas mer i undervisningen, exempelvis genom att dra uppmärksamhet till den när inputen analyseras. Inlärare löper annars risk att bortse från morfologin, i synnerhet när L2 är mer komplext än L1, vilket ju är fallet med just species om man jämför svenska med finska. Det har ytterligare konstaterats att informanternas problem med grundläggande grammatik även stör förståelsen av inputen, dvs. att problemet ligger i kompetensen och inte bara i performansen (se DeKeyser 2005 för en sammanfattning). Bortfall av speciesmarkörerna i L2-svenskan hos finskspråkiga inlärare torde således inte enbart bero på tidspressen som upplevs i en kommunikationssituation, utan även på brist på sensitivitet för speciesmarkering. Om det redundanta elementet dessutom är semantiskt sett nytt eller abstrakt blir inläringen ännu svårare (DeKeyser 2005). Det grammatiska elementets betydelse har alltså relevans för både de receptiva och de produktiva kunskaperna.

Formell svårighet handlar framför allt om komplexitet, dvs. antalet val inläraren måste göra när det gäller att välja de målspråksenliga morfem/allo-morferna och placera dem på de rätta ställena i ett yttrande. Själva substantivböjningen i svenskan är enklare än i flera andra språk, men enligt Philipsson (2004: 125) är de svenska nominalfraserna ”tämligen komplexa”, även när de inte innehåller t.ex. adjektivattribut. Det svenska systemet ter sig ”asymmetriskt och svårtillgängligt”: den obestämda artikeln är ett fristående grammatiskt ord, medan definit betydelse hos substantiven anges med ändelser (Philipsson 2004: 126). Dessutom kongruerar artiklarna i genus, och bestämdhetsändelsens form beror även på substantivets deklination (SAG 1999, del 2, Pronomen § 58; Substantiv § 43). Om de morfem en inlärare bör hålla reda på dessutom har en svår betydelse (såsom speciesmorfologin ur en finskspråkig svenskinlärarens synvinkel) kan man förvänta sig svårigheter i inläringen.

Synnerligen komplexa är nominalfraser med *dubbel bestämdhet*⁵, vilka består av en framförställd bestämd artikel⁶ samt adjektivattribut och substantiv i

-
4. Det är dock aldrig ett obligatoriskt element i språket och infödda talare reagerar inte på bortfall av det, varför man fortfarande anser att finskan är ett språk utan artiklar (Juvonen 2000).
 5. Termen dubbel bestämdhet förekommer inte i SAG, men den är en vedertagen term inom forskningen i L2-svenska (se t.ex. Philipsson 2004, Tingbjörn 1979, jfr Faarlund m.fl. 2006). Därför använder även jag termen i denna studie.

bestämd form (se ex. 1–3). Som Lahtinen (1993b) uttrycker det, innehåller dubbel bestämdhet flera punkter där man kan välja fel:

1. <i>den</i> ART.SG.UTR.DEF	<i>röda</i> ADJ-SG.DEF	<i>väskan</i> NOUN-SG.UTR.DEF
2. <i>det</i> ART.SG.NEUTR.DEF	<i>bästa</i> ADJ-SG.DEF	<i>stället</i> NOUN-SG.NEUTR.DEF
3. <i>de</i> ART.PL.DEF	<i>smutsiga</i> ADJ-PL.DEF	<i>väggarna</i> NOUN-PL.UTR-DEF

Som exemplen 1–3 visar innehåller dubbel bestämdhet *redundans*, dvs. den definitiva betydelsen uttrycks med flera morfem än vad som är semantiskt nödvändigt (DeKeyser 2005). DeKeyser (2005) klassificerar redundansen som en typ av komplexitet i förhållandet mellan formen och betydelsen, men jag anser att den också kan beröra bara formen: dubbel bestämdhet innehåller redundans, men den förekommer enbart i definit betydelse, dvs. den är semantiskt sett enkel.

Redundans förekommer i svenskan också i nominalfraser med pronomina attribut, dvs. fraser där adjektiviska pronomen står som attribut (SAG 1999, del 2, Pronomen § 13):

4. <i>den här</i> DEM.SG.UTR	<i>bilen</i> NOUN-SG.UTR.DEF
5. <i>det här</i> DEM.SG.NEUTR	<i>huset</i> NOUN-SG.NEUTR.DEF
6. <i>de här</i> DEM.PL	<i>bilarna/husen</i> NOUN-PL.UTR/NEUTR.DEF

Exemplen ovan innehåller två element med definit betydelse – det demonstrativa pronomenet *den/det/de här* och substantivets bestämdhetsändelse – medan t.ex. frasen *denna bil* innehåller enbart ett sådant element, det definitiva attributet *denna* (se ex. 7–9 nedan). Båda fraserna är dock semantiskt definitiva, dvs. de syftar på en definit referent (SAG 1999, del 3, Nominalfraser § 4, § 6). Om en

-
6. Den fristående artikeln kan utebli i vissa kontexter, om adjektivattributet står i superlativ eller ersätts t.ex. med ett ordningstal, t.ex. *minsta barnet, första tåget* (SAG 1999, del 3, Nominalfraser, § 5), och substantivets bestämdhetsändelse uteblir i determinativt bruk, dvs. om nominalfrasen innehåller ett restriktivt efterställt attribut, t.ex. *den intressanta bok som vi läste...* (SAG 1999, del 2, Pronomen, § 59). Dessa drag gör dubbel bestämdhet ännu mer förbryllande för L2-inlärare, men behandlas inte mer ingående här pga. att testet i denna undersökning inte handlar om den fristående artikeln eller determinativt bruk.

fras med ett semantiskt definit pronomen också innehåller ett adjektivattribut ökar frasens komplexitet och graden av redundans ytterligare, eftersom adjektivet bör stå i bestämd form på ett likadant sätt som i dubbel bestämdhet (t.ex. *den här nya bilen*, SAG 1999, del 2, Adjektiv § 62).

Den tredje svårighetstypen, komplext förhållande mellan formen och betydelsen, är oberoende av de två förstnämnda. Både formen och betydelsen kan vara enkla i sig, men ett ogenomskinligt förhållande mellan dem leder till problem i inläringen t.o.m. i frekventa strukturer (DeKeyser 2005). Exempelvis de svenska possessiva pronomenen (t.ex. *min*, *din*) och genitivattributen (t.ex. *Annas*, *flickans*) – två underkategorier av definitiva attribut – är enkla ord med konkret betydelse, som dessutom har direkta motsvarigheter i finskan. Dessutom bildas nominalfraser som innehåller sådana (se ex. 7–9) med hjälp av en enkel, undantagslös regel [ett possessivt pronomen/ett genitivattribut+ett eventuellt adjektivattribut i bestämd form+substantiv i obestämd form utan artikel].⁷ Problemet ligger dock i att dessa attribut i svenskan alltid gör referenten unikt identifierbar, dvs. frasens betydelse är definit (SAG 1999, del 3, Nominalfraser § 10). Därför är korrelationen mellan substantivformen och betydelsen låg. Intressant nog ökar ett eventuellt adjektivattribut inte frasens komplexitetsgrad när komplexiteten ligger i förhållandet mellan substantivformen och betydelsen: adjektiven står alltid i bestämd form i semantiskt definitiva nominalfraser (t.ex. *min/Annas/denna nya bil*; SAG 1999, del 2, Adjektiv § 62); förhållandet mellan adjektivattribut och betydelse är entydigt:

- | | |
|---|---|
| 7. <i>min/Annas/denna</i>
POSS ATTR.SG.UTR/
GEN ATTR/DEM.SG.UTR | <i>bil</i>
NOUN.SG.UTR.INDEF |
| 8. <i>mitt/ Annas/detta</i>
POSS ATTR.SG.NEUTR/
GEN ATTR/DEM.SG.NEUTR | <i>hus</i>
NOUN.SG.NEUTR.INDEF |
| 9. <i>mina/ Annas/dessa</i>
POSS ATTR.PL/
GEN ATTR/DEM.PL | <i>bilar/hus</i>
NOUN.PL.UTR/NEUTR.INDEF |

Finskspråkiga L2-inlärare har i tidigare undersökningar visat en tydlig tendens att utgå från att definit betydelse alltid uttrycks med bestämd form (Nyqvist 2013, 2015, 2018ab). I Nyqvist (2013) blev denna normavvikelse i samband med possessiva pronomen och pronomenet *samma* t.o.m. signifikant vanligare

7. Samma regel tillämpas på vissa definitiva pronomina attribut, såsom *denna* och *samma* (SAG 1999, del 2, Pronomen § 62, § 198).

med tiden. Det är sannolikt att informanterna under undersökningstiden blev allt mer medvetna om den definitiva betydelsen hos nominalfrasen och drog den (felaktiga) slutsatsen att definitiv betydelse alltid förutsätter en bestämd form. I det följande kallar jag genitivattribut och de olika typerna av pronomina attribut med den sammanfattande termen *PRG-attribut*.

Fraser med ett PRG-attribut som konstrueras med ett obestämt substantiv (se ex. 7–9) är formellt enkla eftersom de saknar redundans. En formellt sett mer komplex nominalfras med redundans (se ex. 4–6 ovan) kan dock te sig som mindre komplex vad gäller förhållandet mellan form och betydelse: det definitiva attributets betydelse och substantivets form går hand i hand. En annan typ av komplext förhållande innebär att olika former har samma betydelse (DeKeyser 2005), vilket är fallet med pronomenen *denna/detta/dessa* och *den/det här* – som dessutom konstrueras med sinsemellan olika former av substantiv, dvs. obestämd resp. bestämd form.

2.2 Komplexa nominalfraser i tidigare forskning i nordiska språk som L2

Nominalfraserna i L2-svenskan har undersökts mycket. Både dubbel bestämdhet och PRG-attribut nämns i flera undersökningar men dessa frastyper har pga. sin låga frekvens vanligen inte stått i fokus.

Gemensamt för tidigare forskning kring dubbel bestämdhet är att strukturen vanligen är så lågfrekvent att analyserna enbart sker på gruppnivå. Den är svår för inlärare oberoende av ålder, L1, färdighetsnivån och inlärningsmetoden (Axelsson 1994, Lahtinen 1993a, Sundman 1995, Nyqvist 2018abc). Analyserna har vanligen behandlat enbart singularis, troligen pga. att pluralis, i denna konstruktion, är ännu mer lågfrekvent. Hos finskspråkiga inlärare förekommer många slags normavvikelser, framför allt ändelsebortfall i substantivet (typen *den röda *bok*, för *den röda boken*, Axelsson 1994, Lahtinen 1993a, Sundman 1995, Nyqvist 2018b), men även artikelbortfall (typen **röda boken*, Nyqvist 2018a). Likartade normavvikelser förekommer även hos L2-inlärare av norska (som alltså har en identisk struktur, Faarlund m.fl. 2006; Jin 2007, Nordanger 2017). Den komplexa strukturen bereder problem ännu efter flera års inläring (Axelsson 1994, Lahtinen 1993a, Sundman 1995, Nyqvist 2018abc), men Axelsson (1994) har även funnit evidens för att explicit undervisning gynnar tilläggnandet av dubbel bestämdhet: hennes finskspråkiga informanter som fått svenskundervisning redan innan de flyttade till Sverige presterade på en högre nivå än informanterna som inte lärt sig svenska i hemlandet.

Vad gäller PRG-attributen har normavvikelser i både adjektivattributets och substantivets form rapporterats av Sundman (1995; se även Nyqvist 2018abc,

2013, 2015), framför allt i samband med semantiskt definitiva attribut som konstrueras med obestämda substantiv (t.ex. *samma bok*). De attribut som konstrueras med substantivets bestämda form (t. ex. *den här boken*) behärskas vanligen på en högre nivå (Nyqvist 2018abc), dvs. ett komplext förhållande mellan form och betydelse är ett större problem än formell komplexitet. Substantivets form är den vanligaste normavvikelsen även i fraser som också innehåller ett adjektivattribut, men dessa är så få i materialet att det inte är möjligt att dra slutsatser som kunde tillämpas i en större inlärarygrupp (Nyqvist 2018abc).

3 Material och metod

Undersökningen bygger på ett binärt skriftligt flervalstest som jag själv har konstruerat (se Bilaga 1). Det består av sammanlagt 28 semantiskt definitiva komplexa nominalfraser. I varje fras finns ett val mellan två former (t.ex. *det röda äpplet/det röda äpple*), och informanternas uppgift är att välja den målspråkenliga formen av substantivet, eftersom problematiken kring substantivformen är gemensam för dubbel bestämdhet och fraser med PRG-attribut: de sistnämnda innehåller vanligen inte några fristående artiklar. Dessutom har tidigare forskning visat att informanterna oftare har problem med substantivets form än med t.ex. adjektivets form (Axelsson 1994, Lahtinen 1993a, Nyqvist 2018ab). Testet bygger alltså på resultaten från tidigare forskning. Tabell 1 sammanfattar de frastyper som förekommer i testet:

Tabell 1: De undersökta frastyperna i testet

Frastyyp	Antalet fraser	av vilka		
		utrum sg	neutrum sg	utrum pl
Dubbel bestämdhet	12	4	4	4
Genitivattribut + adjektiv + substantiv	4	2	1	1
Possessivattribut + adjektiv + substantiv	4	3	1	-
<i>Den/det här</i> + substantiv	4	3	1	-
<i>Denna/detta/dessa</i> + substantiv	4	2	1	1
Totalt	28	14	8	6

Tolv av fraserna behandlar dubbel bestämdhet (se ex. 1–3 i 2.1.), som representerar formell komplexitet (jfr DeKeyser 2005). Utrum singularis, neutrum singularis och utrum pluralis har alla fyra belägg. Utrala och neutrala substantiv behandlas i singularis separat eftersom neutrala substantiv inte är lika vanliga i svenskan som de utrala (SAG 1999, Del 2, Substantiv § 40). Därför är det

möjligt att de är svårare för informanterna (jfr Lahtinen 1998). Singularis och pluralis undersöks separat pga. att tidigare forskning har visat att bestämd form pluralis är krävande även utan adjektivattribut (Sundman 1995, se även Nyqvist 2018abc, 2015, 2013). I pluralis förekommer enbart utrala substantiv, då testet inte kunde göras alltför långt.

Sammanlagt 16 fraser innehåller ett urval semantiskt definitiva PRG-attribut. Både fraser med possessivt pronomen eller genitivattribut som attribut + adjektivattribut (t.ex. *din bästa vännen/din bästa vän*; *Floridas långa stränderna/Floridas långa stränder*) har fyra belägg i testet. Både possessiva pronomen och genitivattribut kallas i SAG possessiva attribut (SAG 1999 del 1), men jag behandlar dem separat i det aktuella testet pga. att tidigare forskning har visat skillnader i korrektheten (Nyqvist 2013, 2015, 2018ab). Dessa fraser innehåller även adjektivattribut, eftersom sådana fraser är få i fri produktion. (Nyqvist 2018ab, 2013.) Dessa PRG-attribut representerar komplext förhållande mellan formen och betydelsen (jfr DeKeyser 2005).

Resten av fraserna med PRG-attribut innehåller demonstrativa pronomenen *den/det här* och *denna/detta/dessa*. Båda pronomenen har fyra belägg, av vilka de flesta står i singularis. Pronomenen är synonyma men konstrueras i standardsvenskan med sinsemellan olika substantivformer: *den/det här* med substantivets bestämda form, medan *denna/detta/dessa* med den obestämda formen (t.ex. *den här gången/*den här gång*, **denna gången/denna gång*), dvs. även dessa representerar ett komplext förhållande mellan formen och betydelsen (jfr DeKeyser 2005). Dessutom fungerar fraserna med *den/det här* som exempel på formell komplexitet och redundans.

Fraserna i testet innebär inte val mellan indefinit och definit betydelse: i dubbel bestämdhet väljer informanterna mellan en målspråksenlig och en formellt sett bristfällig bestämd nominalfras. I fraserna med PRG-attribut framgår speciesbetydelsen av attributet, som också bestämmer substantivets form, och därför aktualiseras inget egentligt speciesval. I det följande kommer jag att enbart koncentrera mig på nominalfrasernas form och bortse från de semantisk-pragmatiska aspekterna.

Ett grammatiktest är naturligtvis inte ett lika autentiskt material som uppsatser, men ett test är ett fungerande sätt att få tillräckligt med belägg på fras typer som har visat sig vara lågfrekventa i spontant språkbruk, och kan trots allt anses reflektera inlärares kunskapsnivå vid en viss tidpunkt (R. Ellis & Barkhuizen 2005). Ett test möjliggör dock för informanterna en maximal koncentration på grammatiken, varför resultaten inte är direkt jämförbara med resultaten från spontana material. I det aktuella testet bildar informanterna själva

inte ordformerna, utan väljer mellan två givna alternativ, dvs. testet mäter receptiva kunskaper, och informanterna har i varje fras haft en 50-procentig chans att gissa rätt. Dessutom innehåller nominalfraserna i testet bara högfrekventa substantiv och täcker bara ett litet urval av olika pronomen. Dessutom ger flervalstest ingen information om hur informanten ser på den struktur hen *inte* väljer. Det är nämligen möjligt att informanten ser båda alternativen som acceptabla men detta framkommer inte p.g.a. att hen är tvungen att välja ett alternativ. Man bör inte heller dra slutsatser om inlärares praktiska språkkunskaper utgående från ett flervalstest. Det vore trots allt en överdrift att utgå från att inlärare enbart utnyttjar sina explicita kunskaper när de skriver ett test: informanterna anlitar även sitt ”språköra” i grammatiktest (jfr Nyqvist 2013).

Det pseudolongitudinella materialet har insamlats under våren 2016 från finskspråkiga språkbadselever i årskurserna 6 och 9 ($n = 77$ resp. 86) i Österbotten, Åboland och Nyland, dvs. i de tre landskap där svenskt språkbad erbjuds i Finland. Under testet fick informanterna hjälp med ordförrådet men inte med grammatiken. Informanterna har inlett sin svenskinläring i dagvården, vanligen vid 4–5 års ålder⁸, dvs. sjätteklassarna har lärt sig svenska i 8–9 år och niondeklassarna i 11–12 år. Mängden undervisning de har fått på svenska har varierat i de olika årskurserna.⁹

Informanternas svar analyseras på grupp- och individnivå. Eftersom ett av syftena med analysen är att redogöra för vilka nominalfrastyper som är de svåraste för informanterna beräknas andelarna för korrekta bedömningar (i forts. *korrekthetsprocenten*) för de olika frastyperna (t.ex. dubbel bestämdhet i singularis/pluralis) på gruppnivå genom att dividera antalet informanter som valt den målspråksenliga formen med det totala antalet informanter. Svårighetshierarkierna bildas sedan utgående från fallande korrekthetshierarkier enligt Browns (1973) princip, som utgår från att en hög korrekthetsprocent innebär tidig inläring och lätt struktur. Statistikprogrammet SPSS 25 används för att utreda vilka skillnader mellan frastyper och årskurser som är statistiskt signifikanta. Medelvärdena för korrekthetsprocenten för de olika frastyperna jämförs med hjälp av *t*-testet (gränsvärdet $p < 0,05$), som i princip förutsätter normalfördelning, förutom när samplet är tillräckligt stort (Kim & Kwak 2017).

Materialet analyseras på individnivå med hjälp av implikationell skalning, som i sin traditionella form visar antalet informanter som behärskar de under-

8. För genomsnittliga startåldrar i olika delar av Finland, se Bergroth (2007).

9. 85 % i årskurserna 1–2, 60 % i årskurserna 3–4, 50 % i årskurserna 5–6 och 45 % i årskurserna 7–9 (Bergroth & Björklund 2013).

sökta strukturerna och om behärskning av t.ex. dubbel bestämdhet i pluralis förutsätter behärskning av något annat, såsom dubbel bestämdhet i singularis. Om detta är fallet, kan man dra slutsatsen att singularis behärskas före pluralis och en inlärningsgång (dvs. en svårighetshierarki) kan etableras.

Behärskningen kan operationaliseras med hjälp av både systematiskt bruk (t.ex. tre korrekta belägg kan tolkas som tecken på att en viss form har etablerat sig i inlärarespråket; se t.ex. Rahkonen & Håkansson 2008) och procentuella kriterier (t.ex. Pallotti 2007). Då det aktuella materialet är ett grammatikalitetsbedömningsstest som inte handlar om behärskning i ordets vanliga betydelse, används implikationsskalor framför allt för att bilda svårighetshierarkier som jämförs med dem på grupp nivå.

Eftersom systematiskt bruk vanligen förknippas med spontant språkbruk utgår denna studie från ett procentuellt kriterium, som i flera tidigare undersökningar har varit 80 % eller 50 % (t.ex. Philipsson 2007). Jag använder dock 75-procentigt bruk som kriterium av två skäl: för det första möjliggör mitt test inte användning av 80-procentigt bruk. För det andra anser jag att 50-procentigt bruk vore ett alltför lågt kriterium med tanke på hur erfarna svenskinlärare språkbadsleverna är. Om en informant alltså väljer den målspråksenliga formen i tre av de fyra beläggen på en given typ av nominalfras uppfyller hen kriteriet för *75-procentigt (korrekt) bedömning*, som i denna studie ses som en slags ekvivalent för behärskning. Även antalen informanter som bedömer de undersökta frastyperna 75-procentigt korrekt jämförs med hjälp ett statistiskt test. I detta fall utnyttjas Pearsons χ^2 (gränsvärdet $p < 0,05$), eftersom det inte handlar om att jämföra medelvärdena och eftersom det inte kräver normalfördelat material.

4 Analys

I denna sektionen presenterar och analyserar jag informanternas testsvar först på grupp nivå i 4.1. Mer detaljerad information om den individuella variationen i form av en analys på individ nivå ges i 4.2.

4.1 Grupp nivå

Tabell 2 visar korrekthetsprocenten för de olika frastyperna enligt svårighetshierarkin (fallande korrekthetsprocent) i SB6. Skillnaden mellan utrala och neutrala substantiv vid dubbel bestämdhet i singularis är nonsignifikant i båda årskurserna ($p=0,418$ resp. $p=0,070$), varför de i fortsättningen behandlas som en sammanhållen kategori. Neutrala substantiv i svenskan är sällsyntare än de

utrala (SAG 1999, Del 2, Substantiv § 40), och de har tidigare konstaterats vara svårare än de utrala (Lahtinen 1998). Därför finns det skäl att förmoda att resultatet vore annorlunda i ett spontant material.

Tabell 2: Medelvärden för korrekthetsprocenten för de olika frastyperna på gruppnivå

Frastyp	Antalet fraser	Korrekthets-procent	
		SB6	SB9
Dubbel bestämdhet, singularis	8	85 %	89 %
Genitivattribut + adjektiv + substantiv	4	78 %	81 %
<i>Den/det här</i> + substantiv	4	71 %	85 %
Possessivattribut + adjektiv + substantiv	4	68 %	74 %
Dubbel bestämdhet, pluralis	4	62 %	65 %
<i>Denna/detta/dessa</i> + substantiv	4	46 %	51 %

Korrekthetsprocenten i det aktuella materialet avviker tydligt från tidigare forskning: korrekthetsprocenten för dubbel bestämdhet i singularis är betydligt högre än i Axelsson (50 %, 1994) och i språkbadselevernas uppsatser¹⁰ (71 % i årskurs 6, 60 % i årskurs 9, Nyqvist 2018ab). Axelssons material är dock spontant och muntligt, medan Nyqvists (2018ab) material handlar om fri skriftlig produktion. Med andra ord härstammar de från situationer där informanterna främst har fokuserat på kommunikation, medan det aktuella materialet består av grammatikalitetsbedömning.

Dubbel bestämdhet i singularis bedöms av SB6 korrekt i en signifikant större utsträckning än de övriga frastyperna förutom i genitivattribut ($p < 0,001$ i alla), medan *denna/detta/dessa* är signifikant svårare än alla de lättare typerna med undantag av dubbel bestämdhet i pluralis ($p < 0,001$ i alla). Även genitivattributen är signifikant lättare än de tre svåraste frastyperna ($p < 0,001$ i alla).¹¹ Att genitivattribut är lättare än possessivattribut är något överraskande, eftersom situationen har varit det motsatta i uppsatserna (Nyqvist 2018ab, 2013). Genitivattributen kan tänkas vara svårare än possessivattributen eftersom de bildas produktivt, medan de possessiva attributen utgör en handfull korta, högfrekventa ord som antagligen har tillägnats i ett tidigt skede. Skillnaden mellan dessa två attributstyper torde ha neutraliserats i detta test där informanterna inte har behövt bilda attributen själva. Att dubbel bestämdhet i singularis är signi-

10. Det bör noteras att informanterna i Nyqvist 2018ab och denna studie inte är desamma.

11. Bonferroni-korrekturen har tillämpats i dessa beräkningar, gränsvärdet för $p = 0.05/14$ (där 14 är antalet gjorda t -test).

fikant lättare än i pluralis var att vänta, då frastypen är formellt komplex även i singularis, men de plurala fraserna är ännu mer komplexa pga. att substantivet då innehåller två ändelser (jfr DeKeyser 2005). Dessutom har bestämd form pluralis även utan adjektivattribut visat sig vara en lågfrekvent och problematisk form, där bortfall av bestämdhetsändelsen är vanligt (jfr Nyqvist 2018ab, 2013, Sundman 1995).

Svårighetshierarkin är så gott som densamma i SB9 med undantag av *den/det här* + substantiv, där korrekthetsprocenten stiger signifikant från SB6 till SB9 ($p < 0,001$) så att frastypen är den näst lättaste för SB9. I övrigt är skillnaderna mellan årskurserna små och nonsignifikanta. Dubbel bestämdhet i singularis är även i SB9 signifikant lättare än resten av frastyperna med undantag av *den/det här* och genitivattribut ($p < 0,001$ i alla), och *denna/detta/dessa* är återigen signifikant svårare än alla andra frastyper förutom dubbel bestämdhet i pluralis ($p < 0,001$ i alla).¹²

Dubbel bestämdhet och *den/det här* är alltså formellt komplexa, medan komplexiteten hos *denna/detta/dessa* ligger i förhållandet mellan formen och betydelsen (jfr DeKeyser 2005), dvs. de formellt komplexa frastyperna verkar bereda färre svårigheter för informanterna. PRG-attributet *denna/detta/dessa* är dessutom typiskt för skriftspråket, medan *den/det här* förekommer i både tal och skrift och därför torde vara mer frekvent i inputen och klassrumsspråket. Förhållandet mellan form och betydelse är komplext även hos possessiva attribut och genitivattribut, men de är mer frekventa i språket, vilket torde ha gynnat tillägnet av dem.

4.2 Individnivå

I detta avsnitt analyserar jag mitt material på individnivå i två delar. Först analyseras dubbel bestämdhet (utrum och neutrum i singularis samt utrum i pluralis; se förkortningarna *utr_sg*, *neutr_sg* och *pl* i tabellerna 3–4) och sedan de olika PRG-attributen (possessiv- respektive genitivattribut+adjektivattribut+substantiv samt *den/det här* och *denna/detta/dessa*+substantiv; se förkortningarna *poss+adj*, *gen+adj*, *den här* och *denna* i tabellerna 5–6) med hjälp av kriteriet 75-procentig (korrekt) bedömning. Inlärningsgångarna i SB6 och SB9 jämförs kvalitativt med dem från inlärare utan språkbadsbakgrund som testats med samma test (i forts. *kontrollgruppen*, Nyqvist 2018c). Kontrollgruppens implikationstabeller kan läsas i Bilaga 2.

12. Bonferroni-korrekturen har tillämpats i dessa beräkningar, gränsvärdet för $p = 0,05/14$ (där 14 är antalet gjorda *t*-test).

De undersökta formerna har i tabellerna 3–6 ordnats vågrätt från den lättaste (dvs. den med flest informanter som väljer den 75-procentigt) till den svåraste. Informanterna har ordnats lodrätt utgående från hur många av de undersökta strukturerna de bedömer 75-procentigt korrekt. På grund av det stora antalet informanter representerar raderna i tabellerna inte enskilda individer utan grupper: alla informanter som använder samma strukturer 75-procentigt står på samma rad.¹³

Plustecknen (+) anger att informanterna uppfyller kriteriet för 75-procentig bedömning, medan minustecknet (-) visar att det inte har uppfyllts. Avvikelserna, dvs. former där 75-procentig bedömning har nåtts trots att detta inte skett i en lättare form, anges med utropstecken efter plustecknet (+!).¹⁴ Skalornas validitet har testats genom att kalkylera reproduktionskoefficient- och skalabilitetsvärdena (förkortningarna C_{rep} och C_{scal} i tabellerna 3–6; skalorna är valida när C_{rep} överstiger 0,9 och C_{scal} överstiger 0,6; Hatch & Lazaraton 1991). Tabell 3 sammanfattar 75-procentig bedömning av dubbel bestämdhet i SB6:

Tabell 3: 75-procentig bedömning av dubbel bestämdhet i SB6

grupp	utr_sg	neutr_sg	pl	totalt	%	avvikelser
1	+	+	+	36	47 %	0
2	+	+	-	24	31 %	0
3	+	-	+!	3	4 %	3
4	+	-	-	3	4 %	0
5	-	-	+!	2	3 %	2
6	-	+!	-	4	5 %	4
7	-	-	-	5	6 %	0
	66	64	41	77	100 %	9
	86 %	83 %	53 %			

$C_{rep} = 0,97$, $C_{scal} = 0,96$

Knappt hälften av sjätteklassarna (36 informanter) bedömer alla frastyper 75-procentigt, men de flesta informanter bedömer dubbel bestämdhet 75-procentigt endast i singularis. 75-procentig bedömning är så gott som lika vanligt med utrala och neutrala substantiv. Däremot är 75-procentig bedömning signifikant sällsyntare i pluralis än i både utrum och neutrum singularis ($p < 0,001$ i båda). Fem informanter bedömer ingen av frastyperna 75-procentigt korrekt.

13. Exempelvis grupp 1 i Tabell 3 (SB6) omfattar 36 informanter och grupp 1 i tabell 4 (SB9) 38 informanter (se spalten "Totalt") som bedömt alla tre former 75-procentigt.

14. Exempelvis i grupp 3 i Tabell 3 har tre informanter nått 75-procentig korrekthet i pluralformen trots att de inte har gjort det i neutrum singularis.

Skalan är valid (se värden för C_{rep} och C_{scal} ovan), dvs. följande svårighets-hierarki kan fastställas trots avvikelserna: utrum i singularis > neutrum i singularis > utrum i pluralis. Hierarkin är alltså densamma som i SB6 på grupp-nivå och i kontrollgruppen (Nyqvist 2018c). Tabell 4 sammanfattar 75-procentig bedömning av dubbel bestämdhet i SB9:

Tabell 4: 75-procentig bedömning av dubbel bestämdhet i SB9

Grupp	utr_sg	neutr_sg	pl	totalt	%	avvikelser
1	+	+	+	38	44 %	0
2	+	+	-	27	31 %	0
3	+	-	+!	4	5 %	4
4	+	-	-	7	8 %	0
5	-	-	+!	2	2 %	2
6	-	+!	+!	3	3 %	6
7	-	+!	-	3	3 %	3
8	-	-	-	2	2 %	0
	76	71	47	86	100 %	15
	88 %	83 %	55 %			

$C_{rep} = 0,94$, $C_{scal} = 0,91$

SB9 uppvisar samma tendenser som SB6: knappt hälften (44 %) av informanterna bedömer alla former 75-procentigt. Majoriteten av informanterna bedömer dock dubbel bestämdhet 75-procentigt i singularis. Även denna skala är valid, och svårighetshierarkin är likadan som på grupp-nivå och i SB6 och i kontrollgruppen (Nyqvist 2018c). 75-procentig bedömning är återigen signifikant sällsyntare i pluralis än i både utrum och neutrum singularis ($p < 0,001$ i båda). Två informanter bedömer ingen av formerna 75-procentigt korrekt.

Andelen informanter som bedömer dubbel bestämdhet 75-procentigt i SB9 är nonsignifikant högre än i SB6. Informanterna bedömer singularis redan i SB6 på en så hög nivå att 75-procentigt bruk inte kan stiga signifikant i SB9 (s.k. takeffekt). Andelen är dock relativt låg i pluralis ännu i SB9 (55 %), dvs. dubbel bestämdhet i pluralis bereder svårigheter ännu i slutet av språkbadet. Svårighetshierarkin på individnivå är alltså densamma i SB6, SB9 och kontrollgruppen. Hierarkin förefaller logisk, eftersom bestämd form pluralis med sina två ändelser är mer komplex än bestämd form singularis. Att dubbel bestämdhet är svårare i pluralis än i singularis har dokumenterats i uppsatser av språkbadsleverna även i årskurs 9 (Nyqvist 2018a). I tidigare forskning har bestämd form pluralis dessutom oftare berett problem för inlärare än bestämd form singularis även i fraser *utan* adjektiv. (jfr Nyqvist 2013, 2015, 2018ab;

Sundman 1995.) Tabell 5 sammanfattar 75-procentig bedömning av de olika typerna av PRG-attributen i SB6:

Tabell 5: 75-procentig bedömning av PRG-attributen i SB6

Grupp	gen+adj	den här	poss+adj	denna	totalt	%	avvikelser
1	+	+	+	+	7	9 %	0
2	+	+	+	-	26	34 %	0
3	+	+	-	+!	3	4 %	3
4	+	+	-	-	10	13 %	0
5	+	-	+!	+!	2	3 %	4
6	+	-	-	+!	4	5 %	4
7	+	-	+!	-	5	6 %	5
8	+	-	-	-	3	4 %	0
9	-	+!	+!	+!	1	1 %	3
10	-	-	+!	+!	2	3 %	4
11	-	+!	-	+!	1	1 %	2
12	-	+!	+!	-	1	1 %	2
13	-	+!	-	-	8	10 %	1
14	-	-	-	-	4	5 %	0
	60	57	44	20	77	100 %	28
	78 %	74 %	57 %	26 %			

$C_{rep} = 0,91$, $C_{scal} = 0,88$

Bara sju informanter i SB6 bedömer alla fyra frastyper 75-procentigt korrekt: de flesta informanterna uppfyller inte kriteriet för 75-procentigt bruk i *denna/detta/dessa*. Detta är föga överraskande med tanke på att korrekthetsprocenten för attributet är låg även på gruppnivå (46 % i SB6) och motsvarande PRG-attribut har haft låga korrekthetsprocent även i uppsatserna (Nyqvist 2018ab). Fyra informanter bedömer inte en enda av de undersökta PRG-attributen 75-procentigt.

Skalan är valid trots de 28 avvikelserna, varför svårighetshierarkin genitivattribut > *den/det här* > possessivattribut > *denna/detta/dessa* kan fastställas och hierarkin är densamma som i SB6 på gruppnivå. *Denna/detta/dessa* bedöms 75-procentigt signifikant mer sällan än de övriga PRG-attributen ($p < 0,001$ i alla). Även possessivattribut bedöms 75-procentigt av signifikant färre informanter än genitivattribut ($p = 0,006$).¹⁵

När PRG-attributen jämförs med dubbel bestämdhet i SB6, kan man konstatera att signifikant flera bedömer dubbel bestämdhet i singularis (oberoende av genus) 75-procentigt än possessivattribut och *denna/detta/dessa* ($p < 0,001$ i

15. Bonferroni-korrekturen har tillämpats i dessa beräkningar, gränsvärdet för $p = 0,05/6$ (där 6 är antalet gjorda test).

båda). Genitivattributen bedöms däremot 75-procentigt av signifikant flera än dubbel bestämdhet i pluralis ($p=0,001$).¹⁶ Med andra ord bereder strukturerna där *förhållandet mellan formen och betydelsen* är komplext mer problem än de formellt komplexa strukturerna. Det enda undantaget av detta är dubbel bestämdhet i pluralis, som antagligen är lågfrekvent i inputen (jfr Goldschneider & DeKeyser 2005). Tabell 6 sammanfattar 75-procentig bedömning av de undersökta PRG-attributen i SB9:

Tabell 6: 75-procentig bedömning av PRG-attributen i SB9

Grupp	den här	gen+adj	poss+adj	denna	totalt	%	avvikelser
1	+	+	+	+	17	20 %	0
2	+	+	+	-	32	37 %	0
3	+	+	-	+!	7	8 %	7
4	+	+	-	-	6	7 %	0
5	+	-	+!	+!	1	1 %	2
6	+	-	-	+!	1	1 %	1
7	+	-	+!	-	4	5 %	4
8	+	-	-	-	5	6 %	0
9	-	+!	+!	+!	2	2 %	6
10	-	-	+!	+!	1	1 %	2
11	-	+!	-	+!	2	2 %	4
12	-	-	+!	-	3	3 %	3
13	-	+!	-	-	3	3 %	3
14	-	-	-	-	2	2 %	0
	73	69	60	32	86	100 %	32
	85 %	80 %	70 %	37 %			

$C_{rep} = 0,91$, $C_{scd} = 0,88$

Var femte informant i SB9 bedömer alla de undersökta PRG-attributen 75-procentigt, vilket är nästan signifikant ($p=0,055$) flera än i SB6. Även denna skala är valid, varför svårighetshierarkin *den/det här* > genitivattribut > possessivattribut > *denna/detta/dessa* kan fastställas. Hierarkin är alltså densamma som på gruppnivå i SB9 och i kontrollgruppen (Nyqvist 2018c) och avviker från SB6 i och med att flera informanter uppfyller kriteriet i *den/det här* än i genitivattributet. Gemensamt för båda årskurserna är att *denna/detta/dessa* bedöms 75-procentigt av signifikant färre informanter än de övriga PRG-attributen ($p<0,001$).¹⁷

16. Bonferroni-korrekturen har tillämpats i dessa beräkningar, gränsvärdet för $p = 0.05/6$ (där 6 är antalet gjorda test).

17. Bonferroni-korrekturen har tillämpats i dessa beräkningar, gränsvärdet för $p = 0.05/6$ (där 6 är antalet gjorda test).

Andelen informanter som använder de olika PRG-attributen 75-procentigt är nonsignifikant högre i SB9 än i SB6. 75-procentig bedömning av dubbel bestämdhet i singularis är även i SB9 vanligare än 75-procentig bedömning av något av de undersökta PRG-attributen. Skillnaden är dock signifikant enbart i jämförelse mellan dubbel bestämdhet i singularis och *denna/detta/dessa* ($p < 0,001$ i både utrum och i neutrum). I dubbel bestämdhet i pluralis är 75-procentig bedömning däremot signifikant sällsyntare än 75-procentig bedömning av *den/det här* och genitivattributen ($p < 0,001$ i båda).¹⁸

5 Sammanfattning och diskussion

Syftet med denna artikel har varit att jämföra hur finskspråkiga språkbads elever i årskurserna 6 och 9 bedömer grammatikaliteten hos fraser med formellt komplex dubbel bestämdhet och ett urval PRG-attribut med definit betydelse som konstrueras med ett substantiv i obestämmd form (dvs. där förhållandet mellan formen och betydelsen är komplext, jfr DeKeyser 2005). Gemensamt för alla dessa strukturer är att de innehåller element med låg saliens och många av dem är lågfrekventa i inputen, vilket är typiskt för svåra strukturer (jfr Goldschneider & DeKeyser 2005). Frågor som denna studie strävat efter att besvara är 1) vilka frastyper som är de lättaste resp. svåraste för informanterna på grupp- resp. individnivå och 2) hurdana skillnader som förekommer mellan de två årskurserna.

Korrektetsprocenten på grupp-nivå är (med undantag av fraserna med *denna/detta/dessa*) högre i det aktuella materialet än i tidigare undersökningar (t.ex. Axelsson 1994, Nyqvist 2013, 2015, 2018ab), vilket åtminstone delvis beror på att det aktuella materialet består av ett test medan tidigare forskning har utnyttjat framför allt spontant material. Trots olika eliciteringsmetoder är det dock möjligt att se vissa kvalitativa tendenser som upprepas i de olika undersökningarna.

Vad gäller korrektheten i det aktuella materialet förekommer det liknande tendenser på grupp- och individnivå. Den formellt komplexa dubbla bestämdheten i singularis är i båda årskurserna signifikant lättare än de flesta övriga frastyper på grupp-nivå, medan *denna/detta/dessa* (där förhållandet mellan formen och betydelsen är komplext, jfr DeKeyser 2005) är signifikant svårare än de flesta övriga frastyper. Liknande resultat upprepas även på individnivå, där

18. Bonferroni-korrekturen har tillämpats i dessa beräkningar, gränsvärdet för $p = 0.05/8$ (där 8 är antalet gjorda test).

majoriteten av informanterna bedömer dubbel bestämdhet i singularis 75-procentigt, medan bara få gör detta i fraser med *denna/detta/dessa*.

Typiskt för materialet är att det förekommer signifikanta skillnader i korrektheten mellan de olika undersökta frastyperna. Däremot förekommer det bara sällan signifikanta skillnader mellan de två årskurserna: det enda undantaget är fraserna med *den/det här*, där niondeklassare på gruppnivå når en signifikant högre korrekthetsprocent än sjätteklassare; på individnivå är *den/det här* det lättaste PRG-attributet i SB9 (det näst lättaste i SB6). Detta kan i flera fall förklaras med takeffekten, men också med att vissa frastyper är så komplexa att de fortfarande bereder problem för informanterna i slutet av språkbadet. Då skillnaderna i korrektheten mellan årskurserna på både grupp- och individnivå sällan är signifikanta medan skillnaderna mellan de olika frastyperna i de flesta fallen är det, verkar de språkliga faktorerna spela en mer central roll för svårigheten än de individuella (jfr Housen & Simoens 2016).

När fraser med dubbel bestämdhet respektive PRG-attribut jämförs är svårighetshierarkin likadan i båda årskurserna. Val av korrekt form är både på grupp- och individnivå vanligt i dubbel bestämdhet i singularis, *den/det här* samt i possessiv- och genitivattributen. Betydligt färre väljer den rätta formen vid dubbel bestämdhet i pluralis och i *denna/detta/dessa*. Samma hierarki har dokumenterats i kontrollgruppen (Nyqvist 2018c). Bland de lättaste och de svåraste frastyperna finns alltså både formellt komplexa fraser och fraser med ett komplext förhållande mellan form och betydelse. Att dubbel bestämdhet är svårare i pluralis kan förklaras med en högre komplexitetsgrad, då pluralformen innehåller två olika ändelser medan singularis bara har en. Fenomenet har dokumenterats även i tidigare forskning (Nyqvist 2013, 2015, 2018ab). Dessutom har bestämd form pluralis visat sig vara sällsynt och problematisk även *utan* adjektivattribut för t.o.m. avancerade inlärare (Sundman 1995). Frekvensens centrala roll syns även i fraserna med PRG-attribut: att possessiv- och genitivattribut i båda årskurserna är lättare än *denna/detta/dessa* kan förklaras med att de förstnämnda är vanligare i inputen (jfr. Goldschneider & DeKeyser, se även Nyqvist 2018c). Således är det oklart om DeKeyser (2005) syn på frekvensens roll i strukturer som manifesterar ett komplext förhållande mellan formen och betydelsen (se 2.1 ovan) stämmer i det aktuella materialet.

Gemensamt för båda årskurserna och för både grupp- och individnivå är alltså att de formellt komplexa fraserna (med undantag av dubbel bestämdhet i pluralis) oftare bedöms korrekt än frastyperna där komplexiteten ligger i förhållandet mellan formen och betydelsen. Eftersom finskan saknar morfologiskt species (Vilkuna 1992) är det möjligt att finskspråkiga svenskinlärare i början

av inläringen starkt förknippar bestämd form med definit betydelse, vilket som bekant inte är hela sanningen. Därför tar det naturligtvis tid och energi att tillägna sig de kontexter där denna koppling inte stämmer och att ersätta den med mer målspråksenliga representationer (jfr (Bybee 2008, R. Ellis 2008).

Den explicita undervisningens inverkan på andraspråksinläringen har debatterats länge, men nuförtiden är forskare eniga om att explicit undervisning hjälper en att lägga märke till element i inputen (Ranta 2008, se även DeKeyser 2005), dvs. den höjer deras saliens, vilket är en viktig förutsättning för inläringen av grammatiska element. Ett annat problem med de undersökta strukturerna är deras låga frekvenser i inputen: strukturer med en högre frekvens skulle skapa starkare minnesrepresentationer (Goldschneider & DeKeyser 2005, DeKeyser 2005). Detta kan dock åtgärdas genom att förse inlärare med input där de strukturer som ska tillägnas ofta upprepas: det faktum att inlärare stöter på flera belägg på en struktur förstärker inläringen (*skewed input*, Goldberg & Casenhiser 2008).

Antalen informanter i denna studie och i Nyqvist (2018c) är begränsade, och ett flervalstest återspeglar inte informanternas praktiska språkkunskaper, vilket trots allt är språkundervisningens huvudsakliga ändamål. Det vore likväl en överdrift att utgå från att inlärare enbart utnyttjar sina explicita kunskaper när de skriver ett test: informanterna anlitar även sitt ”språköra” i ett flervalstest (jfr Nyqvist 2013). Det skulle dock vara värdefullt att hitta ett sätt att undersöka dessa nominalfraser i ett spontant material som skulle elicitera tillräckligt med belägg för att utreda hur dessa frastyper ter sig i praktiskt språkbruk och avslöja om dessa resultat kan generaliseras till en större population.

Språkbadet betonar praktisk kommunikation framför grammatisk korrekthet, men det betyder inte att metoden helt kan bortse från grammatiken. En enkät med infödda talare har visat att normavvikelserna speciellt i samband med PRG-attributen upplevs som rätt störande pga. att de ger ett slarvigt intryck, även om de inte äventyrar begripligheten (Nyqvist 2014). Därför är det viktigt att vidareutveckla språkbadspedagogiken och hjälpa språkbads eleverna mot ett ännu mer målspråksenligt språkbruk. Tidigare forskning har målat en relativt dystert bild av framför allt dubbel bestämdhet: enligt Axelsson (1994) är det möjligt att svenskinlärare i formell undervisning kanske aldrig lär sig strukturen. Språkbadet erbjuder dock en möjlighet till riklig input och meningsfull kommunikation, och tillsammans med effektiv explicit undervisning är det realistiskt att förvänta sig att eleverna når en hög kompetensnivå även i de mest komplexa nominalfraserna.

Litteratur

- Axelsson, Monica. 1994. *Noun Phrase Development in Swedish as a Second Language. A Study of Adult Learners Acquiring Definiteness and the Semantics and Morphology of Adjectives*. Stockholm: Centrum för tvåspråkighet, Stockholms universitet.
- Bergroth, Mari. 2007. *Kielikylpyperheet valokeilassa. Taustat ja odotukset*. Vaasan yliopiston julkaisuja. Selvityksiä ja raportteja 145. <http://urn.fi/URN:ISBN:978-952-476-204-5>
- Bergroth, Mari. 2015. *Kotimaisten kielten kielikylpy*. Vaasan yliopiston julkaisuja. Selvityksiä ja raportteja 202. Vasa: Vasa universitet. <http://urn.fi/URN:ISBN:978-952-476-617-3>
- Bergroth, Mari & Siv Björklund. 2013. Kielikylpyohjelman tutkimustuloksia Suomessa. I Tainio, Liisa & Heidi Harju-Luukkainen (red.): *Kaksikielinen koulu – tulevaisuuden monikielinen Suomi/Tvåspråkig skola – ett flerspråkigt Finland i framtiden*. Suomen kasvatustieteellinen seura. Kasvatusalan tutkimuksia 62, 91–114.
- Brown, Roger. 1973. *A First Language: The Early Stages*. Cambridge: MA: Harvard University Press.
- Bybee, Joan. 2008. Usage-based Grammar. I Robinson, Peter & N.C. Ellis (red.): *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, 216–236.
- DeKeyser, Robert. 2005. What Makes Learning Second Language Grammar Difficult? A Review of Issues. *Language Learning* 55–S1, 1–25. <https://doi.org/10.1111/j.0023-8333.2005.00294.x>
- Ellis, Nick C. 2016. Salience, Cognition, Language Complexity, and Complex Adaptive Systems. *Studies in Second Language Acquisition* 38-2, 341–351. <https://doi.org/10.1017/S027226311600005X>
- Ellis, Rod. 2008. *The study of second language acquisition*. Andra upplagan. Oxford: Oxford University Press.
- Ellis, Rod & Gary Barkhuizen. 2005. *Analysing Learner Language*. Oxford: Oxford University Press.
- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 2006. *Norsk referansegrammatikk*. Fjärde upplagan. Oslo: Universitetsforlaget.
- Goldberg, Adele & Devin Casenhiser. 2008: Construction learning and second language acquisition. I Robinson, Peter & N.C. Ellis (red.): *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, 197–215.

- Goldschneider, Jennifer M. & Robert DeKeyser. 2005. Explaining the "Natural" Order of L2 Morpheme Acquisition in English: A Meta-analysis of Multiple Determinants. *Language Learning* 51–S1, 1–50.
<https://doi.org/10.1111/j.0023-8333.2005.00295.x>
- Haspelmath, Martin. 1998. How Young is Standard Average European? *Language Sciences* 20–3, 271–287.
[https://doi.org/10.1016/S0388-0001\(98\)00004-7](https://doi.org/10.1016/S0388-0001(98)00004-7)
- Hatch, Evelyn Marcussen & Anne Lazaraton. 1991. *The Research Manual: Design and Statistics for Applied Linguistics*. Cambridge: Newbury House.
- Housen, Alex & Hannelore Simoens. 2016. Introduction: Cognitive Perspectives on Difficulty and Complexity in L2 Acquisition. *Studies in Second Language Acquisition* 38–S2, 163–175.
<https://doi.org/10.1017/S0272263116000176>
- Jarvis, Scott & Terence Odlin. 2000. Morphological type, spatial reference and language transfer. *Studies in Second Language Acquisition* 22–4, 535–556.
<https://doi.org/10.1017/S0272263100004034>
- Jarvis, Scott. 2002. Topic continuity in L2 English article use. *Studies in Second Language Acquisition*, 24–3, 387–418. DOI:
<https://doi.org/10.1017/S0272263102003029>
- Jin, Fufen. 2007. *Second Language Acquisition and Processing of Norwegian DP Internal Agreement*. Doctoral thesis, NTNU.
<https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/243687>
- Juvonen, Päivi. 2000. *Grammaticalizing the definite article: a study of adnominal determiners in a genre of spoken Finnish*. Stockholm: Stockholms universitet.
- Kim, Jong Hae & Sang Kuy Kwak. 2017. Central limit theorem: the cornerstone of modern statistics. *Korean Journal of Anesthesiology* 70–2, 144–156.
<https://doi.org/10.4097/kjae.2017.70.2.144>
- Lahtinen, Sinikka. 1993a. Om nominalfrasens struktur i finska gymnasisters inläraarsvenska och deras behärskning av typerna "den gula bilen" och "en gul bil". I Golden, Anne & Anne Hvenekilde (red.): *Rapport fra det andre forskersymposiet om Nordens språk som andraspråk*. Oslo: Oslo universitet, avdelning for norsk som andrespråk. Institutt for lingvistiske fag, 181–190.
- Lahtinen, Sinikka. 1993b. Om nominalfrasens struktur och feltyperna "en (gul) bilen" och "det gula bilen" i finska gymnasisters inläraarsvenska. I Muittari, Veikko & Matti Rahkonen (red.): *Svenskan i Finland 2*. Jyväskylä: Jyväskylä universitet, 85–98.

- Lahtinen, Sinikka. 1998. *Genuskongruens och genus i finska gymnasisters inlärarsvenska*. Jyväskylä: Jyväskylä universitet.
<http://urn.fi/URN:ISBN:978-951-39-8435-9>
- Laury, Ritva. 1996. *Sen kategoriasta – onko suomessa jo artikkeli? Virittäjä* 100–2, 162–180.
- Nordanger, Marte. 2017. *The encoding of definiteness in L2 Norwegian. A study of L1 effects and universals in narratives written by L1 Russian and L1 English learners*. Doktoravhandling, Universitetet i Bergen.
<http://bora.uib.no/handle/1956/15565>
- Noreen, Adolf. 1904. *Vårt språk. Nysvensk grammatik i utförlig framställning* 5. Lund: Gleerups.
- Nyqvist, Eeva-Liisa. 2013. *Species och artikelbruk i finskspråkiga grundskoleelevers inlärarsvenska. En longitudinell undersökning i årskurserna 7–9*. En doktorsavhandling publicerad i elektronisk form, Åbo universitet.
- Nyqvist, Eeva-Liisa. 2014. Hurdana species- och artikelfel bör motarbetas? I Lindström, Jan; Sofie Henricson, Anne Huhtala, Pirjo Kukkonen, Hanna Lehti-Eklund, & Camilla Lindholm (red.): *Svenskans Beskrivning* 33. Nordica Helsingiensia 37, 353–362.
- Nyqvist, Eeva-Liisa. 2015. Species och artikelbruk hos vuxna svenskinlärare. *NordAnd* 2/2015, 75–100.
- Nyqvist, Eeva-Liisa. 2018a. Definiteness and use of articles in written Swedish by Finnish-speaking immersion pupils at the end of immersion: a comparison with non-immersion students. *Journal of Immersion and Content-based Language Education* 6–1, 57–84.
- Nyqvist, Eeva-Liisa. 2018b. Species och artikelbruk hos finska språkbadslever i årkurs 6. *Folkmålsstudier* 56, 73–104.
- Nyqvist, Eeva-Liisa. 2018c. Mastering complex Swedish NPs: A comparison of non-immersion pupils and immersion L1 Finnish pupils. *Journal of European Second Language Association*, 14–23.
- Nyqvist, Eeva-Liisa. 2020. Interrogativa bisatser hos 12- och 15-åriga språkbadslever. En jämförelse med traditionell undervisning. I Wide, Camilla; Ingman, Eva; Lankinen, Katri & Vaakanainen, Veijo (red.): *Svenskan i Finland* 18. Åbo: Åbo universitet, 164–180.
<https://www.utupub.fi/handle/10024/150838>
- Nyqvist, Eeva-Liisa & Peter Lundkvist. 2020. De lokala prepositionerna hos finska svenskinlärare i språkbad och traditionell undervisning. I Haapamäki, Saara, Ludvig Forsman & Linda Huldén (red.) *Svenskans beskrivning* 37. Svenska språket vid Åbo Akademi, 215–229.

- Odlin, Terence. 2003. Cross-Linguistic Influence. I Catherine Doughty & Michael H. Long (red.): *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing, 436–486.
- Pallotti, Gabriele. 2007. An operational definition of the emergence criterion. *Applied Linguistics*, 28–3, 361–382.
<https://doi.org/10.1093/applin/amm018>
- Philipsson, Anders. 2004. Svenskans morfologi och syntax i ett andraspråksperspektiv. I Hyltenstam, Kenneth & Inger Lindberg (red.): *Svenska som andraspråk – i forskning, undervisning och samhälle*. Lund: Studentlitteratur, 117–151.
- Philipsson, Anders. 2007. *Interrogative Clauses and Verb Morphology in L2 Swedish. Theoretical Interpretations of Grammatical Development and Effects of Different Elicitation Techniques*. Stockholm: Stockholm University.
<http://su.diva-portal.org/smash/record.jsf?pid=diva2%3A197193&dsid=469>
- Rahkonen, Matti & Gisela Håkansson. 2008. Production of Written L2-Swedish – Processability or Input Frequencies? I Kessler, Jörg-Ulrich (red.): *Processability Approaches to Second Language Development and Second Language Learning*. Newcastle: Cambridge Scholars Publishing, 135–161.
- Ranta, Leila. 2008. Metalinguistic knowledge and oral production. I Jasone Cenoz, & Nancy H. Hornberger (red.): *Encyclopedia of language and education. Volume 6: knowledge about language*. 2nd edition. Springer, 205–216.
- SAG 1999 = Teleman, Ulf, Hellberg, Staffan & Erik Andersson. 1999. *Svenska Akademiens Grammatik. Vol. 1–3*. Stockholm: Nordstedts Ordbok.
- Sundman, Marketta. 1995. Species och artikelbruk hos avancerade inlärare av svenska. I Maija Kalin & Sirkku Latomaa (red.): *Nordens språk som andraspråk 3*. Jyväskylä: Högskolornas språkcentral, 220–229.
- Tingbjörn, Gunnar. 1979. Kontrastiv minigrammatik. I Kenneth Hyltenstam, (red.) *Svenska i inlärarperspektiv. Kontrastiv analys och språktypologi*. Malmö: Liber, 41–78.
- Trenkic, Danjela & Nattama Pongpairoj. 2013. Referent salience affects language article use. *Bilingualism: Language and Cognition* 16–1, 152–166.
<https://doi.org/10.1017/S1366728912000156>
- Vilkuna, Maria. 1992. *Referenssi ja määräisyys suomenkielisten tekstien tulokinnassa*. Helsingfors: Finska Litteratursällskapet.

Summary

Complex Noun Phrases in Swedish by Finnish Immersion Students

This article explores how Finnish-speaking immersion students in the 6th and 9th grades (n = 77 and n = 86, respectively) judge the grammaticality of some types of complex Swedish noun phrases (NPs), i.e., double definiteness, NPs with possessive/genitive and adjective attributes, and NPs with the demonstrative pronouns *den här* and *denna* (“this”). They represent two types of complexity: *formal complexity* and *complexity of the relationship between form and meaning*.

The differences between the grades are usually statistically nonsignificant. The analyses at both the group level and the individual level show that the formally complex NPs are easier than the ones with a complex relationship between form and meaning. The frequent NPs are mastered at a higher level than the less frequent ones. Yet, there are informants in both grades who do not fulfil the criterion in any of the studied NPs.

Keywords: Noun Phrase, Definiteness, Complexity, Swedish as a Second Language, Immersion

Eeva-Liisa Nyqvist
Postdoktoral forskare i nordiska språk
Finska, finsk-ugriska och nordiska avdelningen
Humanistiska fakulteten
Helsingfors universitet
PB24
00014 Helsingfors universitet
eeva.liisa.nyqvist@helsinki.fi

Bilaga 1 : Nominalfraserna i testet med klassificering för frastyp (frastyperna ingick inte i det ursprungliga testet)

Namn: _____

Välj det rätta alternativet:

1. Den röda väskan/den röda väska är min.	dubbel bestämdhet, utr.sg
2. Har du hört om Kalles spännande äventyret/ Kalles spännande äventyr?	genitivattribut
3. Denna gång/denna gången vet jag vad jag måste göra.	denna/detta/dessa
4. Det röda äpplet/det röda äpple hade försvunnit från bordet.	dubbel bestämdhet, neutr.sg
5. Jag fick bråttom med att skriva den här texten/den här text.	den/det här
6. De smutsiga väggarna/de smutsiga väggar ser inte så bra ut.	dubbel bestämdhet, pl.
7. Jag har fått den här tavla/den här tavlan av min mormor.	den/det här
8. De gröna skogar/De gröna skogarna ser vackra ut.	dubbel bestämdhet, pl.
9. Jag älskar Floridas långa stränderna/Floridas långa stränder.	genitivattribut
10. Jag besökte de finaste museerna/ de finaste museer i Berlin.	dubbel bestämdhet, pl.
11. Den nya datorn/den nya dator är mycket snabbare än den gamla.	dubbel bestämdhet, utr.sg
12. Har din gamla tv/din gamla tv:n gått sönder?	possessivattribut
13. Har ni redan läst dessa böcker/dessa böckerna?	denna/detta/dessa
14. Har du inte sett det här foto/det här foto tidigare?	den/det här
15. Vem är din bästa vännen/din bästa vän?	possessivattribut
16. Den snygga flicka/Den snygga flickan i bilden är min flickvän.	dubbel bestämdhet, utr.sg
17. Jag älskar vår fina sommarstugan/vår fina sommarstuga.	possessivattribut
18. Det stora huset/det stora hus är stadshuset.	dubbel bestämdhet, neutr.sg
19. Har du sett mitt gula häftet/mitt gula häfte?	possessivattribut
20. Jag är lite rädd för grannens svarta hund/grannens svarta hunden.	genitivattribut
21. Det gamla skåp/det gamla skåpet borde målas på nytt.	dubbel bestämdhet, neutr.sg
22. Jag skulle vilja träffa världens bästa fotbollspelaren/ världens bästa fotbollsspelare.	genitivattribut
23. Det unga barnet/det unga barn kan ännu inte tala.	dubbel bestämdhet, neutr.sg
24. Jag har längre drömt om den här resan/den här resa.	den/det här
25. De söta jordgubbar/De söta jordgubbarna är det bästa med sommaren!	dubbel bestämdhet, pl.
26. Vad har ni gjort i skolan denna veckan/denna vecka?	denna/detta/dessa
27. På detta sätt/detta sättet lyckas vi mycket bättre.	denna/detta/dessa
28. Den gamla mannen/den gamla man är min morfar.	dubbel bestämdhet, utr.sg

Bilaga 2. Implikationstabellerna från kontrollgruppen (från Nyqvist 2018c)

Dubbel bestämdhet

Group	non-neuter			in total	%	deviations
	sg	neuter sg	pl			
1	+	+	+	23	52%	0
2	+	+	-	10	23%	0
3	+	-	+	3	7%	3
4	+	-	-	4	9%	0
5	-	+	-	1	2%	1
6	-	-	-	3	7%	0
	40	34	26	44	100%	4
	91 %	77 %	59 %			

C_{rep} 0.97, C_{act} 0.92

[

PRG-attributen

Group	den här	gen+adj	poss+adj	denna	in total	%	deviations
1	+	+	+	+	3	7%	0
2	+	+	+	-	20	45%	0
3	+	+	-	+	2	4.5%	2
4	+	+	-	-	4	10%	0
5	+	-	+	+	2	4.5%	4
6	+	-	-	+	1	2%	1
7	+	-	+	-	2	4.5%	2
8	+	-	-	-	2	4.5%	0
9	-	+	+	+	2	4.5%	0
10	-	+	+	-	1	2%	2
11	-	-	+	-	2	4.5%	2
12	-	+	-	-	2	4.5%	2
13	-	-	-	+	1	2%	0
	36	34	32	11	44	100%	15
	82 %	77 %	73 %	25 %			

C_{rep} 0.91, C_{scal} 0.84



Flaskehalshypotesen: Syntaks og morfologi blant norske andrespråksinnlærere av engelsk

Isabel Nadine Jensen¹ og Marit Westergaard²

UiT Norges Arktiske Universitet^{1,2} og NTNU Norges teknisk-naturvitenskapelige universitet²

Flaskehalshypotesen ('Bottleneck Hypothesis', Slabakova 2008, 2013) beskriver og forklarer hva som er vanskelig og hva som er enkelt i andrespråks-tilegnelse. Formålet med dette er å forstå prosessen knyttet til å lære seg et nytt språk etter førstespråket. Hypotesen argumenterer for at det er morfologiske konstruksjoner som er mest utfordrende å tilegne seg. Til sammenligning skal konstruksjoner som tilhører andre lingvistiske moduler, som syntaks og semantikk, være betraktelig enklere å tilegne seg. I denne artikkelen diskuterer vi resultater fra to ulike studier som direkte tester hypotesens prediksjoner for norske innlærere av engelsk som andrespråk. Resultatene viser delvis støtte for flaskehalshypotesen, men demonstrerer også at andre faktorer, som f.eks. påvirkning fra førstespråket og variasjon i innputtet må tas i betraktning dersom man ønsker å vurdere vanskelighetsgraden av å tilegne seg ulike deler av andrespråkets grammatikk.

Nøkkelord: Bottleneck Hypothesis, flaskehalshypotesen, morfologi, syntaks, samsvarsbøying, ordstilling, engelsk, norsk, andrespråkstilegnelse, V2

1 Introduksjon

Denne artikkelen presenterer to studier som undersøker hva som er vanskelig å tilegne seg i engelsk som andrespråk¹ (S2) blant førstespråkstalere av norsk. Vi måler vanskelighetsgraden ved å teste deltakernes evne til å identifisere ugrammatisk morfologi og syntaks i engelsk i en akseptabilitetstest.

1. Begrepene 'førstespråk' og 'andrespråk' refererer til den kronologiske rekkefølgen på språkene.

Vårt mål med disse studiene er å undersøke flaskehalshypotesens ('Bottle-neck Hypothesis', Slabakova 2008; 2013) prediksjoner om andrespråksinnlæring av engelsk blant innlærere med norsk som førstespråk (S1). Resultatene fra den første studien, som heretter kalles studie 1, er tidligere diskutert i I. N. Jensen et al. (2020) og finner delvis støtte for flaskehalshypotesen ved å vise at norske innlærere av engelsk har større problemer med å identifisere ugrammatisk subjekt-verb-kongruens (morfologi) enn ordstilling i ikke-subjektsinitiale deklarativsetninger (syntaks). Studie 2 baserer sine prediksjoner på resultatene i I.N. Jensen et al. (2020) men utvider undersøkelsen av flaskehalshypotesen ved å også teste ordstillinger i setninger med setningsadverbialer og preteritumsformen av svake verb. Disse variablene ble først diskutert i forbindelse med flaskehalshypotesen i M. Jensens (2017) mastergradsavhandling, men inkluderte ikke statistiske analyser. I denne artikkelen analyserer vi M. Jensens data i tillegg til nye data som er samlet inn ved bruk av M. Jensens metode.

Denne artikkelen er bygd opp på følgende måte: del 2 diskuterer bakgrunnen for studiene med fokus på faktorer som kan bidra til større eller mindre utfordringer i andrespråkstilegnelse av grammatikk, inkludert flaskehalshypotesens prediksjoner. Del 3 forklarer forskningsspørsmålene. Del 4 og 5 presenterer henholdsvis studie 1 og studie 2. Til slutt diskuteres resultatene fra studiene i del 6 og oppsummeres i del 7.

2 Bakgrunn

Flaskehalshypotesen er basert på en sammenligning av hvordan strukturer som tilhører ulike lingvistiske moduler, som syntaks, semantikk og morfologi, tilegnes (Slabakova, 2008, 2013). I følge denne hypotesen er morfologi flaskehalsen i andrespråkstilegnelse. Det vil si at elementer som tilhører denne modulen er mest utfordrende å tilegne seg. Hypotesen argumenterer for at ferdigheter i semantikk og syntaks utvikles i takt med eksponering for andrespråket, mens ferdigheter i morfologi i større grad stagnerer. Årsaken til utfordringene med morfologi knyttes til et argument om at morfemer representerer samlinger av morfologiske, syntaktiske og fonologiske trekk. Et eksempel er samsvarsmarkøren i engelsk, morfemet *-s*, som anses å representere en samling av abstrakte morfosyntaktiske trekk, [tempus], [aspekt] og [kongruens], i tillegg til trekk som sikrer at subjektet er til stede i setningene, at subjektet er i nominativ og at verbet ikke flytter ut av verbfrasen. Disse trekkene har konsekvenser for akseptabiliteten og betydningen til hele setningen, og alle må

tilegnes for å oppnå en målspråkslik grammatikk på andrespråket. Dette er ifølge flaskehalshypotesen særlig utfordrende når det finnes uoverensstemmelser mellom kombinasjonen av disse trekkene i S1 og S2. For å undersøke flaskehalshypotesens prediksjoner, tester vi evnen til å identifisere ugrammatisk syntaks og morfologi blant andrespråkstalere av engelsk med norsk som førstespråk. Dersom flaskehalshypotesens prediksjoner er riktige, vil det være vanskeligere å identifisere ugrammatisk morfologi enn ugrammatisk syntaks.

I begge studiene (studie 1 og 2, se del 4 og 5) bruker vi ordstilling som eksempel på et syntaktisk fenomen. Ordstillingen i norsk og engelsk er ulik. Engelsk omtales som et subjekt-verb-objekt-språk (SVO), mens norsk blir omtalt som et SVO-språk med V2 (V2 står for 'verb second', dvs. at det finite verbet plasseres i andre posisjon av setningen). Denne forskjellen ser vi i ikke-subjektsinitiale deklarativer hovedsetninger, hvor ordstillingen verb-subjekt (VS) er dominerende i norsk, mens subjekt-verb (SV) dominerer i engelsk. Eksempel (1a, b) illustrerer denne forskjellen.

- | | | |
|-----|---|--------|
| (1) | a. I går drakk <i>studentene</i> vin. | [XVSO] |
| | b. Yesterday <i>the students</i> drank wine. | [XSVO] |

I likhet med ikke-subjektsinitiale deklarativsetninger synliggjøres også V2-fenomenet i subjektsinitiale deklarativsetninger med setningsadverbial. I disse setningene har norsk ordstillingen subjekt-verb-adverb, mens engelsk som hovedregel har verbet etter adverbet, slik at ordstillingen blir subjekt-adverb-verb. Eksempel (2a, b) illustrerer forskjellene mellom norsk og engelsk. Vi tester sensitivitet for ugrammatisk ordstilling i setninger med adverbial i studie 2.

- | | | |
|-----|---|--------|
| (2) | a. Studentene drakk <i>ofte</i> vin. | [SVXO] |
| | b. The students <i>often</i> drank wine. | [SXVO] |

En tradisjonell forklaring på V2-fenomenet i norsk innenfor generativ syntaks er at verbet flytter fra VP (verbfrasen) til CP². Flyttingen skjer på grunn av et trekk i C-posisjonen som tiltrekker seg det finite verbet (se f.eks. Vikner 1995) og resulterer i at verbet plasseres i andre posisjon i setningen. Ikke-finite

2. CP refererer til 'complementizer phrase' (Åfarli & Eide 2003).

leksikalske verb flytter derimot ikke i engelsk, men forblir i VP-en. Dermed må førstespråkstalere av norsk *avlære* verbflytting når de lærer seg engelsk.

Vi valgte ordstilling som syntaktisk fenomen på bakgrunn av tidligere studier som viser at dette er utfordrende for andrespråksinnlærere. For eksempel viser Westergaard (2003) at verbplassering i engelsk som andrespråk er vanskelig for førstespråkstalere av norsk på tidlige stadier av tilegnelsesprosessen. Det morfologiske fenomenet vi valgte er subjekt-verb-kongruens (også referert til som samsvarsbøyning) i studie 1 og 2. Engelsk krever at samsvar mellom subjekt og verb markeres med morfemet *-s* på verbet når subjektet er tredje person entall. Samsvar mellom subjekt og verb markeres derimot ikke i norsk. Eksempel (3) og (4) illustrerer forskjellen mellom de to språkene.

- (3) a. Per og Mari **snakker** norsk.
 b. Per **snakker** norsk.
- (4) a. Peter and Mary **speak** English.
 b. Peter **speaks** English.

Tidligere forskning har vist at subjekt-verb-kongruens er utfordrende å tilegne seg på tvers av alder og ferdighetsnivå i engelsk som andrespråk; se f.eks. Haznedar (2001) som undersøkte førstespråkstalere av tyrkisk, og Lardiere (1998) som undersøkte en førstespråkstaler av kinesisk (se også Ionin & Wexler 2002). Flere studier viser også at andrespråksinnlærere i alle aldre gjør færre syntaktiske enn morfologiske feil (se f.eks. White 2003). Dette støtter hypotesen om *syntaks-før-morfologi*, introdusert av White (2003), dvs. at kunnskap om syntaktiske prosesser er på plass før kunnskap om morfologiske prosesser.

Barn som snakker engelsk som førstespråk tilegner seg vanligvis *-s*-morfemet konservativt, dvs. at når de har lært det, så bruker de det kun der det er nødvendig. Underbruk av *-s*-morfemet forekommer altså oftere enn overbruk. I andrespråkstilegnelse av engelsk har man imidlertid funnet både over- og underbruk blant innlærere med ulike førstespråk (se f.eks. Dröschel 2011; Vettorel 2014). Garshol (2018) har studert korpusdata fra norske innlærere av S2 engelsk og viser at 45 % av de totale samsvarsfeilene er utelatelse av *-s*, mens 55 % er overproduksjon. Korpuset som ble undersøkt bestod av 430.000 ord, produsert skriftlig av 15–16 åringer (N = 185) over en ettårsperiode.

I studie 2 tester vi også sensitivitet for ugrammatiske preteritumsformer. I likhet med subjekt-verb-kongruens er tempus også en morfologisk kategori. Mønsteret for regelmessig bøyning av preteritumsformen av verb er likt på

norsk og engelsk, ved at begge språkene markerer preteritum ved å legge til et suffiks på verbet, som vi ser i eksemplene (5) og (6).

- (5) a. Per og Mari spiste egg til frokost i går.
b. Per nådde ikke opp til den øverste hylla.
- (6) a. Peter and Mary painted their house yesterday.
b. Mary walked to the bus stop this morning.

Det er imidlertid andre faktorer enn tilhørighet til lingvistisk modul (syntaks eller morfologi) som kan påvirke vanskelighetsgraden av å tilegne seg et trekk, slik som lærbarhet (positiv og negativ evidens i innputtet), synlighet i innputtet, frekvens og instruksjon. I de neste avsnittene diskuterer vi noen av disse faktorene i lys av trekkene som testes i studie 1 og 2: subjekt-verb-kongruens, preteritum og ordstilling.

Tidligere generative studier (Trahey & White 1993, Mazurkewich & White 1984, Inagaki 2006, Gabriele 2009, Montrul & Yoon 2009: 308) har vist at det er lettere å tilegne seg et nytt trekk, dvs. et trekk som finnes i andrespråket og ikke i førstespråket, enn å kvitte seg med (avlære) et trekk som finnes i førstespråket og ikke i andrespråket. Dette forklares med at negativ evidens, dvs. informasjon om hva som er ugrammatisk i et språk, ikke er tilgjengelig i innputtet fra målspråket. Det er med andre ord umulig for en andrespråksinnlærer å bruke innputtet til å finne ut hvilke trekk som *ikke* finnes i målspråket. Trahey & White (1993) kalte denne avlæringsprosessen for *preemption*. Nye trekk, derimot, synliggjøres gjennom positiv evidens, dvs. informasjon om hva som er grammatisk i et språk.

Den vanligste ordstillingen i engelsk er subjekt-verb-objekt (SVO). Denne ordstillingen kamuflerer en viktig forskjell mellom norsk og engelsk syntaks, nemlig at norsk har V2-regelen (se eksempel (1a, b)). For norske innlærere kan derfor setninger med ordstillingen SVO gi inntrykk av at engelsk fungerer på samme måte som norsk. Enkelte studier i andrespråkstilegnelse viser at V2-fenomenet kan bli overført til engelsk til tross for at det regnes som et markert trekk ved norsk. Dette kommer blant annet til uttrykk i feil ordstilling i både ikke-subjektsinitiale og subjektsinitiale deklarativsetninger i engelsk (for eksempel **Yesterday went John to the shop* og **I eat always eggs for breakfast*). Westergaard (2003) viser at denne overføringen av V2-regelen til engelsk er vanlig blant norske innlærere på (svært) tidlige stadier av tilegnelsesprosessen. Bohnacker (2006) viser også at V2-ordstillingen kan overføres fra S1 svensk til S2 tysk. Dette er basert på en sammenligning av ordstilling i ikke-sub-

jektsinitiale deklarativsetninger blant to grupper med tysk-innlærere. Den ene gruppen hadde svensk som S1 og lærte seg tysk som S2. Den andre gruppen hadde også svensk som S1, engelsk som S2 og tysk som tredjespråk (S3). Mens S3-innlærerne ofte brukte SVO-ordstilling (slik som man også gjør i engelsk), var S2-gruppen konsekvent i sin bruk av V2 i tysk. Westergaard (2003) argumenterer for at det er viktig å eksponere innlærerne for setninger som tydelig illustrerer kontrasten mellom de to språkene, slik som for eksempel XSV (*Yesterday John went to the shop*). Følgelig må V2-ordstillingen avlæres, og det forventes at dette er relativt utfordrende for norske innlærere av engelsk. Vi refererer til denne faktoren som *lærbarhet* (se tabell 1).

En annen faktor er *innputtfrekvens* (Ellis 2002). Morfemet *-s* som markerer samsvar mellom subjekt og verb i engelsk forekommer ofte i innputtet, fordi det må være til stede i alle setninger i presens med subjektet i tredje person entall. I COCA (Corpus of Contemporary American English, Corpus.byu.edu 2018a), som er et korpus bestående av 520 millioner ord, forekommer samsvarsmarkøren *-s* totalt 6.198.523 ganger. Det vil si at dette morfemet er til stede i så mye som 37,5 % av alle relevante kontekster (leksikalske verb i presens; se Vedlegg A). Når man ser bort fra skriftlige kilder i COCA, forekommer *-s* i 23,4 % av alle relevante kontekster (944.638 ganger; se Vedlegg B). Denne hyppige forekomsten betyr at innlærere av engelsk ofte er eksponert for *-s*-morfemet, og at det er tvilsomt at eventuelle problemer med å tilegne seg subjekt-verb-kongruens er forårsaket av mangel på innputt.

Som tidligere nevnt, argumenterer Westergaard (2003) for at innlærere må eksponeres for XSV-ordstilling for å avlære seg V2. Hun påpeker at XVS-setninger brukes relativt hyppig i norsk (ca. 30 %, se også Lightfoot 1999; Westergaard 2008), men relativt sjelden i engelsk. Dette støttes av Yang (2001), som har undersøkt moderne engelsk i Penn Treebank-korpuset.³ Han viser at rundt 10 % av alle deklarativsetninger i engelsk har ordstillingen XSV eller SXV. Videre har Charles Yang (personlig kommunikasjon) utført et søk i Pearl Sprouse-korpuset⁴ (se Vedlegg C), hvor han fant XSV i 6,8 % av alle deklarativsetninger (3 114 av totalt 36 117 setninger). Dette illustrerer at forekomsten av samsvarsmarkøren *-s* (37,5 %) er betydelig høyere enn ikke-V2 ordstilling i engelsk (6,8 %).

En tredje faktor er *synlighet i innputtet*, dvs. at grammatiske morfemer har en tendens til å bli forkortet og sammenflettet med materialet rundt fordi de er

3. Førstespråkstalere av engelsk (tekst og tale).

4. Førstespråkstalere av engelsk (tale).

de mest frekvente elementene i et språk (Zipf 1949; Bybee 2000). De er dermed ikke spesielt fremtredende i innputtet sammenlignet med verb eller substantiv. For eksempel har samsvarsmarkøren *-s* ingen vokal, den utgjør ikke en stavelse, den er svært ofte del av en konsonantklynge, og i motsetning til flertallsmarkøren *-s* er den semantisk redundant (se f.eks., Goldschneider & DeKeyser 2005). I lys av disse observasjonene kan man stille spørsmål ved om det i det hele tatt er mulig å sammenligne ordstilling med subjekt-verb-kongruens når det gjelder synlighet i innputtet. Vi har ikke kjennskap til noen slike direkte sammenligninger, men Cintrón-Valentín & Ellis (2016) sammenligner samsvarbøyning med blant annet preposisjonsfraser og temporale adverb. De argumenterer for at samsvarbøyning er mindre fremtredende i innputtet fordi samsvarsmarkører er kortere og i mindre grad vektlagt i talestrømmen og dermed vanskeligere å legge merke til.

I studie 1 gikk vi igjennom lærematerialet som ble brukt i deltakernes engelskundervisning for å undersøke graden av eksplisitt informasjon om samsvarsregler, tempus og ordstilling. Materialet bestod av lærebøker med tekst og oppgaver, i tillegg til tjenesten NDLA (Nasjonal Digital Læringsarena), som er en nettbasert samling av læremateriale laget for elever i videregående skole. NDLAs engelskside har totalt 1644 oppgaver og 750 enheter med lesemateriale (NDLA 2018b). Vår undersøkelse viser at ingen av lærebøkene eksplisitt forklarer engelsk ordstilling, mens samsvarbøyning er beskrevet i alle lærebøkene (se Vedlegg D). NDLA beskriver ordstilling i én av de 750 leseenhetene, mens subjekt-verb-kongruens beskrives ti ganger.

Det finnes to grunner til at oppmerksomhet i lærematerialet er en viktig faktor for våre forskningsspørsmål. For det første bidrar eksplisitt informasjon til økt oppmerksomhet rettet mot det aktuelle språklige trekket. For det andre viser dette klar evidens for trekket. Cintrón-Valentín og Ellis (2016) viste at lite synlige morfosyntaktiske trekk kan læres gjennom form-fokusert instruksjon, dvs. eksplisitt oppmerksomhet mot trekkene, blant annet fordi denne undervisningsformen bidrar til at innlærerne vektlegger visse trekk i språket som de ellers ville ignorert. Tabell 1 illustrerer faktorene som er diskutert her og som kan påvirke vanskelighetsgraden av tilegnelse.

Tabell 1: Faktorer som påvirker hvor enkelt eller vanskelig det er å tilegne seg samsvarsregler, tempus og ordstilling i engelsk som andrespråk av innlærere med norsk som førstespråk. Kursiv font indikerer at konstruksjonen skal være lettere å tilegne seg.

	Grammatisk modul	Innflytelse fra S1	Lærbarhet	Frekvens	Synlighet	Læremidler
Kongruens	Morfologi	Ingenting å overføre	<i>Læres fra positiv evidens</i>	<i>Høy frekvens</i>	Lav synlighet	<i>Beskrevet i klasseromsundervisning</i>
Preteritum (svak)	Morfologi	<i>Kan overføres</i>	<i>Læres fra positiv evidens</i>	<i>Høy frekvens</i>	Lav synlighet	<i>Beskrevet i klasseromsundervisning</i>
XSVO	<i>Syntaks</i>	Ja, V2 kan overføres og må avlæres	Avlæring gjennom XSV i innputt og negativ evidens	XSV er mindre frekvent	<i>Høyere synlighet</i>	Vanligvis ikke beskrevet
SAdvV	<i>Syntaks</i>	Ja, V2 kan overføres og må avlæres	Avlæring gjennom XSV i innputt og negativ evidens	XSV er mindre frekvent	<i>Høyere synlighet</i>	Vanligvis ikke beskrevet

Ifølge flaskehalshypotesen skal subjekt-verb-kongruens utgjøre den største utfordringen for innlærerne (se andre kolonne av tabell 1). Det er imidlertid slik at også de andre faktorene i tabellen kan påvirke den totale vanskelighetsgraden, og tabellen viser at ikke alle faktorene trekker i samme retning. For eksempel vil avlæring av V2-ordstilling være avhengig av eksponering for ikke-V2 i ikke-subjektsinitiale deklarativer (XSV), mens kongruens må læres. Frekvens og fravær av instruksjon tilsier at ikke-V2 i engelsk skal være utfordrende. På den andre siden er samsvars-s og preteritum lite fremtredende i innputt sammenlignet med ordstilling. Derfor, hvis kongruens viser seg å være vanskeligere enn ordstilling, vil dette være på tross av faktorer som læring kontra avlæring, frekvens og instruksjon.

3 Forskningsspørsmål og prediksjoner

Vi stilte følgende forskningsspørsmål i begge studiene:

Forskningsspørsmål 1: Er det vanskeligere å identifisere ugrammatisk syntaks enn ugrammatisk morfologi for norske S2-innlærere av engelsk?

Forskningsspørsmål 2: Forbedres evnen til å identifisere ugrammatisk syntaks og morfologi i takt med økt generelt ferdighetsnivå?

Flaskehalshypotesen argumenterer for at det er lettere å tilegne seg syntaks enn morfologi. Dersom dette stemmer, forventet vi for det første at det ville være vanskeligere for deltakerne å identifisere ugrammatisk morfologi enn ugrammatisk syntaks. Vi forventet med andre ord gjennomsnittlig lavere differanseskår for variabler som representerer morfologi (utregningen av differanseskår er beskrevet i del 4). For det andre forventet vi at det ville være mindre forskjeller mellom deltakere på ulike ferdighetsnivåer når det gjelder morfologi enn syntaks. Det vil si at vi forventet en jevnere differanseskår for morfologi på tvers av generelt ferdighetsnivå i engelsk. Grunnen til denne forventningen er flaskehalshypotesens argument om at morfologi ikke bare er vanskeligere å lære, men også tar lengre tid.

M. Jensen (2017: 47) påpeker at flaskehalshypotesen hevder at preteritum er vanskelig å tilegne seg i et andrespråk, i og med at tempus i engelsk blir fremhevet som vanskelig for mange innlærere med forskjellige S1 (se Slabakova, 2013). Morfologi har imidlertid først og fremst vist seg å være problematisk når det finnes en forskjell mellom S1 og målspråket (se del 2, Slabakova, 2013). Det er kun subjekt-verb-kongruens som representerer en forskjell mellom norsk og engelsk morfologi, og flaskehalshypotesen blir dermed ikke direkte falsifisert selv om preteritum skulle være lettere å tilegne seg enn kongruens. Det er imidlertid problematisk for flaskehalshypotesen dersom en syntaktisk konstruksjon er vanskeligere å tilegne seg enn en morfologisk konstruksjon. I lys av flaskehalshypotesen forventet vi derfor gjennomsnittlig høyere differanseskår for syntaktiske variabler og preteritum sammenlignet med subjekt-verb-kongruens. Vi forventet også mindre forskjeller mellom deltakere på ulike ferdighetsnivåer når det gjelder subjekt-verb-kongruens på den ene siden og de øvrige variablene på den andre siden.

4 Studie 1

4.1 Metode og materiale

Studien bestod av tre deler. Den første delen var en akseptabilitetstest, den andre delen var en ferdighetstest og den tredje delen var et spørreskjema om språklig bakgrunn. Alle delene ble gjennomført elektronisk og skriftlig. I akseptabilitetstesten ble deltakerne bedt om å vurdere engelske setninger på en Likert-skala

fra én til fire, hvor én betydde at setningen var helt uakseptabel på engelsk og fire betydde at setningen var helt akseptabel. Deltakerne hadde også mulighet til å svare *vet ikke*. Instruksjonene ble gitt på norsk, både muntlig og skriftlig. Seks variabler var inkludert i akseptabilitetstesten, illustrert ved eksempler i (7) til (10).

- (7) *Ikke-subjektsinitiale deklarativsetninger med leksikalske verb*
 a) *Yesterday **went the teacher** to the shop.
 b) Yesterday **the teacher went** to the shop.
- (8) *Ikke-subjektsinitiale deklarativsetninger med hjelpeverb*
 a) *Every day **should the students** bring their books to school.
 b) Every day **the students should** bring their books to school.
- (9) *Subjekt-verb-kongruens (flertall)*
 a) ***The teachers gives** their students a lot of homework.
 b) **The teachers give** their students a lot of homework.
- (10) *Subjekt-verb-kongruens (entall)*
 a) ***The brown dog play** with the yellow football.
 b) **The brown dog plays** with the yellow football.

Det var totalt 36 setningspar i testen (seks setninger per variabel)⁵. Et setningspar bestod av én ugrammatisk setning og én tilsvarende grammatisk setning. Dette vises i eksemplene ovenfor hvor (a) viser den ugrammatiske setningen (b) viser den grammatiske versjonen. Alle testsetningene var omtrent like lange (10–12 stavelser) og de leksikalske elementene ble valgt ut på grunnlag av COCAs liste over de 5000 mest frekvente ordene (corpus.byu.edu 2018b). Ikke-subjektsinitiale deklarativsetninger hadde enten sterke verb i preteritum eller modalverb + infinitiv for å unngå at deltakerne gjorde sine akseptabilitets-

5. I I. N. Jensen (2016, se også I. N. Jensen et al. (2020)) testes også samsvar mellom verb og subjekt med preposisjonsfrase (f.eks. *The teacher with the black shoes walks...*). Dette blir ofte kalt langdistanse-kongruens. I slike tilfeller forekommer ofte såkalte 'attraction errors', som vil si at verbet samsvarer med substantivet i preposisjonsfrasen istedenfor subjektet, fordi dette står nærmere verbet (se f.eks., Bock & Miller 1991, Cunnings 2017). Flere studier (f.eks. Foote 2011, Almor et al 2001, Cunnings 2017) hevder at årsaken til denne typen feil ikke er relatert til kunnskap om kongruens, men heller til prosesseringsvansker. For eksempel argumenterer Keating (2005) for at det er vanskelig å beholde informasjon om numerus (entall/flertall) på subjektet i arbeidsminnet når flere elementer (f.eks. en preposisjonsfrase) skiller subjektet og verbet. Vi valgte derfor kun å fokusere på subjekt-verb-kongruens uten preposisjonsfraser i denne artikkelen.

vurderinger basert på kongruens. De deklarativer setningene som ble brukt for å teste kongruens inneholdt svake verb.

Vi var interessert i deltakernes evne til å identifisere ugrammatiske setninger. Vi målte dette ved å trekke verdien en deltaker ga til en grammatisk setning (1–4 på Likert-skalaen) fra verdien gitt til den ugrammatiske setningen (også 1–4). Dette gir oss en 7-punkts skala fra –3 til 3. For eksempel fikk vi verdien –3 dersom en deltaker ga verdien 1 til den grammatiske setningen og verdien 4 til den ugrammatiske setningen i et setningspar; verdien 0 fikk vi dersom begge setningene i et par fikk samme verdi, og en positiv verdi fikk vi når den grammatiske setningen ble vurdert som bedre enn den ugrammatiske setningen. Dette målet kaller vi differanseskår.

Den generelle ferdighetstesten var en standardisert luketest der deltakerne ble bedt om å fylle inn manglende ord eller fraser i en engelsk tekst. Deltakerne kunne velge blant tre svaralternativer. Subjekt-verb-kongruens med svake verb og ordstilling ble ikke testet i denne ferdighetstesten. Deltakerne fikk ett poeng per riktig ord/frase, og den høyeste mulige poengsummen var 40.

Til slutt svarte deltakerne på spørsmål om alder og språklig bakgrunn. Vi ekskluderte deltakere som hadde flere eller andre førstespråk enn norsk fra analysen.

4.2 Deltakere

Seksti elever mellom 11 og 18 år deltok i studien. Elevene ble rekruttert fra offentlige norske skoler i forskjellige deler av landet, fra syvende klasse på barneskolen (N=25), samt første (N=27) og andre klasse (N=17) på videregående skole. Skolene valgte selv ut hvilke klasser som skulle delta. De hadde alle lært engelsk på skolen fra og med første klasses trinn, dvs. fra 5–6-årsalderen. Deltakernes generelle ferdighetsnivå i engelsk ble målt ved hjelp av ferdighetstesten (se del 4.1). Poengsummene per deltaker varierte fra 11 til 38 av 40 mulige poeng (gjennomsnitt = 27,23).

4.3 Resultater

Resultatene fra studie 1 er tidligere diskutert i I. N. Jensen et al. (2020). I denne her artikkelen fokuserer vi på de variablene som best belyser flaskehalshypotesen, det vil si XSVO og lokal subjekt-verb-kongruens (se også del 4.1).

I den statistiske analysen brukte vi som nevnt differanseskår som avhengig variabel (se del 4.1). Dersom deltakerne hadde valgt alternativet *vet ikke* ble dette behandlet som en manglende verdi, og begge versjoner av setningen ble tatt ut av analysen. Vi brukte en lineær blandet modell ('linear mixed model')

(Bates et al. 2015). En blandet modell tar i betraktning både faste effekter (*fixed effects*) og tilfeldige effekter (*random effects*). Faste effekter refererer til variabler som kan forklare variasjon i dataene. I vår modell var disse variablene *konstruksjon* (samsvar mellom verb og flertallssubjekt, samsvar mellom verb og entallssubjekt, XSV-ordstilling med hjelpeverb og XSV-ordstilling med leksikalsk verb) og *ferdighetsnivå* (11–38 av 40 mulige poeng). Tilfeldige effekter representerer avvik fra prediksjonene våre som skyldes faktorer som ikke kan kontrolleres i eksperimentet. Vi inkluderte *deltaker* og *setning* som tilfeldige effekter. *Deltaker* kontrollerte for individuelle forskjeller og *setning* kontrollerte for forskjeller mellom de ulike testsetningene.

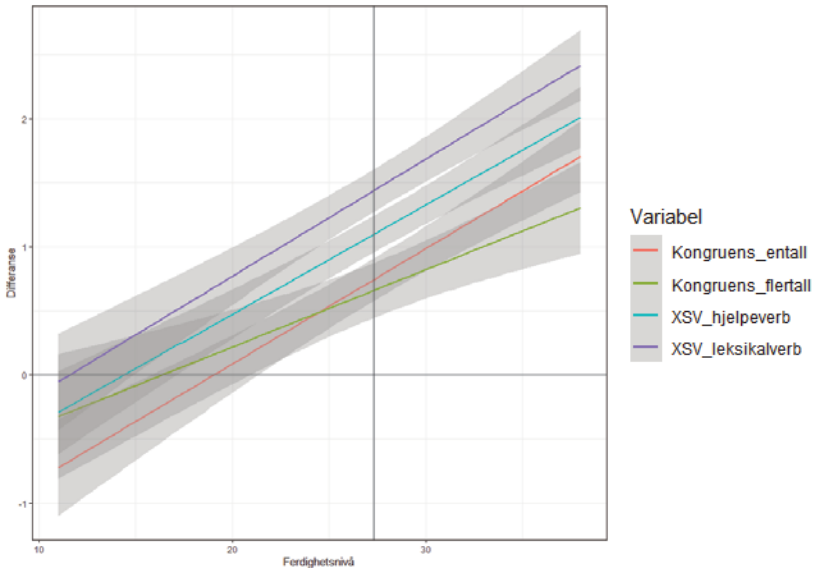
I tillegg til at vi forventet at *konstruksjon* og *ferdighetsnivå* hadde en effekt på differanseskåren, forventet vi også en interaksjon mellom *konstruksjon* og *ferdighetsnivå*. Det vi si at vi forventet at deltakernes ferdighetsnivå i engelsk påvirket differanseskåren i ulik grad basert på hvilken konstruksjon det var snakk om. Vi testet om *konstruksjon*, *ferdighetsnivå* og interaksjonen mellom disse to variablene hadde en signifikant effekt på differanseskårene med sannsynlighetstester (Pinheiro & Bates 2000; Winter 2013). Dersom *p*-verdien (sannsynlighetsverdien) var lavere enn 0,05 ble effekten regnet som statistisk signifikant.

Resultatene viste en signifikant effekt av *ferdighetsnivå* ($\chi^2(8) = 177,89$, $p < ,0001$). Det vil si at det finnes en signifikant forskjell mellom deltakere på ulike ferdighetsnivåer. I Figur 1 viser X-aksen deltakernes poengsummer på ferdighetstesten, mens Y-aksen viser differanseskåren. Desto høyere poengsum innlærerne fikk, desto høyere differanseskår. Resultatene viste også en signifikant effekt av *konstruksjon* ($\chi^2(7) = 11,25$, $p < ,0001$). Det betyr at det er en signifikant forskjell mellom konstruksjonene, altså at den gjennomsnittlige differanseskåren varierer basert på hvilken konstruksjon det er snakk om.

Post-hoc-tester (*t*-tester) viste hvordan konstruksjonene skilte seg fra hverandre. De to variablene som representerer kongruens var begge signifikant forskjellige fra ordstilling: Den gjennomsnittlige differanseskåren for samsvar mellom verb og flertallssubjekt er signifikant lavere enn ikke-subjektsinitiale deklarativer med leksikalske verb (SE = ,08, $t = 10,33$, $p < ,0001$) og med hjelpeverb (SE = ,08, $t = 5,82$, $p < ,0001$). Det samme gjelder for den gjennomsnittlige differanseskåren for samsvar mellom verb og entallssubjekt sammenlignet med leksikalske verb (SE = ,08, $t = 8,97$, $p < ,0001$) og hjelpeverb (SE = ,78, $t = 4,46$, $p < ,0001$).

Vi fant også en signifikant interaksjon for *ferdighetsnivå* og *konstruksjon* ($\chi^2(2) = 168,17$, $p < ,0001$). En signifikant interaksjon betyr at konstruksjonene

er forskjellige fra hverandre når det gjelder hvordan ferdighetsnivået påvirker differanseskåren. Dette kan vi se i Figur 1 ved å sammenligne hvor bratte linjene er. For eksempel er linjen som representerer ordstilling med leksikalske verb tydelig brattere enn den som representerer samsvar mellom verb og flertalls-subjekt. Interaksjonen mellom ferdighetsnivå og samsvarsbøyning med flertallssubjekt er signifikant forskjellig fra de andre interaksjonene: ferdighetsnivå og samsvar mellom entallssubjekt og verb ($SE = ,01, t = 2,86, p = ,004$), ferdighetsnivå og ordstilling i setninger med leksikalske verb ($SE = ,01, t = 3,37, p = ,001$) og ferdighetsnivå og ordstilling i setninger med hjelpeverb ($SE = ,01, t = 2,37, p = ,018$). I Figur 1 visualiseres dette ved at linjen for samsvarsbøyning med flertallssubjekt (Kongruens_flertall) er slakere enn de andre linjene.



Figur 1: Differanseskår per konstruksjon i studie 1 (-3-3) og ferdighetsnivå (11-38).

5 Studie 2

I M. Jensens (2017) mastergradsavhandling ble flaskehalshypotesen testet ved hjelp av fire variabler. To av variablene representerte morfologi: Preteritum og

subjekt-verb-kongruens. De to andre variablene representerte syntaks: Ordstilling i ikke-subjektsinitiale deklarativsetninger og ordstilling i subjektsinitiale deklarativsetninger med setningsadverbial. Ifølge M. Jensen (2017) var formålet med mastergradsavhandlingen å teste flaskehalshypotesen ved bruk av flere variabler enn de som ble testet i I. N. Jensen (2016). Det ble imidlertid ikke foretatt en statistisk analyse. M. Jensen (2017: 33) begrunnet dette med at det var for få deltakere i studien. Tendenser i resultatene tyder likevel på at det finnes potensielle utfordringer for flaskehalshypotesens prediksjoner. Vi ønsket derfor å utforske disse resultatene statistisk. Vi brukte derfor M. Jensens metode for å samle inn ytterligere data (N = 37) fra samme populasjon. Resultatene ble slått sammen med resultatene i M. Jensen (2017, dvs. N = 67) og presenteres her ved hjelp av statistiske modeller.

5.1 Metode og materiale

Studien bestod av de samme tre delene som i studie 1: en akseptabilitetstest, en ferdighetstest og et spørreskjema om språklig bakgrunn. Setningene ble delt inn i fire variabler: ikke-subjektsinitiale deklarativsetninger, subjektsinitiale deklarativsetninger med setningsadverbial, tempus (preteritum) og subjekt-verb-kongruens; se (11)–(14). De to førstnevnte variablene representerte syntaks og de sistnevnte representerte morfologi.

- (11) *Ikke-subjektsinitiale deklarativsetninger*
- *Last night **opened the girl** a present from her dad.
 - Last night **the girl opened** a present from her dad.
- (12) *Subjektsinitiale deklarativsetninger med setningsadverbial*
- *The girl **played always** soccer with her brother.
 - The girl **always played** soccer with her brother.
- (13) *Preteritum*
- *The baker **bake** a cake **two hours ago**.
 - The baker **baked** a cake **two hours ago**.
- (14) *Subjekt-verb kongruens*
- ***The teacher talk** about mathematics and numbers.
 - The teacher talks** about mathematics and numbers.

Testen inneholdt fem setningspar per variabel, totalt 20 setningspar (40 setninger). Et setningspar bestod av én grammatisk setning og én tilsvarende ugrammatisk setning. Igjen brukte vi differanseskår på akseptabilitetstesten som avhengig variabel. Det vil si at vi undersøkte hvordan differanseskåren varierte

mellom de ulike konstruksjonene (ordstilling i ikke-subjektsinitiale deklarativsetninger, ordstilling i deklarativsetninger med setningsadverbial, tempus og subjekt-verb kongruens). Ferdighetstesten og bakgrunnsskjemaet ble gjennomført etter akseptabilitetstesten. Deltakerne fikk det samme bakgrunnsskjemaet som i studie 1 og en forkortet versjon av ferdighetstesten (se del 4.1). Maksimal poengsum på ferdighetstesten var 29.

5.2 Deltakere

Det deltok 67 personer i alderen 9–16 i studien. Som tidligere nevnt ble 30 av disse deltakerne testet i forbindelse med M. Jensens (2017) mastergradsavhandling. Deltakerne var fordelt på tre klassetrinn fra ulike deler av landet: fjerde klasse (N=15), åttende klasse (N=27) og første klasse på videregående skole (N=25). Større variasjon i alder var ønskelig for å teste flaskehalshypotesen på tvers av ulike ferdighetsnivåer. Skolene valgte selv ut hvilke klasser som skulle delta, og alle deltakerne hadde norsk som eneste S1 og engelsk som S2, tilegnet i en klasseromskontekst fra seksårsalderen. På ferdighetstesten oppnådde deltakerne 19,4 i gjennomsnittlig poengsum av totalt 29 mulige poeng, og poengsummene varierte fra 6 til 27.

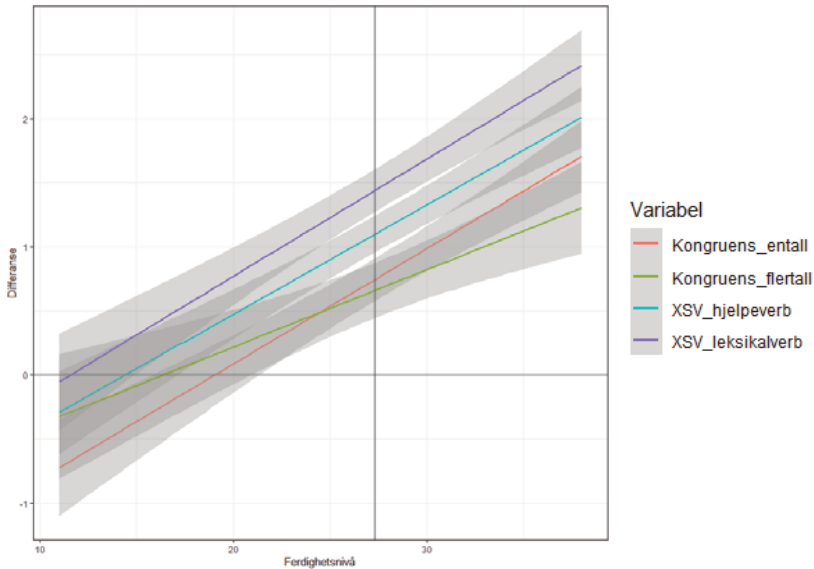
5.3 Resultater

Studie 2 hadde samme forskningsdesign som studie 1, og vi brukte derfor den samme statistiske modellen for å analysere resultatene (se del 4.3 for begrunnelse av bruken av en lineær blandet modell). Den avhengige variabelen var differanseskår (se også 4.1), dvs. verdien gitt til den grammatiske setningen (1–4) minus verdien gitt til den ugrammatisk versjonen av setningen (1–4). En positiv verdi betød at den grammatiske setningen var rangert som mer akseptabel enn den ugrammatisk (i et setningspar) og en negativ verdi var et resultat av at den ugrammatisk setningen hadde blitt vurdert som bedre enn den grammatisk setningen. Som faste effekter brukte vi variablene *ferdighetsnivå* (6–27 av 29 mulige poeng) og *konstruksjon* (subjekt-verb-kongruens, tempus, ordstilling i ikke-subjektsinitiale deklarativsetninger og ordstilling i subjektsinitiale deklarativsetninger med setningsadverbial). Vi så også etter en interaksjon mellom *ferdighetsnivå* og *konstruksjon*. Som tilfeldige effekter brukte vi variablene *deltaker* og *setning*.

Vi fant en signifikant hovedeffekt av *ferdighetsnivå* ($\chi^2(8) = 138,40, p < ,0001$) og *konstruksjon* ($\chi^2(7) = 78,02, p < ,0001$). Dette betyr at deltakernes ferdighetsnivå påvirker differanseskåren, slik at differanseskåren øker i takt med poengsummen på ferdighetstesten, som vi ser i Figur 2. Vi ser også at differanse-

skåren er signifikant forskjellig fra konstruksjon til konstruksjon. De laveste differanseskårene finner man ved subjekt-verb-kongruens og adverb-verb-ordstilling (S-Adv-V). Disse to variablene er ikke signifikant forskjellig fra hverandre når det kommer til gjennomsnittlig differanseskår (SE = ,09, $t = -1,75$, $p = 0,08$), men de er signifikant forskjellige fra de resterende variablene ($p < ,05$).

Vi fant også en signifikant interaksjon mellom *ferdighetsnivå* og *konstruksjon* ($\chi^2(2) = 132,50$, $p < ,0001$). Mer spesifikt er interaksjonen mellom ferdighetsnivå og subjekt-verb-kongruens signifikant forskjellig fra interaksjonen mellom ferdighetsnivå og tempus (SE = ,015, $t = 2,77$, $p = 0,006$). I Figur 2 vises dette ved at linjen som representerer subjekt-verb-kongruens er tydelig slakere enn linjen som representerer preteritum.



Figur 2: Differanse mellom verdier gitt til setningsparene for undervariablene (-3–3) og ferdighetsnivå (6–27).

6 Diskusjon

I denne delen diskuterer vi resultatene fra studie 1 og 2 i lys av forskningsspørsmålene og flaskehalshypotesens prediksjoner. Forskningsspørsmålene er gjentatt her (se også seksjon 3):

Forskningsspørsmål 1: Er det vanskeligere å identifisere ugrammatisk syntaks enn ugrammatisk morfologi for norske S2-innlærere av engelsk?

Forskningsspørsmål 2: Forbedres evnen til å identifisere ugrammatisk syntaks og morfologi i takt med økt generelt ferdighetsnivå?

Flaskehalshypotesen hevder at morfologiske trekk er vanskeligere å lære enn syntaktiske trekk. Imidlertid kan faktorer som lærbarhet, frekvens og instruksjon, som diskutert i del 2 (se tabell 1), også spille inn og trekke i motsatt retning. Resultatene fra studie 1 viser at de fleste deltakerne var gode til å akseptere grammatiske setninger, men dette betyr ikke nødvendigvis at deltakerne har tilegnet seg den grammatiske konstruksjonen. For å kunne hevde dette, må deltakerne også avvise ugrammatiske setninger. Derfor ble evnen til å identifisere ugrammatikalitet brukt som et mål på tilegnelse i våre studier.

Som Figur 1 viser, starter gjennomsnittsforskjellen mellom de grammatiske og ugrammatiske setningene under null for deltakerne med de laveste generelle ferdighetsskårene i engelsk, men klatrer over null på mer avanserte nivåer. Dette gjelder både ordstilling og subjekt-verb-kongruens. Vi ser med andre ord en positiv assosiasjon mellom generelt ferdighetsnivå og avvisning av ugrammatikalitet for begge konstruksjonene.

Videre undersøkte vi om en av konstruksjonene er mindre problematisk enn den andre. Som Figur 1 viser, er innlærerne i gjennomsnitt flinkere til å identifisere ugrammatisk ordstilling enn ugrammatisk subjekt-verb-kongruens. Dette ser vi ved et signifikant høyere gjennomsnitt for ordstilling. Dette er i tråd med flaskehalshypotesens prediksjoner for forskningsspørsmål 1.

Vi ser også forskjeller når det gjelder interaksjon mellom ferdighetsnivå og konstruksjon som får konsekvenser for flaskehalshypotesens prediksjoner om forskningsspørsmål 2. Kurven som representerer samsvar mellom entallssubjekt og verb (se figur 1) oppfører seg ikke signifikant forskjellig fra de syntaktiske variablene. Dette er potensielt problematisk for flaskehalshypotesen, ettersom den forventer en slakere kurve for morfologiske variabler, som ville indikert et mer vedvarende problem. Problemer med subjekt-verb-kongruens kan imidlertid komme til uttrykk på to måter: overaksept eller underaksept. Underaksept betyr aksept av ukorrekt fravær av *-s*, det vil si i setninger med entallssubjekter, slik som for eksempel **The dog run*. Overaksept betyr aksept av ukorrekt tilstedeværelse av *-s*, dvs. i setninger med flertallssubjekter, slik som **The dogs runs*. Studie 1 viser at både overaksept og underaksept forekommer, men interaksjonen mellom underaksept (samsvar mellom entallssubjekt og verb) og ferdighetsnivå er altså ikke signifikant forskjellig fra de andre variablene. Imid-

lertid er kurven som viser samsvar mellom flertallssubjekt og verb signifikant slakere enn de andre kurvene. Man ser med andre ord at overaksept av *-s*-morfemet er mest problematisk for deltakerne på tvers av ferdighetsnivå. Denne variabelen representerer morfologi og resultatet kan derfor til en viss grad sies å støtte flaskehalshypotesens prediksjoner for forskningsspørsmål 2. Imidlertid er det et problem for flaskehalshypotesen at underaksept ikke er signifikant forskjellig fra de syntaktiske konstruksjonene.

Disse observasjonene peker mot at avlæring av V2-regelen kan være mindre problematisk enn innlæring av kongruens i S2 engelsk, demonstrert ved overaksept av *-s*-morfemet. Dette er tilfellet til tross for at de andre faktorene som er nevnt i tabell 1 trekker i motsatt retning: samsvarsmarkøren er mer frekvent enn XSV-ordstilling i engelsk, den er lærbar fra positiv evidens i innputt og forklart i lærematerialet.

I likhet med studie 1, viser studie 2 at subjekt-verb-kongruens (med entall-subjekt) er vanskeligere enn ordstilling i ikke-subjektsinitiale deklarativsetninger. Dette illustreres med en signifikant høyere gjennomsnittsverdi for den sistnevnte variabelen. Subjekt-verb-kongruens skiller seg imidlertid ikke signifikant fra den andre syntaktiske variabelen som ble introdusert i studie 2, nemlig ordstilling i subjektsinitiale deklarativsetninger med setningsadverbial, som altså viser seg å være vanskelig å avlære. Denne variabelen er i likhet med subjekt-verb-kongruens signifikant forskjellig fra ordstilling i ikke-subjektsinitiale deklarativsetninger. Dette er problematisk for flaskehalshypotesen, som argumenterer for at lingvistisk modul er avgjørende for hva som er vanskelig i et andrespråk. Heller ikke faktorene i tabell 1 (innflytelse fra S1, lærbarhet, frekvens, synlighet og tilstedeværelse i lærematerialet) gir noen forklaring på at disse to syntaktiske variablene er signifikant forskjellige. Vi spekulerer i om vanskelighetene med å plassere verbet i relasjon til adverbet kan skyldes variasjon i engelsk ordstilling. For eksempel er V2-ordstilling grammatisk i engelsk når det finite verbet er et hjelpeverb (f.eks. *The students have always liked my lectures.*)

Videre viser studie 2 at deltakerne ikke har problemer med tempus (se figur 2). Dette er en morfologisk variabel og bør derfor i utgangspunktet være vanskeligere å tilegne seg sammenlignet med syntaks, ifølge flaskehalshypotesen. Realiseringen av preteritum ved svake verb er imidlertid svært lik på norsk og engelsk. Hypotesen understreker at de mest utfordrende (morfologiske) trekkene er de som representerer en uoverensstemmelse mellom S1 og S2, og resultatet motsier dermed ikke flaskehalshypotesens prediksjoner for forskningsspørsmål 1 og 2. Imidlertid understreker dette resultatet viktigheten av å

ta hensyn til innflytelse fra S1 når den generelle vanskelighetsgraden av et trekk i S2 skal evalueres (se også Jensen et al. 2020). En annen faktor som kan spille inn her er at preteritumsmarkøren *-ed* er mer nyttig for kommunikasjon enn samsvarsmarkøren *-s*. Fravær av preteritumsmarkøren påvirker betydning, mens fravær av samsvarsmarkøren kun påvirker grammatikken.

Når det gjelder interaksjonen mellom ferdighetsnivå og konstruksjon, viser studie 2 at subjekt-verb-kongruens er signifikant forskjellig fra tempus. Dette er den eneste signifikante forskjellen mellom variablene. Dette illustreres med en brattere kurve for preteritum i figur 2. Studie 1 viste at problemer med subjekt-verb-kongruens i størst grad reflekteres gjennom overaksept. I studie 2 testes kun underaksept (entallssubjekter). Det er derfor ikke overraskende at interaksjonen mellom ferdighetsnivå og subjekt-verb-kongruens ikke er signifikant forskjellig fra de syntaktiske variablene i studie 2.

Resultatene fra studie 1 kan sies å støtte flaskehalshypotesen av to grunner: For det første er ugrammatisk morfologi vanskeligere å identifisere enn ugrammatisk ordstilling. For det andre viser interaksjonene mellom ferdighetsnivå og konstruksjon at overaksept av *-s*-morfemet er et vedvarende problem som ikke er direkte avhengig av generell ferdighet i språket. Dette er problematisk for innlærerne til tross for flere faktorer som skulle ha gjort læringsoppgaven lettere, slik som positiv evidens for subjekt-verb-kongruens i innputtet, at *-s*-morfemet er mer frekvent enn setninger som illustrerer ikke-V2-ordstilling i engelsk og at kongruens blir eksplisitt undervist i skolen (se tabell 1).

Samtidig er det også klart at funn fra studie 2 tyder på at en konstruksjons tilhørighet til lingvistisk modul (syntaks eller morfologi) ikke kan forklare all atferd blant andrespråkstalere av engelsk med norsk som førstespråk. Studie 2 viser at det er en klar forskjell i vanskelighetsgrad mellom ulike typer av morfologiske eksponenter og ulike typer syntaktiske strukturer. I lys av funnene fra studie 1 og 2 vil vi hevde at flaskehalshypotesen kan si noe om hva som er vanskelig i et andrespråk på generell basis, men at vanskelighetsgraden blir ytterligere påvirket av mange andre faktorer.

7 Oppsummering/konklusjon

I denne artikkelen har vi presentert funn fra to studier som testet flaskehalshypotesen (Slabakova, 2008, 2013). Vi testet norske andrespråksinnlærere av engelsk i deres evne til å identifisere ugrammatisk morfologi og syntaks. Flas-

kehalshypotesens prediksjon er at morfologi er vanskeligere å tilegne seg enn syntaks. Denne prediksjonen ble delvis støttet av våre to studier.

I den første studien viste resultatene at deltakerne hadde større problemer med subjekt-verb-kongruens enn SVO-ordstilling i ikke-subjektsinitiale deklarativsetninger. Vi så at den vanligste feilen når det gjelder subjekt-verb-kongruens er overaksept av *-s*-morfemet. Resultatene fra studie 2 bekreftet disse funnene. Denne studien viste imidlertid også at morfologi er uproblematisk dersom det finnes likheter mellom første- og andrespråket (tempus) og at ordstilling i setninger med setningsadverbialer er like vanskelig som subjekt-verb-kongruens. Dette er problematisk for flaskehalshypotesen fordi ordstilling representerer syntaks. Vi konkluderte derfor at tilhørighet til lingvistisk modul, slik som for eksempel morfologi eller syntaks, ikke alene kan forklare hva som er vanskelig eller lett å tilegne seg i et andrespråk. Fremtidige studier bør undersøke hvilke andre faktorer som påvirker vanskelighetsgraden, gjerne på tvers av ulike språkkombinasjoner.

Vedlegg

Vedlegg A: Leksikalske verb i COCA (Corpus.byu.edu, 2018a)

Verbtype	Frekvens	Prosent av alle leksikalske verb i presens (N=16 535 156)	Prosent av alle leksikalske verb (N=61 188 902)
Tredje person entall	6 198 523	37,5 %	10,1 %
Ikke-tredje person entall	10 336 633	62,5 %	16,9 %

Merk: Ikke-tredje person entall = Første og andre person entall og flertall, samt tredje person flertall.

Vedlegg B: Leksikalske verb i muntlig COCA (Corpus.byu.edu, 2018a)

Verbtype	Frekvens	Prosent av alle presens leksikalske verb (N=4 042 909)	Prosent av alle leksikalske verb (N=13 175 959)
Tredje person entall	944 638	23,4 %	7,2 %
Ikke-tredje person entall	3 098 271	76,6 %	23,5 %

Merk: Ikke-tredje person entall = Første og andre person entall og flertall, samt tredje person flertall.

Vedlegg C: Frekvens og prosent av deklarativer i Pearl Sprouse-korpuset; søk utført av Charles Yang, som vi herved takker for bidraget.

Setningstype	Frekvens	Prosent av alle deklarativer (N = 46 117)
Ikke-V2-setninger	4449	9,6 %
SXV	1115	2,4 %
XSV	3114	6,8 %

Merk: Tallene inkluderer deklarativsetninger med verb (men ikke spørsmål, imperativ eller interjeksjon). Ikke-V2-setninger inkluderer både SXV og XSV.

Vedlegg D: Antall oppgaver og artikler i lærematerialet som fokuserer på subjekt-verb-kongruens og ordstilling.

Materiale	Klassetrinn	Kongruens			Ordstilling		
		Oppgaver	Artikler	Totalt	Oppgaver	Artikler	Totalt
Røkaas et al. (2011)	12	17	3	20	0	0	0
Rugset & Ulven (2011)	11	0	1	1	0	0	0
NDLA (2018b)	11–13	10	10	20	1	1	2
Solberg & Unnerud (2015a,b)	7	2	1	3	0	0	0

Referanser

- Almor, Amit, Maryellen Coles MacDonald, Daniel Kempler, Elaine S. Andersen, Lorraine K. Tyler. 2001. Comprehension of long distance number agreement in probable Alzheimer's disease. *Language and Cognitive processes* 16, 35–63.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67–1, 1–48.
- Bock, Kathryn & Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23–1, 45–93.
- Bohnacker, Ute. 2006. When Swedes begin to learn German: from V2 to V2. *Second Language Research* 22–4, 443–486.
- Brown, Roger. 1973. *A first language: the early stages*. Cambridge, MA: Harvard University Press.
- Bybee, Joan. 2000. *Mechanisms of Change in Grammaticalization: The Role of Frequency*. Albuquerque, NM: University of New Mexico.

- Cintrón-Valentín, Myrna C. & Nick C. Ellis. 2016. Salience in Second Language Acquisition: Physical Form, Learner Attention, and Instructional Focus. *Frontiers in Psychology* 7, 1–21.
- Corpus.byu.edu. 2018a. *Corpus of Contemporary American English*. Hentet fra <https://corpus.byu.edu/coca/>.
- Corpus.byu.edu. 2018b. *Word frequency data*. Hentet fra <https://www.word-frequency.info/sample.asp#simple>.
- Cunnings, Ian. 2017. Parsing and Working Memory in bilingual sentence processing. *Bilingualism: Language and Cognition* 20–4, 659–678.
- Dröschel, Yvonne. 2011. *Lingua Franca English: The Role of Simplification and Transfer*. Bern: Peter Lang.
- Ellis, Nick C. 2002. Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143–188.
- Garshol, Lenka. 2018. We Agreed to Disagree: Subject-Verb (Dis)Agreement Patterns in L2 English of Norwegian learners. *3rd Norwegian Graduate Student Conference in Linguistics and Philology: Abstracts*. Oslo: University of Oslo.
- Gabriele, Alison. 2009. Transfer and transition in the SLA of aspect. *Studies in Second Language Acquisition* 31, 371–402.
- Goldschneider, Jennifer M. & Robert M. DeKeyser. 2015. Explaining the Natural Order of L2 Morpheme Acquisition in English: A Meta-analysis of Multiple Determinants. *Language Learning*, 55(S1), 27–77.
- Haznedar, Belma. 2001. The Acquisition of the IP system in child L2 English. *Studies in second Language Acquisition* 23, 1–39.
- Inagaki, Shunji. 2006. Manner-of-motion verbs with locational/directional PPs in L2 English and Japanese. I Roumyana Slabakova, Silvina Montrul & Philippe Prévost (red.): *Inquiries in Linguistic Development*. Amsterdam: John Benjamins, 41–68.
- Ionin, Tania & Kenneth Wexler. 2002. Why is ‘is’ easier than ‘-s’? Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research* 18(2), 95–136.
- Jensen, Isabel Nadine, Roumyana Slabakova, Marit Westergaard & Björn Lundquist. 2020. The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners’ knowledge of syntax and morphology in L2 English. *Second Language Research*, 36(1), 3–29.
- Jensen, Isabel Nadine, Roumyana Slabakova & Marit Westergaard. 2017. The Bottleneck Hypothesis in L2 Acquisition: A Study of L1 Norwegian Speak-

- ers' Knowledge of Syntax and Morphology in L2 English. *Proceedings of the 41st annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- Jensen, Mirjam. 2017. *Investigating the Bottleneck Hypothesis in Second Language Acquisition: The acquisition of narrow syntax and functional morphology among Norwegian L2 learners of English* (Mastergradsavhandling). UiT Norges Arktiske Universitet, Tromsø.
- Keating, Gregory D. 2005. *Processing gender agreement across phrases in Spanish: Eye movements during sentence comprehension* (Doktorgradsavhandling). University of Illinois at Chicago, Chicago.
- Lardiere, Donna. 1998. Dissociating syntax from morphology in a divergent end-state grammar. *Second Language Research* 14, 359–375.
- Lightfoot, David. 1999. *The development of language: Acquisition, change, and evolution*. Oxford: Wiley-Blackwell.
- Mazurkewich, Irene & Lydia White. 1984. The acquisition of dative alternation: unlearning overgeneralizations. *Cognition* 16, 261–283.
- Montrul, Silvina & James Yoon. 2009. Putting parameters in their proper place. *Second Language Research* 25–2, 291–311.
- NDLA: Open Educational Resources For Secondary Schools. 2018b. *NDLA: Engelsk*. Hentet fra <https://ndla.no/en/node/42>.
- Pinheiro José & Douglas Bates. 2000. *Mixed-Effects Models in S and SPLUS*. New York: Springer.
- Pollock, Jean-Yves. 1989. Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry* 20–3, 365–424.
- Rugset, Audun & Eva Ulven. 2011. *Gateways*. Oslo: Aschehoug.
- Røkaas, Felicia, Julia Kagge, Ellen-Cathrine Kaspersen, Jeanne L. Skanke & Eva Haugum. 2011. *Worldwide*. Oslo: Aschehoug.
- Slabakova, Roumyana. 2008. *Meaning in the Second Language*. Berlin: Mouton de Gruyter.
- Slabakova, Roumyana. 2013. What is easy and what is hard to acquire in a second language: A generative perspective. I García Mayo, Maria del Pilar, Maria Junkal Gutiérrez Mangado & Maria Martínez-Adrián (red.): *Contemporary approaches to second language acquisition*. Amsterdam/Philadelphia: John Benjamins, 5–28.
- Slabakova, Roumyana. 2016. *Second Language Acquisition*. New York, NY: Oxford University Press.
- Solberg, Cecilie & Hege Dahl Unnerud. 2015a. *Stairs 7: Textbook*. Oslo: Cappelen Damm.

- Solberg, Cecilie & Hege Dahl Unnerud. 2015b. *Stairs 7: Workbook*. Oslo: Cappelen Damm.
- Trahey, Martha & Lydia White. 1993. Positive Evidence and Preemption in the Second Language Classroom. *Studies in Second Language Acquisition* 15-2, 181–204.
- Vettorel, Paola. 2014. *English as a Lingua Franca in Wider Networking: Blogging Practices*. Berlin: De Gruyter Mouton.
- Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. New York/Oxford: Oxford University Press.
- Westergaard, Marit. 2003. Unlearning V2. Transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian children. *EUROSLA Yearbook* 3, 77–101.
- White, Lydia. 2003. *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.
- Yang, Charles. 2001. Internal and external forces in language change. *Language Variation and Change* 12-3, 231–250.
- Zipf, George K. 1949. *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley.
- Åfarli, Tor A. & Kristin M. Eide. 2003. *Norsk generativ syntaks*. Oslo: Novus forlag.

Summary

The Bottleneck Hypothesis (Slabakova 2008, 2013) argues that acquisition of morphology is the most difficult part of acquiring a second language. In this article, we discuss two studies that test the Bottleneck Hypothesis, by investigating Norwegian L1 learners' ability to identify ungrammaticality in morphology and syntax in English at different levels of the acquisition process. The results show that subject-verb agreement is more difficult than word order in non-subject-initial declarative clauses, and that, despite increased general proficiency in English, the participants frequently overaccept the *-s* morpheme. However, we also see that morphology can be relatively easy for second language learners if it represents a match between the L1 and the L2 and that certain syntactic constructions can be significantly more challenging. This illustrates that, although the Bottleneck Hypothesis can say something general about what is difficult in an L2, it does not explain all behavior demonstrated by Norwegian L2 learners of English.

Isabel Nadine Jensen
Stipendiat i lingvistikk
Institutt for språk og kultur
UiT Norges Arktiske Universitet
Postboks 6050
9037 Tromsø
isabel.n.jensen@uit.no

Marit Westergaard
Professor i engelsk lingvistikk
Institutt for språk og kultur
UiT Norges Arktiske Universitet
Postboks 6050
9037 Tromsø

Institutt for språk og litteratur
NTNU Norges teknisk-naturvitenskapelige universitet
7491 Trondheim
marit.westergaard@uit.no



‘I are in New York’: om tilegnelsen av samsvarsbøyning i engelsk blant innlærere med norskspråklig bakgrunn¹

Kristin Killie

UiT Norges arktiske universitet

Artikkelen diskuterer tilegnelsen av samsvarsbøyning (mellom subjekt og verbal) ved *be* i engelsk blant 12–13- og 15–16-åringer med norsk språkbakgrunn. Det legges fram data fra the *Corpus of Young Learner Language* som tyder på at en del 12–13-åringer overgeneraliserer formen *are* i presens, mens innlærere fra begge aldersgrupper overgeneraliserer *was* i preteritum. Mens overgeneraliseringa av *are* trolig skyldes fonetisk og grafemisk likhet med den norske formen *er*, kan overgeneraliseringa av *was* skyldes denne formens høye frekvens. Det argumenteres også for at drilling, frekvens i input og output og ‘chunk-læring’ påvirker tilegnelsen av *be*-formene. Dataene støtter idéen om at tilegnelsen av *be*-paradigmet, og av samsvarsbøyning generelt, er en langsom prosess (Serratrice 2001) som utgjør en ‘flaskehals’ innen språktilegnelse (Slabakova 2019).

Nøkkelord: samsvarsbøyning, engelsk, andrespråktilegnelse, transfer, overgeneralisering

1 Innledning

Denne artikkelen tar for seg hvordan innlærere av engelsk som L2 med norsk språkbakgrunn tilegner seg samsvar mellom subjekt og verbal i engelsk, eller

1. Takk til redaktører og fagfeller for nyttige tilbakemeldinger og til the Research Group for Language Testing and Assessment for å ha gjort CORYL-korpuset gratis tilgjengelig på nett.

‘samsvarsbøyning’, for korthets skyld.² For disse er samsvarsbøyning mellom subjekt og verbal et utfordrende aspekt ved engelsk da dette fenomenet ikke eksisterer i norsk (Hauge & Angelsen 2018: 274–275). Innlærere med norsk språkbakgrunn må dermed ikke bare tilegne seg de relevante engelske verbformer, men også konseptet samsvarsbøyning, på et bevisst eller ubevisst nivå. Artikkelen fokuserer på tilegnelse av samsvarsbøyning ved verbet *be*. Her er det en rekke former med ulike stammer, og det kan virke krevende å bevege seg fra ord- til systemkunnskap. Dette kan forklare at en del norskspråklige innlærere tyr til overgeneraliseringsstrategier. Ifølge Garshol (2018: 45–46) overgeneraliserer mange 15–16-åringer *are* i presens og *were* i preteritum. I denne studien er jeg interessert i om dette også gjelder elevene i the *Corpus of Young Learner Language* (CORYL). Følgende to hypoteser er testa:

- 1) En del innlærere med norsk språkbakgrunn som lærer engelsk, generaliserer formen *are* i presens.
- 2) En del innlærere med norsk språkbakgrunn som lærer engelsk, generaliserer formen *were* i preteritum.

Artikkelen har følgende organisering: Del 2 gir relevant bakgrunnsinformasjon om tilegnelsen av samsvarsbøyning i engelsk, generelt og blant norske L2-innlærere. Del 3 redegjør for korpus og metode. Dataene presenteres så i del 4 og diskuteres i del 5, før del 6 kort oppsummerer de viktigste funnene.

2 Å tilegne seg samsvar mellom subjekt og verbal

2.1 Samsvarsbøyning i engelsk – hva må læres, og hvorfor er det så vanskelig?
Engelsk har i presens tallkongruens mellom subjektet og det finite verbet (med unntak av modalverb). I flertall brukes grunnformen av verbet (*The girls live*), mens en i entall legger til en *-s* (*The girl lives*). Når subjektet er et personlig pronomen, får man i tillegg personkongruens. Det er da kun i 3. person entall man legger til *-s*-suffikset; altså får vi *I live* og *you live*, men *he/she/it lives*.

2. Merk at termen L2 i denne artikkelen også kan bety L3 eller andre språk som læres etter L1. Jeg skiller ikke mellom tilegnelse og (inn)læring i denne artikkelen da jeg mener det i praksis er umulig å skille mellom de to prosessene i en norsk kontekst, der innlærere får så mye input utenfor klasserommet. Termen *engelsk* er å forstå som ‘standardengelsk’ da regionale varianter av engelsk kan ha helt andre systemer (Cheshire & Milroy 1993, DeBose 2015; McCafferty 2004; Oancea 2017; Wolfram 2004).

Verbet *be* har det mest kompliserte paradigmet. Her brukes ulike verbstammer til å signalisere en forskjell i person og tall, såkalt suppletivisme: *I am, you are, he/she/it is, we are, you are, they are*. *Be* skiller seg også ut ved å kreve samsvarsbøyning i preteritum, der formen *was* brukes i 1. og 3. person entall, mens *were* brukes ellers.

Samsvarsbøyning er utfordrende å tilegne seg i alle språk, og dette gjelder både første- og andrespråkstilegnelse (Breiteneder 2005, Brown 1973, Dröschel 2011, Dulay & Burt 1974, Fisher 1985, Haznedar 2001, Housen 2002, Jensen, Slabakova, Westergaard & Sundquist 2020, Keeney & Wolfe 1972, Radford 1990, Slabakova & Gajdos 2008, Slabakova 2016, 2019, White 2003). Ifølge Slabakova (2016: 397) er dette fordi samsvarsbøyning tilhører den komplekse språkkomponenten funksjonell morfologi, som sies å utgjøre selve ‘flaskehalsen’ (the ‘bottleneck’) i språktilegnelse (Slabakova 2016: 389–403, 2019). Den funksjonelle morfologien inneholder informasjon om grammatiske kategorier som tempus, aspekt, grammatisk tall og kjønn, kasus osv. Når en lærer et nytt språk, må en tilegne seg både generell og spesifikk informasjon om hvordan de ulike grammatiske kategoriene realiseres i det aktuelle språket (Lardiere 2009, Slabakova 2016: 175). Dette er langt mer komplisert enn å tilegne seg leksikalske uttrykk og syntaktiske kunnskaper som for eksempel grunnleggende ordstilling (Jensen et al. 2020, Slabakova 2019). Når den funksjonelle morfologien skal læres, må språkinnlæreren foreta en utvelgelse (‘selection’) av hvilke grammatiske kategorier og trekk som er relevante for det aktuelle språket (f.eks. [Person: {1, 2, 3}], [Number {Sg, Pl}]), og finne ut av hvordan de ulike trekkene kombineres med hverandre og med de leksikalske elementene (‘configuration’) (Lardiere 2009). Hvert morfem kan inneholde et knippe med ulike grammatiske trekk, men sammensetningen av de ulike knippene varierer fra språk til språk. Når det gjelder engelsk samsvarsmorfologi på verb, uttrykker de aktuelle morfemene tempus, aspekt, modus, person og tall. Bøyningsmorfemene som brukes på norske verb i presens, uttrykker derimot ikke person og tall. Altså er det bare delvis semantisk overlapp mellom de relevante morfemene i de to språkene. Dermed må norske innlærere som lærer engelsk, foreta det Lardiere (2009) kaller rekonfigurasjon (‘reconfiguration’ eller ‘feature re-assembly’). Dette er en mer krevende oppgave enn å starte med blanke ark og lære et nytt trekk som er uttrykt gjennom egne morfemer (Ionin & Montrul 2010: 907, Sprouse 2006).

I tillegg til den funksjonelle grammatikkens kompleksitet, vil tilegnelsen av samsvarsbøyning i L2 kunne lide under utilstrekkelig input og output. At input er viktig i språktilegnelse, er velkjent (Krashen 1985), og det samme gjelder

output og interaksjon (Swain 1985, 2000). Input og output må forekomme i store nok doser til at det som skal læres, har en viss frekvens (jf. artiklene i Collins & Ellis 2009). Dette gjelder også ved tilegnelsen av funksjonell morfologi. Den senere tids studier av funksjonell morfologi i arvespråk (minoritetsspråk som kun læres og brukes hjemme) avdekker for eksempel vansker med tilegnelsen av funksjonell morfologi, noe som blant annet tilskrives lav frekvens (Slabakova 2016: 163–4, 398). Gitt alt dette er det ikke usannsynlig at lav frekvens i input og output vil kunne ha en negativ effekt på tilegnelsen av samsvarsbøyning i engelsk også blant innlærere med norsk språkbakgrunn. I tillegg kan lav frekvens resultere i prosesseringsvansker også av tilegnede strukturer fordi tilgangen til all den relevante informasjonen ikke er tilstrekkelig automatisert (Slabakova 2016: 398f.). Prosesseringsvanskene vil være størst under høyt tidspress, men vil sannsynligvis kunne spille inn også under skrivning av et andrespråk.

Norske innlærere må altså, på et bevisst eller ubevisst plan, opparbeide seg en forståelse av hva samsvar mellom subjekt og verbal går ut på, og de må lære hvordan dette fenomenet kommer til uttrykk i engelsk. Sistnevnte oppgave innebærer, som vi har sett, blant annet å tilegne seg kunnskaper om ulike verbformer. Dette er enda vanskeligere i et språk som engelsk, der samsvarsbøyning på leksikalske verb kun markeres eksplisitt i 3. person entall, gjennom endelsen *-s*, mens andre person/tall-kombinasjoner sammenfaller med stammen/infinitiven. Dette gjør inputen lite konsistent og dermed lite informativ (Serratrice 2001: 49). I tillegg kommer utfordringene med *be*-paradigmets irregularet.

Valget av riktig verbform betinger også at man besitter ulike typer syntaktiske, semantiske og leksikalske kunnskaper. Man må vite at verbets form normalt bestemmes av kjernen i subjektsfrasen, men at det i noen tilfeller er subjektets mening som er avgjørende, som i (3) under. I tillegg må man ha kunnskaper om tallegenskapene til ulike substantiver og subjekter, for eksempel ubestemte pronomen (4) og utelleglige substantiver (5) og flertalls-substantiver (6) som tilhører en annen tallkategori i ens eget språk osv.

- 3) Gin and tonic *is* a classic drink. (Dypedahl & Hasselgård 2018: 119)
- 4) Everyone *is* hungry; let's eat. (Dypedahl & Hasselgård 2018: 115) ('alle')
- 5) The money *was* not there. (Dypedahl & Hasselgård 2018: 32) ('pengene')
- 6) The police *are* looking for you. (Dypedahl & Hasselgård 2018: 117) ('politiet')

Skal innlærere lykkes med samsvarsbøyning, må de også forstå forskjellen på standard- og substandardengelsk og klare å holde de to typene varieteter fra hverandre ettersom de ulike formene av *be* brukes på andre måter utenfor standardvarietetene (jf. fotnote 2). Dette kan man anta er ei utfordring med den massive eksponeringa dagens unge har for substandardspråk, gjennom film og serier, dataspill, musikk og annen populærkultur (Brevik 2019, Preisler 1999: 242, 244ff).

2.2 Relevant forskning

Killies (2019) data tyder på at L2-innlærere av engelsk med norsk språkbakgrunn gjør langt flere samsvarsfeil når subjektets kjerne er et substantiv enn når den er et personlig pronomen. Killies korpus består av alle samsvarsbøyde former av *be*, *have* og *do* i CORYL. 12–13-åringene bruker riktig verbform ved personlige pronomen i hele 90,4 % av tilfellene, mens treffprosenten kun er 78,1 % når subjektets kjerne er et substantiv. Dette forklares som et resultat av utstrakt drilling av verbparadigmene for *be*, *have* og *do* sammen med personlige pronomen (se også Tan 2005). I tillegg spiller trolig høy frekvens i input og output en rolle: Personlige pronomen opptrer svært hyppig som subjekt, både i engelsk og i andre språk. I Killies (2019) korpus er 58,5 % av 12–13-åringenes subjekter personlige pronomen, mens 30,5 % er substantiver. Tilsvarende tall for 15–16-åringene er 53,6 % og 31,3 %. Garshol (2018) finner at også 15–16-åringene i hennes korpus gjør langt flere feil når subjektet er et substantiv enn når det er et personlig pronomen, og ifølge Slabakova og Gajdos (2008) finner man det samme mønsteret blant innlærere med tysk språkbakgrunn. Forskning viser videre at problemene øker når subjektet blir mer komplekst. Dette gjelder også for førstespråkbrukere av engelsk og skyldes nok i stor grad prosesseringsproblemer (for en diskusjon, se Killie 2019).

Studier som fokuserer på verbets form, har avdekket ulike generaliseringsstrategier. Den vanligste nybegynnerstrategien, uavhengig av språkbakgrunn, er å utelate tredje persons *-s*'en i presens (Austin 2012; Prévost & White 2000; White 2003: 183).³ Dette gjør også innlærere med norsk språkbakgrunn, men Garshol (2018) og Killie (under vurdering) finner at en del elever på ungdomstrinnet og i videregående skole overgeneraliserer tredje persons *-s*'en i presens, altså bruker den i entall og flertall og ved alle personer, som vist i (7) og (8).

(7) All students *starts* at the same level. (Garshol 2018: 44)

3. Dette er vanlig også i førstespråktilegnelse (Wexler 1994).

(8) I *does* this all the time. (Garshol 2018: 44)

At en del norske innlærere av engelsk overgeneraliserer 3. persons *-s-en*, støttes av elisitasjonsdataene i Jensen et al. (2020), som viser at norskspråklige 12–19-åringene ofte godtar ugrammatiske setninger der 3. persons *-s-en* brukes ved flertallssubjekter, som i (7) over.

Her skal vi imidlertid ikke ta for oss overgeneralisering av *-s-* eller null-suffikset ved leksikalske verb, men hvordan norske innlærere uttrykker samsvar ved *be*, altså gjennom suppletivisme. Det har blitt hevdet at L2-innlærere generelt gjør langt færre samsvarsfeil ved *be* enn ved andre verb (Ionin & Wexler 2002, White 2003). Dette har blitt knytta til ulike aspekter ved verbet *be*, som tidlig tilegnelse, høy frekvens, løfting av hjelpeverb ut av VP og *be*-formenes suppletive karakter (Fraser, Bellugi & Brown 1963, Keeney & Wolfe 1972, White 1992, 2003). Ifølge Keeney & Wolfe (1972: 699) skyldes den mål-språksnære produksjonen ved *be* en kombinasjon av høy frekvens og manglende interferens fra substantivendelser. Sistnevnte antas å forekomme ved leksikalske verb, der det nominale flertallssuffikset *-s* potensielt blir forbundet med flertall også ved verb. White (1992), derimot, forbinder samsvarsbøyning med verbløfting i de tidlige fasene av L2-innlæringa. Det faktum at innlærere generelt samsvarsbøyer *be*, men ikke leksikalske verb, tilskrives at *be* fungerer som hjelpeverb, og at det som hjelpeverb løftes ut av VP og opp til I. Ettersom dette ikke skjer ved leksikalske verb, blir ikke disse samsvarsbøyd.

Serratrice (2001) finner imidlertid i sin casestudie av tospråklige Carlo (engelsk/italiensk) at tilegnelsen av *be*-formene går langsomt de første årene. Når Carlo begynner å lage enkle setninger som krever *be*, utelater han først verbet, som er vanlig i tidlig setningsbygging (Brown 1973). Først i alderen 2;4.7 år begynner han å bruke kopula *be* (hjelpeverbet *be* introduseres noe senere), men da kun den sammentrukne formen *it's* (*a*). Serratrice argumenterer for at det her nok ikke dreier det seg om noen systemisk tilegnelse av samsvarsbøyning, men heller om tilegnelse av en 'chunk' som barnet trenger for å navngi objekter (en aktivitet barn bruker mye av tida si på). Som vi skal se, kan slik chunk-læring spille en rolle også i andrespråkstilegnelse.

Garshol (2018) studerer tilegnelsen av samsvarsbøyning blant 119 15–16-åringene med norsk L1 ved tre ulike videregående skoler i Agder. Hun finner at elevene i materialet overgeneraliserer *are* i presens (*She are upset*) og *were* i preteritum (... *his brother were killed too*) og forklarer dette med at formene likner de norske presens- og preteritumsformene. Hun finner videre at elevene oftere gjør feil i preteritum enn i presens (2019: 44–45, 58, 63–64).

3 Korpus og noen metodologiske betraktninger

Dataene i denne studien tatt fra CORYL-korpuset. Korpuset består av et tilfeldig utvalg av tekster fra de nasjonale prøvene i engelsk i årene 2004 og 2005. Tekstene er skrevet av elever i 7. og 10. klasse fra hele landet. Det foreligger ikke opplysninger om elevenes førstespråk, men det antas at de fleste av dem har norsk som sitt første- eller andrespråk (Hasselgren & Sundet 2017: 199). De fleste vil da sannsynligvis ha engelsk som L2 eller L3. I alt inneholder korpuset 272 elevtekster, 136 fra hver av de to aldersgruppene 12–13 år (7. klasse) og 15–16 år (10. klasse), til sammen 129 421 ord. Tekstene er markert for alder, skoletrinn, ferdighetsnivå (CEFR) og kjønn og er kodet for ulike typer feil, som samsvarsfeil, preposisjonsfeil, stavfeil osv. (se Hasselgren & Sundet 2017 for mer informasjon). Ettersom jeg i denne studien ikke kun er interessert i elevenes ukorrekte anvendelse av *be*, men også i de korrekte brukene, har jeg ikke benyttet meg av feilkodinga, men har isteden utført ordsøk. Søkene inkluderte alle korrekte og ukorrekte former av *be*, inkludert negerte former (*aren't*, *wasn't* osv.).

Korpuseksempelene ble kategorisert etter subjektets type: setninger, substantiver og ulike typer pronomen: personlige, demonstrative, ubestemte, interrogative, relative og eiendomspronomen. Det var også en blandet type bestående av minst to kjerner der minst én var et substantiv og minst én et pronomen. Setningene med substantivisk subjekt ble videre klassifisert for type substantiv: tellelig, utellelig eller flertallssubstantiv. I tillegg ble eksemplene klassifisert som korrekt eller ukorrekt.

Dataene kan vanskelig testes for statistisk signifikans da det ikke foreligger systematisk koding av alle verbformer på individnivå. Korpuset er da heller ikke lagd for signifikanstesting (Hasselgren & Sundet 2017: 199). Man kan imidlertid få tilgang til data på individnivå ved å klikke på et treff i konkordansen og *be* om kontekst. Da får en se hele teksten til den aktuelle eleven. Med flere tusen verbformer er det vanskelig å undersøke alle på denne måten. Det man imidlertid fort ser når man klikker seg inn i elevtekstene, er hvor ulike de er, både når det gjelder omfang og innhold. En del elever har skrevet veldig lite tekst og har kun brukt verbet *be* noen få ganger, mens andre har produsert ganske mye tekst og også mange *be*-former.⁴ Det at korpuset ikke er egna for signifikanstesting, betyr imidlertid ikke at det ikke byr på interessante data.

4. I oppgave (1) for 7. trinn gis for eksempel elevene følgende instruksjon: "Look at the picture. What do you see?" (Lie, Hopfenbeck, Ibsen & Turmo 2005: 87). Her har en del elever svart med fulle setninger, mens andre kun har gitt ei liste med nominalfraser og har dermed altså ikke brukt verb i det hele tatt.

Dataene viser en del klare tendenser, og en kan kombinere disse kvantitative dataene med mer kvalitative dyddykk. Som vi skal se i del 4, har jeg, ved å undersøke de ukorrekte formene til 12–13-åringene, kommet frem til et mål på spredning, altså hvordan feilene er fordelt på antall elever. Jeg har også sett nærmere på tekstene til enkeltelever som har produsert mange *be*-former for å vurdere om de overgeneraliserer enkeltformer.

4 Presentasjon av dataene

4.1 Overgeneralisering av *ARE* i presens

Tabell 1 viser korrekte og ukorrekte pronomener-verb-kombinasjoner ved *be* i presens på gruppenivå. Tabellen viser både absolutt og relativ andel forekomster av den aktuelle pronomener-verb-kombinasjonen i forhold til totalt antall av den relevante tall- og personkombinasjonen.

Tabell 1: Kombinasjoner av personlige pronomener og presensformene av BE

	12–13-ÅRINGER	15–16-ÅRINGER
1. PERSON ENTALL		
I am	166/201 (82,6 %)	52/53 (98,1 %)
I are	30/201 (14,9 %)	1/53 (1,9 %)
I is	5/201 (2,5 %)	---
ukorrekte former, total	35/201 (17,4 %)	1/53 (1,9 %)
2. PERSON ENTALL OG FLERTALL		
you are	55/55 (100 %)	111/111 (100 %)
you is (ukorrekte former)	---	---
3. PERSON ENTALL, <i>IT</i>		
it is	207/220 (94,1 %)	132/135 (97,8 %)
it are (ukorrekte former)	13/220 (5,9 %)	3/135 (2,2 %)
3. PERSON ENTALL, <i>HE OG SHE</i>		
s/he is	44/56 (78,6 %)	40/42 (95,2 %)
she are (ukorrekte former)	12/56 (21,4 %)	2/42 (4,8 %)
1. PERSON FLERTALL		
we are	99/100 (99 %)	49/50 (98 %)
we be	---	1/50 (2 %)
we is	1/100 (1 %)	---
ukorrekte former, total	1/100 (1 %)	1/50 (2 %)
3. PERSON FLERTALL		
they are	33/36 (91,7 %)	68/69 (98,6 %)
they is (ukorrekte former)	3/36 (8,3 %)	1/69 (1,4 %)

I tabellen er 3. person entall delt i to kategorier: *he/she* og *it*. *It* er holdt for seg selv fordi dette pronomenet er mye mer frekvent enn *he* og *she*, og det er ifølge Biber, Johansson, Leech, Conrad & Finegan (1999: 334) også det eneste personlige pronomenet som er frekvent i både formelle og uformelle sjangre. *You are*-sekvensene er ikke delt inn i entalls- og flertallsforekomster da elevene ikke gjør en eneste feil her og trolig ikke skiller mellom entall og flertall.

Vi ser at feilratene er lave i presens ved personlige pronomen. De er også gjennomgående lavere for 15–16-åringene enn for 12–13-åringene, noe som tyder på at en del elever får kontroll på pronomen-verb-sekvensene på ungdomstrinnet. Vi kan merke oss at det ikke gjøres en eneste feil ved *you* i presens, og at det gjøres kun to feil ved *we* og fire ved *they*. Treffprosenten er videre høy ved *it*, mens de yngste elevene gjør en del feil ved *s/he* og *I*. Dersom drilling er den eneste faktoren som påvirker innlæringa av *be*-formene, skulle man forvente at andelen ukorrekte former var jevnere fordelt over pronomenene. Det er derfor sannsynlig at også andre årsaker spiller inn. Som man kan utlede av Tabell 1 over, forekommer formen *are* ved hele 85,9 % av feilene (55 av 64 feil) som 12–13-åringene gjør i presens ved personlige pronomen. Til sammenlikning involverer kun 14,1 % av feilene formen *is* (ni tilfeller), mens formen *am* ikke brukes feil en eneste gang. Ei slik stor overvekt av *are* tyder på at en del elever overgeneraliserer denne formen. *Are* er brukt sammen med alle personlige pronomen.

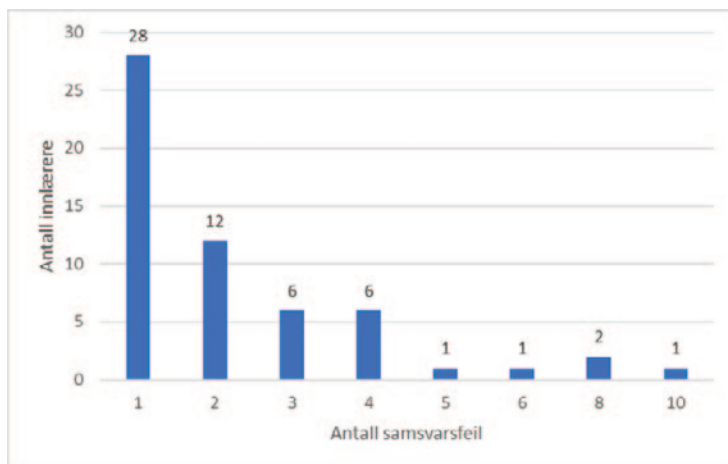
Dersom *are* overgeneraliseres av elevene, må en regne med at dette ikke kun gjøres ved personlige pronomen, men også ved andre typer subjekter. Tabell 2 viser feilratene ved *are* og *is* ved de to mest frekvente subjektstypene: personlige pronomen og substantiver.

Tabell 2: Feilrater for ARE og IS ved personlig pronomen og substantiv som subjekt

	12–13-ÅRINGER		15–16-ÅRINGER	
	pers. pronomen	Substantiv	pers. pronomen	substantiv
<i>are</i>	55/242 (22,7 %)	62/115 (53,9 %)	6/234 (2,6 %)	15/127 (11,8 %)
<i>is</i>	9/260 (3,5 %)	29/371 (7,8 %)	1/173 (0,6 %)	28/266 (10,5 %)

Som vi ser, gjør begge aldersgrupper langt flere feil når subjektet er et substantiv enn når det er et personlig pronomen. Spesielt gjør 12–13-åringene en god del feil ved *are*. For å kunne vurdere i hvilken grad disse feilene er uttrykk for en generell overgeneraliseringstendens i aldersgruppa, eller om det er snakk om

noen få individers språkbruk, trenger vi informasjon om hvordan de 117 ukorrekte formene av *are* fordeler seg på 12–13-åringene i CORYL. De aktuelle formene er distribuert over 57 (av 136) elever som vist i Figur 1.



Figur 1: Distribusjonen av ukorrekte ARE-former blant 12–13-åringene

Vi ser at halvparten av elevene (28 av 57) som har brukt *are* feil, kun har produsert én ukorrekt form. Her finnes det selvsagt ikke grunnlag for å hevde at det foreligger noen overgeneraliseringsstrategi. Imidlertid bruker den andre halvparten av elevene *are* feil mer enn én gang, og noen gjør det en rekke ganger. Dette kan tyde på at *are* generaliseres av en del innlærere, men det er ikke ukomplisert å avgjøre om dette er tilfellet. Antall feil er ikke den eneste relevante faktoren da også korrekte anvendelser av *are* kan være et resultat av generalisering. I tillegg vil en utbredt bruk av *is* blant de samme innlærerne tale mot at de overgeneraliserer *are*. Jeg har derfor analysert den totale presensbruken til de 12–13-åringene som har brukt *are* feil mer enn én gang. Dessverre er, som nevnt over, mange av tekstene i CORYL veldig korte, eller de inneholder så få presensformer av *be* at det er vanskelig å si noe om grammatikken til de aktuelle elevene på dette punktet. Tabell 3 oppgir imidlertid 17 kandidater for overgeneralisering av *are*, basert på følgende kriterier: (1) minst fem forekomster av *are*, der minst to er ukorrekt, eller minst tre ukorrekte forekomster av *are*, uavhengig av det totale antallet *are*-forekomster, kombinert

med (2) ingen overgeneralisering av *is*.⁵ I tabellen er kandidatene rangert primært etter antall ukorrekte forekomster av *are* og sekundært etter antall *are*-forekomster totalt.⁶

Tabell 3: Kandidater som trolig overgeneraliserer ARE

elev-ID	ukorrekte <i>are</i>	korrekte <i>are</i>	<i>are</i> , total	<i>is</i> , total
p42-7	10	4	14	---
ingen ID-1	10	---	10	---
ingen ID-2	9	7	16	4
ingen ID-3	7	4	11	1
p68-7	5	1	6	2
p215-7	4	14	18	3
ingen ID-4	4	8	12	1
ingen ID-5	4	3	7	1
p183-07	4	2	6	---
p75-7	4	1	5	1
p148-07	3	3	6	2
ingen ID-6	3	2	5	---
p44-7	3	1	4	---
p281-07	3	1	4	2
ingen ID-7	3	---	3	---
p97-7	2	3	5	1
ingen ID-8	2	3	5	2

Som vi ser, har 12 elever produsert flere ukorrekte enn korrekte *are*-former. Fire elever har flere korrekte enn ukorrekte forekomster av *are*, men det at de har brukt formen ukorrekt flere ganger, tyder likevel på at det foreligger overgeneralisering, og at hvor godt de treffer, beror på flaks. Én elev har produsert like mange korrekte og ukorrekte *are*-former. Kandidat p215-7 skiller seg ut ved å ha brukt *are* korrekt 3,5 ganger så ofte som han/hun har brukt formen

-
- For eksempel så har elev p248-7 produsert hele ni forekomster av *are* – fire ukorrekte og 5 korrekte. En kan likevel ikke hevde at denne eleven overgeneraliserer *are* da han/hun også bruker *is* hele 14 ganger, alle gangene korrekt og med en rekke ulike subjektstyper.
 - Tallene inneholder også forekomster av *are* med andre typer subjekter, som relativpronomener og ubestemte pronomener, subjektstyper som ikke ellers er diskutert i denne artikkelen (men se Killie 2019). Som det fremgår av tabellen, var ikke ID-nummeret til alle elevene tilgjengelig.

ukorrekt, men også han/hun har brukt *are* ukorrekt flere ganger. Alle forekomstene av *is* i Tabell 3 er korrekte, men det dreier seg stort sett om *it is/it's*, som sannsynligvis er innlært som en fast kombinasjon på grunnlag av høy frekvens, som ved førstespråkstilegnelse (jf. del 2.2). Det er kun tre eksempler på *is* med substantivisk subjekt. (Ett av dem er *my name is* i eksempel 9 under.) Mens det altså er 17 elever som potensielt overgeneraliserer *are*, gir de samme kriteriene kun fire kandidater for overgeneralisering av *is*.

Et klart eksempel på overgeneralisering av *are* er gitt i (9):

- (9) I can See a ball on the left Side of the tv. The *are* three Pichur back of the man, witt a Chikita in his hand. One baby Some play withe his moter. And a red dog Some Loking at the tv.
Heyy.. I *are* inn Hollywodd. The *are* werry manny intrasting peapol hear. I *are* going in the City and Shoping fottball-Shoos. I like looking at fottball Shoops.. I hate driving hear.. greeting me. Jack
Hello my name *is* Jack. and may two friends Adam and Peter. The *are* freadag We *are* gowing to bild a littel house. We *are* Starting to morow. I *are* going to a tre in my hage. and One min Laiter kome Adam and Peter. 7 ower layter *are* we finnist we go to my fader and spørr om he has got Some green maling. And he Say yes but you *are* doing thet to morrow. The next day. We Startet to male, And we hang opp one pirrat flagg. Peter tok his meal now. thet was pitZZa.. thet koming a Snake, deady kome and helpt oss. 3 ower layter we *are* finnist.. We *are* Sleapind there to nayt.. finniSt..
(CORYL, 12–13-åring)

Tallene i Tabell 1 gir inntrykk av at elevene har god kontroll på hvilken verbform som skal brukes ved pronomenene *you*, *we* og *they*, men hvis *are* overgeneraliseres av en del 12–13-åringer, blir tolkninga av tallene usikker. Kanskje dekker den høye treffprosenten ved *you*, *we* og *they* over en *mangel* på kunnskap som resulterer i overgeneralisering av én verbform – *are*?

I sum kan det se ut som om mange elevers produksjon av presensformer er et resultat av ‘chunking’ (som sannsynligvis skyldes drilling og frekvens i input og output, jf. del 2.2), samtidig som det finnes en default-form, nemlig *are*. Dette kan illustreres ved utdraget i (10).

- (10) But I, I *am* djust siting here and *are* eating on a big pizza and *are* drinking on a big coca cola. When I *am* siting here and *are* eating and drink-

ing. I *am* dreiming that a big orange and yellow snake *are* clambing upp the three. (CORYL, 12–13-åring)

Innlæreren her har tydeligvis lært seg chunken *I am*, men en ‘reservestrategi’ (default), nemlig overgeneralisering av *are*, trer inn når andre elementer kommer mellom pronomenet og verbet.

4.2 Hypotese 2: Overgeneralisering av WERE i preteritum

Tabell 4 viser bruken av *be* i preteritum ved personlige pronomen. Som i Tabell 1 er tallene på gruppenivå og viser antall forekomster av de ulike pronomen-verb-kombinasjonene, etterfulgt av prosentandelen i parentes.

Tabell 4: Kombinasjoner av personlige pronomen og preteritumsformene av BE

	12–13-ÅRINGER	15–16-ÅRINGER
1. PERSON ENTALL		
I was	149/156 (95,5 %)	214/226 (94,7 %)
I were (ukorrekte former)	7/156 (4,5 %)	12/226 (5,3 %)
2. PERSON ENTALL AND FLERTALL		
you were	3/3 (100 %)	6/8 (75 %)
you was (ukorrekte former)	---	2/8 (25 %)
3. PERSON ENTALL, IT		
it was	268/271 (98,9 %)	185/187 (98,9 %)
it were (ukorrekte former)	3/271 (1,1 %)	2/187 (1,1 %)
3. PERSON ENTALL, HE OG SHE		
s/he was	76/85 (89,4 %)	87/92 (94,6 %)
s/he were (ukorrekte former)	9/85 (10,6 %)	5/92 (5,4 %)
1. PERSON FLERTALL		
we were	42/74 (56,8 %)	21/30 (70 %)
we was (ukorrekte former)	32/74 (43,2 %)	9/30 (30 %)
3. PERSON FLERTALL		
they were	10/28 (35,7 %)	14/23 (60,9 %)
they was (ukorrekte former)	18/28 (64,3 %)	9/23 (39,1 %)

Feilraten er lav ved *it* og *I* og også ved *he* og *she*, men høy ved *we* og *they*, hvor en del elever ser ut til å generalisere formen *was*. Det er vanskelig å si noe sikkert om verbbruken ved *you* pga. få belegg. Hvis vi sammenlikner Tabell 4 med Tabell 1, ser vi at det gjøres mer feil i preteritum enn i presens. Som nevnt

over gjelder dette også Garshols 15–16-åringer. I Tabell 5 sammenliknes bruken av *was* og *were* ved personlige pronomener og substantiver.

Tabell 5: Feilrater for *WAS* og *WERE* ved pronominale og substantiviske subjekter

	12–13-ÅRINGER		15–16-ÅRINGER	
	pers. pronomen	Substantiv	pers. pronomen	Substantiv
<i>was</i>	50/543 (9,2 %)	51/291 (17,5 %)	20/506 (4 %)	23/270 (8,5 %)
<i>were</i>	19/74 (25,7 %)	18/31 (58,1 %)	19/60 (31,7 %)	17/36 (47,2 %)

Som i presens er andelen feil langt høyere ved substantiviske enn ved pronominale subjekter. Videre er andelen ukorrekte former atskillig høyere ved *were* enn ved *was*. Tallene kan gi inntrykk av at det først og fremst er *were* som generaliseres. Dette er uventet siden vi har sett at *was* i så stor grad brukes ved flertallspronomen, der elevene ifølge Tabell 4 har størst problemer med å velge riktig *be*-form. Den høye feilraten ved *were* i Tabell 5 skyldes trolig at *were* generaliseres av noen innlærere (som nevnt under), i kombinasjon med et lavt antall *were*-forekomster. En kan argumentere for at dersom *were* ble overgeneralisert i stor grad, burde antallet *were*-forekomster ha vært langt høyere. Tabell 6 viser andelen ukorrekte *was*-former i *were*-kontekster og omvendt.

Tabell 6: Feilrater i kontekster som krever *WAS* og *WERE*

	12–13-ÅRINGER		15–16-ÅRINGER	
	pers. pronomen	Substantiv	pers. pronomen	Substantiv
<i>was</i>	50/543 (9,2 %)	51/291 (17,5 %)	20/506 (4 %)	23/270 (8,5 %)
<i>were</i>	19/74 (25,7 %)	18/31 (58,1 %)	19/60 (31,7 %)	17/36 (47,2 %)

Tabellen viser en utstrakt bruk av *was* i *were*-kontekster, men ikke omvendt. Og selv om feilraten er lavest for 15–16-åringene, er den fortsatt svært høy i den eldste aldersgruppa sammenlikna med tallene for presensformene (jf. Tabell 1 og 2). De 101 ukorrekte *was*-formene blant 12–13-åringene er fordelt på 68 av totalt 136 elever. Kriteriene som ble etablert i del 4.1, gir grunnlag for å hevde at 24 av disse elevene generaliserer *was* (av disse 24 har faktisk bare fire elever brukt *were* i det hele tatt). Til sammenlikning er de 37 ukorrekte *were*-formene blant 12–13-åringene fordelt på 17 elever, hvorav kun fire ser ut til å generalisere denne verbformen. Tallene kan tyde på at det er vanligere å

generalisere *was* i preteritum enn å generalisere *were*. Dette er altså den motsatte konklusjonen av den Garshol (2018) trekker på basis av sitt materiale. Kan hende skyldes forskjellen her at generaliseringsstrategier varierer ganske mye. Det er imidlertid vanskelig å vurdere CORYL-dataene opp mot Garshols data ettersom det ikke foreligger tall på hvordan de 204 ukorrekte *were*-formene og de 177 ukorrekte *was*-formene i hennes materiale er fordelt over hennes 199 elever.

Tekstutdraget i (11) viser hvordan en elev overgeneraliserer formen *was* i preteritum:

- (11) One day when i came home from shool, I found the front door wide open. I think with the same that my brother has come home, but on the second time ai think he don't mess so match. I went to the phone to call my mam but the phone *was* broken I go to the silver to se wats has happening with that. But it *was* ther. But many thing *was* stoled. I tok the bike and go to mum at the work. She comes with ones. He cald Jon and Per with the mobile-phone. They *was* on the case. A week later the thifs *was* found. It *was* a lady from my mum old, old clas. And a nother man. They had sold the camera to the man over oss in the house with many Fylle??? pople. (CORYL, 12–13-åring)

Før vi går videre, bør det påpekes at de ukorrekte *be*-formene i denne studien i svært liten grad kan forklares som manglende kontroll på substantiv- og subjektskategorier som de som er illustrert i (3)–(6) over. Korpuset inneholder noen eksempler på ukorrekt bruk av flertallsverb ved ubestemte pronomen (som i **everybody are hungry*) og entallsverb ved flertallssubstantiver (som **the police was* og **(the) people is/was*), men det foreligger helt klart et generelt problem med å velge riktig verbform ved substantiver.

5 Diskusjon

Denne studien har testa følgende hypoteser:

- 1) En del innlærere med norsk språkbakgrunn som lærer engelsk, generaliserer formen *are* i presens.
- 2) En del innlærere med norsk språkbakgrunn som lærer engelsk, generaliserer formen *were* i preteritum.

Hypotese 1 ser ut til å stemme, men da først og fremst for aldersgruppa 12–13 år. I preteritum ser det derimot ut til å være mer vanlig å overgeneralisere *was* enn *were*, i begge aldersgrupper. I denne delen skal vi se på noen mulige forklaringer for disse hovedfunnene.

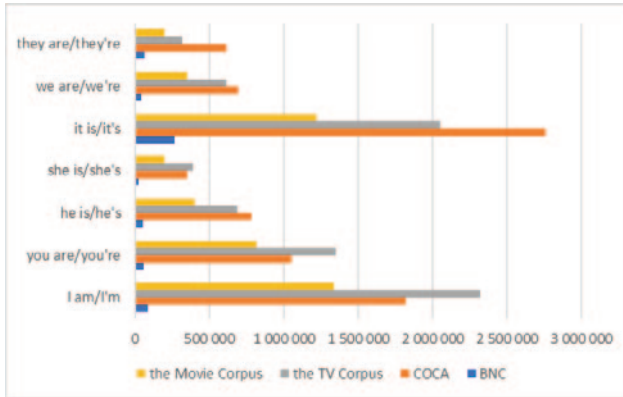
5.1 BE i presens

Overgeneraliseringa av *are* blant de yngste elevene skyldes trolig påvirkning fra norsk. I skriftspråket er *are* den engelske presensformen som er mest lik den (eneste) norske presensformen *er* ettersom begge formene inneholder en vokal etterfulgt av <r>. Når det gjelder uttale, vil det foreligge lydlikhet mellom den norske og den engelske formen for noen grupper av innlærere, men ikke for andre, avhengig av hvordan presensformene uttales i ens egen norske varietet og i den varietetet av engelsk som brukes av innlæreren, og som er dominerende i hennes input. Ifølge Rindal (2010) og Rindal & Piercy (2013) er amerikansk engelsk (General American) den varietetet som dominerer blant norske ungdommer, noe som sannsynligvis skyldes stor eksponering for denne varietetet. I amerikansk engelsk uttales r-en, og det samme gjelder den presensformen av norsk som brukes i talemålene i de mest folkerike områdene sørøst i landet. For en stor andel av norske ungdommer som lærer engelsk, vil det dermed foreligge en likhet mellom *er* og *are* i både skrift- og talespråket. Dette kan få en del innlærere til å konkludere med at *are* er en generell presensform i engelsk, akkurat som *er* er i norsk. En liknende hypotese fremsettes av Garshol (2018: 45–46).⁷

Vi har imidlertid sett at det er forskjeller på treff- og feilratene ved de ulike personlige pronomenene, og her kan det være at frekvens spiller inn. Som forklart over har frekvens sannsynligvis stor innvirkning på læring og automatisering av språkstrukturer. Figur 2 viser hyppigheten av de ulike pronomen–verb-kombinasjonene i fire ulike engelskspråklige korpus.⁸

7. Merk at denne hypotesen betinger at det som teller, er *at r-en* uttales, ikke *hvordan* den uttales, da den amerikanske r-en jo har en annen fonetisk realisasjon enn r-en i norske varieteter.

8. De undersøkte korpusene er the British National Corpus (BNC, 100 millioner ord, ulike sjangre), the Corpus of Contemporary American English (COCA, 560 millioner ord, ulike sjangre), the TV Corpus (325 millioner ord) og the Movie Corpus (200 millioner ord). Korpusene er tilgjengelige på <https://www.english-corpora.org>



Figur 2: Frekvensene av personlige pronomener + BE i presens i fire engelskspråklige korpus

It is/it's er den mest frekvente kombinasjonen av personlig pronomen og *be* i de fire korpusene, med 6 291 899 forekomster. Kombinasjonen er også den mest frekvente i CORYL (Tabell 1). Høy frekvens i input og i innlæreres output forklarer trolig hvorfor koblinga mellom *it* og *is* er så tett at den sjelden brytes, verken av L1- eller L2-innlærere, og heller ikke av innlærere som overgeneraliserer *are*. Her dreier det seg sannsynligvis, for alle innlærergreper, i stor grad om tilegnelse av en fast kollokasjon (*it is*) eller 'chunk' (*it's*) heller enn om systemisk tilegnelse av samsvarsbøyning.

I del 4.1 så vi at *am* er uløselig knytta til *I* ved 12–13-årsalderen. Riktignok bruker en del 12–13-åringar kombinasjonen *I are*, men det ingen som bruker *you am* eller *she am*, og det er heller ingen som kombinerer *am* med et substantivisk subjekt. Dette er sannsynligvis et resultat av drilling og den høye frekvensen i input som vi ser for *I am/I'm* i Figur 2, og som vi må anta også eksisterer i inputen elevene får i klasserommet og på fritiden. Det er grunn til å anta at frekvensen i output også er høy. I norsk skole har man stort fokus på at elevene skal lære å uttrykke egne meninger og beskrive egne opplevelser. Dette er nedfelt som kompetansemål i læreplanene, inkludert læreplanen for engelsk (Utdanningsdirektoratet 2013). Med dette følger en utstrakt bruk av 1. person entall i klasserommet og i engelskfaget generelt. Dette gjaldt også de fleste av oppgavene i de nasjonale prøvene fra 2005, som CORYL-korpuset er basert på (Lie mfl. 2005: 84–87, Hasselgren & Sundet 2017: 198), noe som gjenspeiles i de høye frekvensene av *I am* i Tabell 1. I tillegg er *am* sannsynlig-

vis enklere å lære da formen kun opptrer ved *I* og ikke ved andre typer subjekter, i motsetning til *are* og *is*. Her kan en komme langt med drilling da det ikke vil dukke opp forvirrende input i form av ukjente subjekt-verb-kombinasjoner.

5.2 BE i preteritum

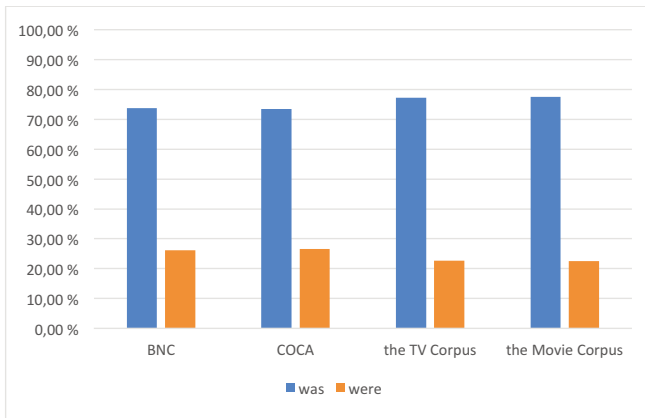
Det at mange norske innlærere ved 15–16-årsalderen fortsatt ikke har klart å matche preteritumsformene av *be* med de personlige pronomenene, kan skyldes at presens introduseres først i skolen. I læreverkene introduseres presensformene opptil flere år før preteritumsformene (Killie & Son under arbeid). Trolig er det i skolen generelt, over mange år, mer fokus på å beskrive ting i presens (seg selv, egne meninger osv.) enn på å snakke om fortida. Følgelig vil tilegnelsen av preteritumsformene lide pga. lav frekvens i input og output.

Det finnes også potensielt forvirrende faktorer i den inputen som elevene får utenfor skolen. Som nevnt i del 2.1 eksponeres norske ungdommer i stor grad for engelske substandard-varieteter, og i disse varietetene finner man andre regler for bruk av *were* og *was* enn i standardvarietetene. Det er heller ikke umulig at forekomster av konjunktiv i inputen kan virke forvirrende. Nå har den engelske konjunktiven i økende grad blitt erstatta av indikativ og uttrykk med modale hjelpeverb (Auer 2009, González-Álvarez 2003, Grund & Walker 2006). Setninger som *if I were you*, *if he were to leave* etc. kan likevel bidra til gjøre reglene for bruken av *were* og *was* uklare. Det virker heller ikke som om konjunktiv gjøres til gjenstand for eksplisitt undervisning i skolen i særlig grad, i hvert fall ikke hvis undervisninga følger lærebøkene for grunnskolen (Killie & Son under arbeid). Konjunktiv diskuteres heller ikke i den læreboka i engelsk grammatikk som, ifølge Caspersen, Bugge & Oppegaard (2017: 41), var mest brukt innen lærerutdanningene i 2016, *Introducing English Grammar* (Dypedahl, Hasselgård & Løken 2012).⁹ Hvis konjunktiven ikke diskuteres i lærerutdanning innen engelsk, vil det ikke være overraskende om verken lærere eller elever har særlig kontroll på den. Vi kan derfor ikke utelukke at elever som legger merke til forekomster av *were*-konjunktiven i input, blir forvirra og kanskje unngår å bruke *were*. Unngåelse ('avoidance') av språkllyder, ord og strukturer/språktrekk som oppleves som vanskelig, antas å være vanlig innen andrespråkstilegnelse (Schachter 1974).

Mens overgeneraliseringa av *are* ble forklart med likhet med norske *er*, er det vanskeligere å finne likhetstrekk mellom *was* og *var*. I skriftspråket deler formene vokalen *a*, men i uttalen finnes det ingen overlapp; ingen av fonemene

9. Dette gjelder også Dypedahl & Hasselgård (2018), som er den nyeste utgaven av boka.

i det engelske ordet finnes i norsk. Kanskje erstatter en del innlærere med norsk språkbakgrunn vokallyden i *was* ([v], [ʌ] eller [ə], avhengig av varietet og trykk) med den norske /a/-lyden, som skaper større lydlikhet mellom de norske og engelske vokalene i *var* og *was*, men det finnes meg bekjent ikke vitenskapelig belegg for å hevde dette. Videre inneholder *were* grafemet <r> og (i amerikansk engelsk og andre rotiske varieteter) fonemet /r/, som gir denne formen et felles grafem/fonem med norske *var*. Generaliseringa av *was* må derfor tilskrives andre årsaker enn likhet med norsk. Unngåelse av *were* er nevnt ovenfor som en mulig faktor. Frekvens er trolig viktig. Figur 3 viser den relative frekvensen av *was* og *were* i de fire undersøkte korpusene.



Figur 3: Den relative frekvensen til preteritumsformene av BE i fire korpus

Som vi ser, er formen *was* langt mer frekvent enn *were* i disse korpusene, og dette vil også gjelde den inputen elevene eksponeres for utenfor skolen. Her kan det nevnes at det i afro-amerikansk engelsk, som norske ungdommer eksponeres mye for gjennom rap, R&B og film, er utbredt bruk av *was* i flertall (Green 2002: 38).¹⁰ Hvis statistisk (probabilistisk) språklæring har noe for seg (se f.eks. Ellis 2006), kan vi se for oss at *was* på grunnlag av input generaliseres som preteritumsform.

10. Dette fenomenet finnes også i britiske varieteter, for eksempel i Hull-dialekten (Cheshire, Kerswill & Williams 2005), men det er kanskje tvilsomt at norske ungdommer i særlig grad eksponeres for britiske substandardvarieteter.

6 Oppsummering og generelle betraktninger

Dataene i denne studien tyder på at 12–13-åringene i CORYL-korpuset overgeneraliserer *are* i presens, mens både 12–13-åringene og 15–16-åringene overgeneraliserer *was* i preteritum. Overgeneraliseringen av *are* er trolig basert på likhet med norske *er*, mens overgeneraliseringen av *was* kan skyldes formens høye frekvens og kanskje til dels også unngåelse av *were*. Elevene er mindre trygge i bruken av preteritumsformene, noe som kan skyldes lav frekvens i input- og i elevenes output. I tillegg foregår det trolig kollokasjon- og chun-klæring ved *it's/it is* og *I am/I'm*. Vi har videre sett at elevene gjør langt færre feil ved personlige pronomen enn ved substantiviske subjekter. Dette ble forklart som et resultat av drilling av pronomen–verb-sekvenser og høy frekvens av disse i input og output. Oppgaven er langt mer utfordrende ved substantiviske subjekter, som kommer i mange ulike former og ikke inngår i drilløvelser.

Vi så i del 2.2 at tilegnelsen av samsvarsbøyning ved *be* blant morsmålsbrukere av engelsk strekker seg over flere år (Serratrice 2001). Dette ser ut til å være tilfellet også blant innlærere av engelsk med norsk språkbakgrunn, men en del faller på plass mellom 12–13- og 15–16-årsalderen, der mange elever tilsynelatende slutter å overgeneralisere *are*.

Det er liten tvil om at samsvarsbøyning utgjør en flaskehals i språktilegnelsen (jf. Slabakova 2019). Dataene vi har sett på her, tyder imidlertid på at det er fullt mulig både å overstyre påvirkning fra førstespråket, å forkaste generaliseringsstrategier og, ikke minst, å tilegne seg trekk som ikke eksisterer i morsmålsgrammatikken (jf. Lardiere 2009).

Referanser

- Auer, Anita. 2009. *The Subjunctive in the Age of Prescriptivism*. Houndmills: Palgrave Macmillan.
- Austin, Jennifer. 2012. Markedness, input frequency, and the acquisition of inflection: Evidence from Basque/Spanish bilingual children. *International Journal of Bilingualism* 17-3, 259–83.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Breiteneder, Angelika. 2005. The naturalness of English as a European lingua franca: The case of the 'third person -s'. *Vienna English Working Papers* 14-2, 3–26.

- Brevik, Lisbeth M. 2019. Gamers, Surfers, Social Media Users: Unpacking the role of interest in English. *Journal of Computer Assisted Learning* 35, 595–606.
- Brown, Roger. 1973. *A First Language: The Early Stages*. London: George Allen and Unwin.
- Caspersen, Joakim, Hanna Bugge & Sigurd Martin Nordli Oppegaard. 2017. Humanister i lærerutdanningen: Valg og bruk av pensum, kompetanse og rekruttering, faglig identitet og tilknytning. En rapport fra Senter for profesjonsstudier. *HiOA Rapport* nr. 2.
- Cheshire, Jenny, Paul Kerswill & Ann Williams. 2005. Phonology, grammar and discourse in dialect convergence. I Auer, Peter, Frans Hinskens & Paul Kerswill (red.): *Dialect Change: Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 135–167.
- Cheshire, Jenny & James Milroy. 1993. Syntactic variation in non-standard dialects: background issues. I Milroy, John & Leslie Milroy (red.): *Real English: The Grammar of English Dialects in the British Isles*. London: Longman, 3–33.
- Collins, Laura & Nick Ellis (red.). 2009. Special issue: Input and second language construction learning: Frequency, form, and function. *The Modern Language Journal*, 93–3.
- DeBose, Charles I. 2015. The systematic marking of tense, modality, and aspect in African American language. I Bloomquist, Jennifer, Lisa J. Green & Sonja L. Lanehart (red.): *The Oxford Handbook of African American Language*. Cary: Oxford University Press.
- Dröschel, Yvonne. 2011. *Lingua Franca English: The Role of Simplification and Transfer*. Bern: Peter Lang.
- Dulay, Heidi C. & Marina K. Burt. 1974. Natural sequences in child second language acquisition. *Language Learning* 24-1, 37–53.
- Dypedahl, Magne, Hilde Hasselgård & Berit H. Løken, 2012. *Introducing English Grammar* (2. utg.). Bergen: Fagbokforlaget.
- Dypedahl, Magne & Hilde Hasselgård. 2018. *Introducing English Grammar* (3. utg.). Bergen: Fagbokforlaget.
- Ellis, Nick C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics*, 27-1, 1–24.
- Fisher, Ulla Thagg. 1985. *The Sweet Sound of Concord: A Study of Swedish Learners' Concord Problems in English*. Ph.d.-avhandling. Lund: Lunds universitet.

- Fraser, Colin, Ursula Bellugi & Roger Brown. 1963. Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior* 2, 121–135.
- Garshol, Lenka. 2018. *I just doesn't know: Agreement errors in English texts by L2 Norwegian learners: Causes and remedies*. Ph.d.-avhandling. Kristiansand: Universitetet i Agder. <https://uia.brage.unit.no/uia-xmlui/handle/11250/2589044>
- González-Álvarez, Dolores. 2003. If He Come vs. If He Comes, If He Shall Come: Some Remarks on the Subjunctive in Conditional Protases in Early and Late Modern English. *Neuphilologische Mitteilungen* 104, 303–313.
- Green, Lisa J. 2002. *African American English: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Grund, Peter & Terry Walker. 2006. The subjunctive in adverbial clauses in nineteenth-century English. I Kytö, Merja, Mats Rydén & Erik Smitterberg (red.): *Nineteenth-Century English: Stability and Change*. Cambridge: Cambridge University Press, 89–109.
- Hasselgren, Angela & Kari Telstad Sundet. 2017. Introducing the CORYL Corpus: What it is and how we can use it to shed light on learner language. *Bergen Language and Linguistics Studies* 7. <https://doi.org/10.15845/bells.v7i0.1107>
- Hauge, Helene & Anja Angelsen. 2018. Developing language awareness. I Bøhn, Henrik, Mange Dypedahl & Gro-Anita Myklevold (red.): *Teaching and Learning English*. Oslo: Cappelen Damm Akademisk, 266–285.
- Haznedar, Belma. 2001. The Acquisition of the IP system in child L2 English. *Studies in Second Language Acquisition* 23, 1–39.
- Housen, Alex. 2002. A corpus-based study of the L2-acquisition of the English verb system. I Granger, Sylviane, Joseph Hung & Stephanie Petch-Tyson (red.): *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching*. Amsterdam/Philadelphia: John Benjamins, 77–116.
- Ionin, Tania & Silvina Montrul. 2010. The role of L1-transfer in the interpretation of articles with definite plurals in L2-English. *Language Learning* 60-4, 877–925.
- Ionin, Tania & Kenneth Wexler. 2002. Why is 'is' easier than '-s'? Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research* 18-2, 95–136.
- Jensen, Isabel Nadine, Roumyana Slabakova, Marit Westergaard & Björn Sundquist. 2020. The Bottleneck Hypothesis in L2 acquisition: L1 Norwe-

- gian learners' knowledge of syntax and morphology in L2 English. *Second Language Research* 36-1, 3–29.
- Keeney, Terrence J. & Jean Wolfe. 1972. The acquisition of agreement in English. *Journal of Verbal Learning and Verbal Behavior* 11, 698–705.
- Killie, Kristin. 2019. The acquisition of subject-verb agreement among Norwegian (teenage) learners of English: focus on the subject. *Journal of Linguistics and Language Teaching* 10-2. https://linguisticsandlanguage teaching.blogspot.com/2019/12/blog-post_25.html
- Killie, Kristin. [under vurdering]. Omission or overgeneralization? Subject-verb agreement errors among Norwegian (teenage) learners of English.
- Killie, Kristin & Minjeong Son. [under arbeid]. Samsvarsbøyning i norske læreverk: implisitt og eksplisitt læring (arbeidstittel).
- Krashen, Stephen. 1985. *The Input Hypothesis: Issues and Implications*. London: Longman.
- Lardiere, Donna. 2009. Some thoughts on a contrastive analysis of features in second language acquisition. *Second Language Research* 25-2, 173–227.
- Lie, Svein, Therese N. Hopfenbeck, Elisabeth Ibsen & Are Turmo. 2005. *Nasjonale prøver på ny prøve: Rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2005*. Universitetet i Oslo/Institutt for lærerutdanning og skoleutvikling. https://www.udir.no/globalassets/upload/nasjonale-prover/5/nasjonale_prover_pa_ny_prove_rapport_ils.pdf
- McCafferty, Kevin. 2004. “Thunder storms is verry dangese in this country they come in less than a minnits notice <...>”: The northern subject rule in Southern Irish English. *English World-Wide* 25-1, 51–79.
- Oancea, Costin. 2017. ‘People speaks like this!’ A few notes on subject-verb agreement in American English dialects. *Analele Universității „Ovidius” din Constanța. Seria Filologie*, XXVIII–1, Section Linguistics, 151–162.
- Preisler, Bent. 1999. Functions and forms of English in a European EFL country. I Bex, Tony & Richard Watts (red.): *Standard English: The Widening Debate*. London: Routledge, 239–267.
- Prévost, Philippe & Lydia White. 2000. Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16-2, 103–133.
- Radford, Andrew. 1990. *Syntactic Theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English*. Oxford: Blackwell.

- Research Group for Language Testing and Assessment. CORYL (Corpus of Young Learner Language). <http://clarino.uib.no/korpuskel/clarino-meta-data?identifier=coryl>
- Rindal, Ulrikke. 2010. Constructing identity with L2: Pronunciation and attitudes among Norwegian learners of English. *Journal of Sociolinguistics* 14-2, 240–261.
- Rindal, Ulrikke & Caroline Piercy. 2013. Being ‘neutral’? English pronunciation among Norwegian learners. *World Englishes* 32-2, 211–229.
- Schachter, Jacqueline. 1974. An error in error analysis. *Language Learning* 24-2, 205–214.
- Serratrice, Ludovica. 2001. Chapter 2: The emergence of verbal morphology and the lead-lag pattern issue in bilingual acquisition. I Cenoz, Jasone, Fred Genesee & International Congress for the Study of Child Language (red.): *Trends in Bilingual Acquisition*. Amsterdam/Philadelphia: John Benjamins, 43–70.
- Slabakova, Roumyana. 2016. *Second Language Acquisition*. Oxford: Oxford University Press.
- Slabakova, Roumyana. 2019. The Bottleneck Hypothesis Updated. I Ionin, Tania & Matthew Rispoli (red.): *Three Streams of Generative Language Acquisition Research: Selected papers from the 7th Meeting of Generative Approaches to Language Acquisition – North America, University of Illinois at Urbana-Champaign*. Amsterdam: John Benjamins, 319–345.
- Slabakova, Roumyana. & Gajdos, Johnathan. 2008. The Combinatorial Variability Hypothesis in the second language. I Bowles, Melissa, Rebecca Foote, Silvia Perpiñán & Rakesh Bhatt (red.): *Selected Proceedings of the 2007 Second Language Research Forum*. Somerville, MA: Cascadilla Proceedings Project, 35–43.
- Sprouse, Rex. A. 2006. Full transfer and relexification: Second language acquisition and creole genesis. I Lefebvre, Claire, Lydia White & Christine Jourdan (red.): *L2 Acquisition and Creole Genesis: Dialogues*. Amsterdam, Netherlands: Benjamins, 169–181.
- Swain, Merrill. 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition* 15, 165–179.
- Swain, Merrill. 2000. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. I Lantolf, James P. (red.): *Sociocultural theory and second language learning*. Oxford: Oxford University Press, 97–114.
- Tan, Aig B. 2005. The use of drill exercises in helping students reduce subject-verb agreement errors in academic writing: A case study in IPBA. *Jurnal*

IPBA 3-2, 67–77.

Utdanningsdirektoratet. 2013. Læreplan i engelsk (ENG1-03). <https://www.udir.no/k106/ENG1-03>

Wexler, Ken. 1994. Optional infinitives, head movement and the economy of derivations. I Lightfoot, David & Norbert Hornstein (red.): *Verb Movement*. Cambridge: Cambridge University Press, 305–350.

White, Lydia. 1992. Long and short verb movement in second language acquisition. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 37-2, 273–286.

White, Lydia. 2003. *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Wolfram, Walt. 2004. The grammar of urban African American Vernacular English. I Kortmann, Bernd & Edgar W. Schneider (red.): *Handbook of Varieties of English*. Berlin: Mouton de Gruyter, 111–132.

Summary

The article discusses the acquisition of subject-verb agreement with *be* by Norwegian learners of English aged 12–13 and 15–16 years. Data from the *Corpus of Young Learner Language* show that many 12–13-year-olds overgeneralize the form *are* in the present tense, while *was* is overgeneralized in the past tense by learners from both age groups. Like Garshol (2018), I argue that *are* is overgeneralized due to phonetic and graphemic similarity with the Norwegian present-tense form *er*. *Was*, by contrast, may be overgeneralized due to its high frequencies. It is further argued that drilling and frequency in input and output generally influence the acquisition of the *be* forms, and that chunk learning plays a role. The data support the idea that the acquisition of the *be* paradigm and of subject-verb agreement is slow (Serratrice 2001) and represents a ‘bottleneck’ in language acquisition (Slabakova 2019).

Kristin Killie
professor i engelsk språk
Institutt for lærerutdanning og pedagogikk
UiT Norges arktiske universitet
9037 Tromsø
kristin.killie@uit.no



Adjektivlæring hos vaksne innlærarar av norsk – ein empirisk studie

Anna-Marie Kjøde Olsen

Universitetet i Bergen

Denne artikkelen utforskar empirisk nokre sider ved læring og bruk av adjektiv hos vaksne innlærarar av norsk som andrespråk. Han er basert på ein kvantitativ tverrsnittstudie av adjektivbruken i 300 tekstar frå Norsk andrespråkskorpus. Eg har samanlikna innlærarmaterialet med tekstar skrivne av førstespråksbrukarar av norsk og i tillegg sett etter samanhengar mellom adjektivbruken og nokre bakgrunnsvariablar. Som mål på adjektivbruken har eg valt frekvens, noko som er motivert av eit bruksbasert teoretisk fundament. Eg ser ein skilnad mellom andrespråksbruken og førstespråksbruken ved at innlærartekstane har ein lågare adjektivfrekvens og mindre variasjon i adjektivbruken sin enn tekstane frå kontrollkorpuset. Vidare er det skilnader mellom innlærarar med tre ulike førstespråksbakgrunnar (somal, tysk og vietnamesisk) som kan tyde på tverrspråkleg påverknad mellom førstespråket og norsk. Det er ikkje sikkert grunnlag for å seie at adjektivfrekvensen varierer med tekstane sitt nivå etter Det felles europeiske rammeverket for språk eller utdanningsbakgrunnen til informantane i denne studien. Det overordna bildet som teiknar seg av adjektivbruken, er den omfattande individuelle variasjonen, noko som er eit venta trekk i innlærarspråk og som finn klangbotn i det teoretiske grunnlaget.

Nøkkelord: adjektiv, adjektivlæring, norsk som andrespråk, bruksbasert, frekvens

1 Innleiing¹

Denne artikkelen presenterer ein kvantitativ tverrsnittstudie av adjektivbruken i tekstar skrivne av vaksne med norsk som andrespråk (S2). Studien nyttar språkbruksdata frå 300 tekstar henta frå Norsk andrespråkskorpus (ASK) (Meurer 2012), og informantane har tre ulike førstespråksbakgrunnar (somali, tysk og vietnamesisk). Målet er å finne ut meir om forekomsten av adjektiv i innlærarnorsk, og kva faktorar adjektivforekomsten kan vise seg å ha ein samanheng med. Adjektiv har i relativt liten grad vore studert i nordisk andrespråkssamanheng ut ifrå eit frekvensperspektiv og leksikalsk perspektiv, og, etter det eg kjenner til, ikkje med basis i same type data eller teoretisk tilnærming som i denne studien. Slik kan studien vere eit empirisk bidrag til utforsking av norsk innlærarspråk basert på språkbruksdata og ut frå eit bruksbasert perspektiv. Studien bruker data frå ulike ferdigheitsnivå i norsk til å skildre utvikling, noko som er ein vanleg framgangsmåte. Vidare føreset studien at den bruken som viser seg i tekstane, gir eit blikk inn i tileigningsprosessen. Dette følgjer av at bruksbaserte tilnærmingar ikkje set noko klart skilje mellom kompetanse og performanse. Både kunnskap om språket og evna til å bruke det er ein del av innlæraren sin kompetanse, og begge komponentane blir etablerte gjennom språkbruk (Mitchell, Myles & Marsden 2013: 7–8).

Forskingsspørsmåla som blir stilte i artikkelen, er:

- 1: Kva slags frekvensmønster kjem til syne i innlærartekstane sin bruk av adjektiv?
- 2: Er det ein kvantitativ skilnad i bruken av adjektiv mellom innlærartekstane og kontrollkorpuset?
- 3: Er det ein kvantitativ skilnad i bruken av adjektiv mellom tekstar på ulike rammeverksnivå?
- 4: Er det ein kvantitativ skilnad i bruken av adjektiv etter kva utdanningsnivå innlæraren har?
- 5: Er det ein kvantitativ skilnad i bruken av adjektiv mellom dei ulike S1-gruppene?

Artikkelen er strukturert slik: Del to omtalar tidlegare forskning, og i del tre skriv eg om adjektiv som ordklasse i norsk og i kontrastivt perspektiv. Del fire tar for seg teoretiske sider ved rolla frekvens har innanfor det bruksbaserte ram-

1. Eg vil gjerne takke ein anonym fagfelle for konstruktive tilbakemeldingar på ein tidlegare versjon av manuset og Jonar Eikeland ved UiB for hjelp med statistikken.

meverket og andre variablar som kan spele inn i bruken av adjektiv. I del fem blir dataa og metoden presentert og i del seks resultatata av undersøkinga. Til slutt blir funna diskuterte i del sju før eg oppsummerer i del åtte.

2 Tidlegare forskning

Det har blitt gjort få studiar av adjektivfrekvens i norsk som andrespråk, men Axelsson (1989, 1994) og Kotsinas (1982) har begge undersøkt adjektiv i munnlege data i svensk S2. Studiane inkluderer adjektiv i både attributiv og predikativ stilling. Kotsinas gjer ei longitudinell kartlegging av ordforrådet til seks voksne, lågt utdanna informantar (op.cit.: 14). Generelt nyttar dei adjektiv (token) med om lag same frekvens (4,2 %) som førstespråksbrukarar (4,4 %), men den individuelle variasjonen er stor (op.cit.: 118). Adjektiva er distribuerte talmessig over få ulike adjektiv (typar) (op.cit.: 40–45). Ho finn at innlærarane med svakast utvikla S2-vokabular, lener seg på høgfrequente ord i større grad enn dei andre, men det kjem ikkje fram om det gjeld adjektiv (op.cit.: 211).

Axelsson sin studie er ein kombinert longitudinell og tverrsnittstudie av 60 voksne S2-innlærere med tre ulike språkbakgrunnar (finsk, polsk og spansk) og to ulike ferdigheitsnivå (Axelsson 1994: 7). Dei spanske og polske innlærarane med eit høgt ferdigheitsnivå i svensk brukte eit høgare tal adjektiv (token) og fleire ulike adjektiv (typar) i intervjuet enn dei på eit lågare ferdigheitsnivå, men hos dei finske informantane var det berre så vidt skilnad mellom ferdigheitsnivåa (op.cit.: 116). Oversikta Axelsson gir, er ikkje justert for talet på ord i intervjuet totalt. Dei finske informantane har samtidig ein lågare frekvens av det Axelsson definerer som «foreign adjectives» enn dei spanske og polske, noko ho tilskriv tverrspråkleg påverknad (op.cit.: 127 & 149). Ho konkluderer med at det ikkje er éin enkelt faktor som bestemmer korleis adjektiv blir lært, men at fleire faktorar, som kommunikative behov, utdanning og personlegdom, tverrspråkleg påverknad, formularlæring og ferdigheitsnivå, verkar saman (op.cit.: 147–150).

I hovudoppgåva si brukte Svendsen frekvens og semantikk for å sjå om det var mogleg å predikere ei læringsløype for adjektiv hos 42 voksne innlærere av norsk som andrespråk (Svendsen 1996: 1). Svendsen gjorde ein tverrsnittstudie med skriftlege elisiteringstestar. Adjektiva i testane var valte både på grunnlag av høg frekvens i ulike frekvensordlister og for å representere ulike semantiske typar. Konklusjonen var at frekvens aleine ikkje kunne predikere ei læringsløype for adjektiv, men at frekvens får høgare forklaringsverdi i kombinasjon med semantikk (op.cit.: 131–132).

Til slutt ser Nilsen (2018) i sin doktorgradstudie på utviklinga av substantivfrasen i tekstar skrivne til Norskprøven.² Tekstane spenner frå A1 til B2 i Det felles europeiske rammeverket for språk (CEFR/Rammeverket) (Europarådet 2001). Ho går ut frå at innlærarane prosentvis tar i bruk stadig fleire ledd i substantivfrasen samtidig i takt med at den språklege erfaringa deira aukar, og indirekte gjeld altså studien attributiv bruk av adjektiv. Resultata så langt kan tyde på at adjektivfrekvensen i substantivfrasar aukar svakt mellom kvart av ferdigheitsnivåa, og at auken ser ut til å vere tydelegast mellom A1 og A2.³

3 Ordklassen adjektiv

3.1 Adjektiv i norsk

Orda i studien er valte basert på at dei er adjektiv i norsk. Adjektiv i norsk er kjenneteikna av at dei lar seg gradbøye og samsvarsbøye i kjønn, tal og bestemtheit (Faarlund, Lie & Vannebo 1997: 345). Semantisk uttrykker dei vanlegvis eigenskapar eller tilstandar (op.cit.: 346). Dei utgjer kjernen i ein adjektivfrase og kan fungere predikativt eller attributivt til eit substantiv op.cit.: 345.). I setninga «*Det vesle huset i skogkanten er raudt og kvitt.*» er *vesle* eit eksempel på attributiv bruk og *raudt* og *kvitt* eksempel på predikativ bruk.

Eg har operasjonalisert adjektivomgrepet etter Norsk referansegrammatikk (Faarlund, Lie & Vannebo 1997) sin definisjon fordi han korresponderer med definisjonen tagginga i ASK bygger på. Definisjonen der er morfologisk fundert og reknar gradbøying som eit tilstrekkeleg kriterium for medlemskap (op.cit.: 22). Det tyder mellom anna at tidsnemnarar (*ofte, fort* og *lenge*), mengdeord (*mange, få*), ordenstal og plasseringsord (*høgre, nedre*) er rekna med. I tillegg inneheld klassen presens partisipp, perfektum partisipp i attributiv stilling og perfektum partisipp i predikativ stilling når dei er kongruensbøyge.

Analyseeininga i den aktuelle studien er ord. Eg bruker leksemdefinisjonen av ord (Sveen, Simonsen & Theil 2000: 74). Medlemmar i eit leksem høyrer til i same ordklasse og har lik tyding, men kan ha ulike bøyingsformer. ASK nyttar *lemma* for leksem, mens eg i teksten held på termen *type* i tråd med Bybee (2008) sin token-/typedistinksjon.

2. Norskprøven – <https://www.kompetansenorge.no/Norsk-og-samfunnskunnskap/Norsk-prove> (05.12.19)

3. Gølin K. Nilsen, personleg kommunikasjon i e-post 14.november 2018.

3.2 Adjektiv i kontrastivt lys

Informantane i den studien har tysk (TY), somali (SO) og vietnamesisk (VI) som førstespråk. Valet av desse språkbakgrunnane er motivert av den ulike avstanden dei har til norsk typologisk og genetisk, og særleg når det gjeld korleis ord som i norsk høyrer til i adjektivklassen, blir klassifiserte. Både generelt og når det gjeld adjektivordklassen meir spesifikt, har tysk langt meir til felles med norsk enn dei to andre språka. Ei side ved slektskapen mellom tysk og norsk som er relevant, er den store mengda kognatar, noko som lettar læringsbyrden både reseptivt (Golden 2014: 112) og produktivt (Jarvis & Pavlenko 2008: 85, Llach 2011: 5). Om vi ser vekk frå kasussystemet, blir ordklassen adjektiv i det store og heile handsama på same måte som i norsk både syntaktisk, morfologisk og semantisk (Janik 2015: 72).

Somali blir rekna som eit syntetisk språk i større grad enn norsk (Husby 2017: 92). Adjektiv blir anerkjent som ein eigen ordklasse i somali av både Husby (2017: 99) og Saeed, i kontrast til grammatikkar som har rekna orda som ei undergruppe av verb (Saeed 1993: 4). Saeed inkluderer både det han kallar «basic adjectives» og «derived adjectives» (ibid.), og sistnemnde er ei stor og ope gruppe som kan avleiast både frå verb og substantiv (op.cit.: 185). Adjektiva i somali kan stå både etterstilt attributivt til substantiv og predikativt i setninga (Husby 2017: 99). Dei kan ha gradbøying og blir bøygd i samsvar med substantivet, unntatt når det gjeld bestemtheit (op.cit.: 100).

Vietnamesisk er på si side typologisk svært ulikt norsk (Rosén 1999: 1). Det kan klassifiserast som eit isolerende språk ettersom orda manglar bøyging. Ord med det vi assosierer med adjektivisk tyding, blir i vietnamesisk vanlegvis sett på som ei undergruppe av verb (Husby 2017, Rosén 1999: 28). Felles for dei er at dei kan stå etter det forsterkande ordet *rât* 'veldig', og sjølv om orda gjerne blir oversett med norske adjektiv, er altså tydinga verbal (Husby 2017: 250). Ordet *lón* tyder til dømes snarare 'å vere stor' enn berre 'stor'. Vietnamesisk har ikkje noko som tilsvare norsk samsvarsbøying, men kan gradbøye uttrykk ved hjelp av eigne funksjonsord (ibid.). Den typologiske avstanden relatert til det adjektiviske kan derfor seiast å vere noko mindre mellom norsk og somali enn mellom norsk og vietnamesisk.

4 Bruksbaserte perspektiv

Det er mange faktorar som kan vere med å påverke tileigninga av eit andrespråk hos voksne. I det følgande skriv eg om frekvens, påverknad frå førstespråket,

utdanningsnivå og ferdigheitsnivå, målt langs CEFR-skalaen, som alle er inkluderte som variablar i denne studien.

4.1 Frekvens

Denne studien plasserer seg innanfor ramma av eit bruksbasert perspektiv. Bruksbaserte tilnærmingar legg stor vekt på innputten si rolle i språklæringa (Ellis & Wulff 2014: 86), og frekvens blir rekna som ein sentral faktor ved innputten som drivar for læringa (Nistov et al. 2018: 116). For det første har innputten innlæraren møter visse strukturelle eigenskapar som gjer han meir eller mindre «prosesserbar» (Berggreen & Tenfjord 1999: 94), og av dei faktorane blir frekvens rekna som den viktigaste. Enkelt sagt går det raskare å identifisere og hente fram ein frekvent språkleg konstruksjon enn ein som er lite frekvent (Ellis 2006: 102). For det andre baserer prosesseringa av innputt seg på umedvitne, statistiske utrekningar, og frekvens er ein heilt nødvendig komponent for å kunne rekne ut kva som er statistisk sannsynleg (Mitchell, Myles & Marsden 2013: 104).

Ulike typar frekvens kan ha ulike effektar i læringsprosessen, og her er skiljet Bybee (2008) har utvikla mellom tokenfrekvens og typefrekvens relevant. Av dei effektane Bybee (2008: 218) set fram, er *bevaringseffekten* den mest relevante i denne studien. Bevaringseffekten er relatert til tokenfrekvens. Høg tokenfrekvens styrkar den kognitive representasjonen av den aktuelle språklege eininga og gjer eininga lettare tilgjengeleg i minnet. Frekvente einingar blir lært raskare og bevart betre (op.cit.: 219). Adjektiv som har høg frekvens i målspråket, vil derfor bli lært raskare enn lågfrekvente. Basert på bevaringseffekten vil vi mellom anna kunne forvente at stor grad av eksponering for målspråket vil vise seg i meir leksikalsk variert adjektivbruk og bruk av meir lågfrekvente adjektiv.

Frekvensseffektar aleine kan likevel ikkje forklare språklæring fullt ut. Det blir mellom anna peikt på at frekvens ikkje ser ut til å ha ein kumulativ effekt på læringa, men følgjer ei bratt læringskurve i starten og flatar gradvis ut (Ellis & Wulff 2014: 78).

4.2 Påverknad frå førstespråket

I eit bruksbasert perspektiv er det òg tilhøve knytte til den assosiative læringsprosessen som er med på å dempe frekvensseffekten (Mitchell, Myles & Marsden 2013: 105), og eitt av desse tilhøva er *learned attention*. Innan innlæraren møter S2, er han allereie kognitivt optimalisert til førstespråket sitt (Ellis & Wulff 2014: 82). På område der S1 og S2 skil seg frå kvarandre, kan resultatet bli at innlæraren ikkje legg merke til trekk i S2 fordi S1 har lært han å oversjå

dei (op.cit.: 83). Slik forklarer bruksbaserte tilnærmingar at sjølv om innlæraren får store mengder innputt, er det likevel ikkje alt som blir til inntak. Førstespråket vil altså kunne påverke korleis effekten av frekvensen i andrespråket blir.

Internasjonal og norsk forskning har vidare vist empirisk at ein ikkje kan sjå vekk frå S1-bakgrunn i S2-studier der innlærarane har ulikt førstespråk (Golden, Jarvis & Tenfjord 2017).

Tverrspråkleg påverknad kan observerast langs ulike aksar (Jarvis & Pavlenko 2008: 20). Særleg relevante for den aktuelle studien er lingvistisk transfer, i motsetnad til transfer på eit konseptuelt nivå, og leksikalsk transfer (ibid.). Når ordkunnskapen i eitt språk påverkar bruken av ord i eit anna språk, kan det vere relatert til ulike sider ved denne ordkunnskapen, for eksempel ordklasse og frekvens (op.cit.: 72–73).

Ei rekke faktorar kan føre til tverrspråkleg påverknad, mellom anna forholdet mellom målspråket og språk innlæraren kan frå før, og kva grad av likskap det er mellom dei (Jarvis & Pavlenko 2008:176). Eit viktig skilje går mellom det som blir kalla objektiv og subjektiv likskap. Kjelda til tverrspråkleg påverknad er primært subjektiv likskap, altså at innlæraren har ei subjektiv oppfatning om likskap mellom S1 og S2, og derfor vel å lene seg på S1 når han lærer og/eller bruker S2 (op.cit.: 178–179). Kva grad av samanfall som faktisk finst mellom språka, den objektive likskapen, spelar inn på kor sannsynleg det er at det som blir overført er positivt eller negativt (op.cit.: 179). Mens positiv transfer kan gjere at eit trekk blir lært raskare, kan negativ transfer på si side mellom anna føre til unngåing (Gujord & Ragnhildstveit 2018: 138). Objektiv likskap fører ikkje i seg sjølv til transfer, og det treng ikkje vere samanheng mellom subjektiv og objektiv likskap, men ein finn likevel langt fleire tilfelle av openberr transfer når målspråket liknar på kjeldespråket (Jarvis & Pavlenko 2008: 177–179).

4.3 Andre kjelder til variasjon

To andre variablar som kan gi opphav til systematisk variasjon i adjektivbruken mellom innlærarane, er utdanningsnivå og ferdigheitsnivået i S2, her forstått som nivåplassering i CEFR. Både desse variablane og informasjon om førstespråk finst i ASK.

Valet av CEFR-variabelen baserer seg på den generelle forventninga om ei gradvis utvikling frå det enkle til det meir komplekse (Berggreen & Tenfjord 1999: 65) og at ferdigheitsnivåa kan spegle ei slik utvikling i adjektivbruk, slik også Nilsen (2018) føreset. Dette samsvarer med tanken om at innlæraren sitt

språklege nettverk gradvis veks fram og aukar i kompleksitet i takt med den språklege erfaringa, noko som harmonerer med fleire teoretiske synspunkt.

Vidare kan utdanningsnivået vere opphav til variasjon på ulike språklege område og kan påverke både kor raskt og kor vellukka ein lærer eit S2 (t.d. Bugge, Carlsen & Gujord, 2014 via Gujord 2017). Andrespråksforskinga har i stor grad konsentrert seg om høgt utdanna⁴ innlærarar, og i studiar som omfattar innlærarar med låg eller manglande utdanning, har ofte ikkje utdanningsvariabelen vore isolert (van de Craats, Kurvers & Young-Scholten 2005: 9). Som ein følge av det, veit vi i dag mindre om utviklinga av eit S2 hos den sistnemnde gruppa, noko som er uheldig når det er ting som tyder på at utdanningsnivået til innlæraren er ein sterkt medverkande faktor for kor vellukka S2-læringa blir (Mitchell, Myles & Marsden 2013: 21). Utdanningsnivå er tatt med i studiane til både Kotsinas (1982), Axelsson (1994) og Svendsen (1996).

5 Metode og data

5.1 Dataa frå ASK

Dataa i studien kjem frå 300 tekstar i det elektroniske innlærarkorpuset ASK (Meurer 2012). Informantanene har anten somali (SO), tysk (TY) eller vietnamesisk (VI) som førstespråk. Dei aktuelle tekstane er svar på Språkprøven i norsk for vaksne innvandrarar (heretter Språkprøven), som skulle måle språkferdigheiter på B1-nivå (Tenfjord, Hagen & Johansen 2009: 55). Alle prøve-svara blei vurderte til bestått. Tekstane har i ettertid blitt plasserte på rammeverksnivå (Carlsen 2012: 31–33), og hamna då på tre ulike nivå, A2, B1 og B2. Det er denne plasseringa eg nyttar.

Tekstane er på om lag 200–400 ord kvar (jf. Tabell 5). Dei omhandlar ulike opne, kvardagslege tema, og til saman svarer dei på 32 ulike oppgåvetitlar. Til kvar tekst er det knytt sjølvrapporterte personopplysningar om mellom anna førstespråksbakgrunn, alder og utdanningsnivå.

Som kontrollkorpus har eg brukt 100 tekstar frå ASK som er skrivne av vaksne med norsk som S1. Tekstane blei til under dei same rammevilkåra som resten av ASK-materialet når det gjeld lengde, tema, hjelpemiddel, tid⁵ osv., men dei er ikkje skrivne i ein autentisk testsituasjon. Tematisk er tekstane svar

4. van de Craats et.al (2005: 8-9) definerer ein lågt utdanna som «an adult who has at most ten years of education in the country of origin», og høg utdanning som meir enn 10 års skulegang.

5. Ingvild Nistov, personleg kommunikasjon juni 2020.

på fire ulike oppgåvetitlar. Så få tekstar med så smal tematikk kan sjølvstekt ikkje gi eit representativt bilde av skriftleg norsk, men dei gir likevel eit visst empirisk samanlikningsgrunnlag for adjektivbruken i same typen tekst.

5.2 Handtering av dataa

Først blei alle dei tagga førekomstane (token) av ordklassen adjektiv søkte opp⁶ og lasta ned som ei typebasert frekvensliste. Feilannotering kan vere ei utfordring i ASK fordi Oslo-Bergen-taggaren (Johannessen et al. 2012) ikkje er konstruert for å handtere innlærarspråk (Tenfjord, Hagen & Johansen 2009: 62). Derfor var det behov for ein nøyte manuell gjennomgang av typelista. Der ein type verka potensielt problematisk, blei enkeltførekomstane søkte opp i kontekst. Når det var grunnlag for det, blei førekomstar og typar fjerna eller slått saman. Døme på endringar eg har gjort i materialet er at⁷:

- adjektiv med to gyldige skrivemåtar er slått saman til eitt (f.eks *gammel* og *gammal*)
- samskrivingar er delte opp og plasserte i rett type (*altforflink*, *mestmulig*)
- feilstavingar er retta opp (*islamisk* og *islamsk*)
- samansette ord som er særskrivne og derfor feilannoterte, er fjerna (*syke* i *sykepleier*)
- typar som tilhøyrer ein annan ordklasse, er fjerna (*butikkpersonal* og *1800-talle*)
- ikkje-adjektiveiske førekomstar som har same form som ein adjektivisk type, er fjerna (substantivførekomstar av *dyr* og *mobil*)
- partisippførekomstar som etter definisjonen er adjektiv, er endra frå infinitivform til partisippform (*leke* → *lekende*, *drepe* → *drept*) og ikkje-adjektiveiske partisipp fjerna
- førekomstar som ser ut til å vere samanblanda ut frå konteksten, er retta opp (ein førekomst av *ferdig* er flytta til *farlig*: «Man bør ikke røyke mye for røyking kan være ferdig for kroppen, det vil ødelegge lunger og brenne tenner (tekst-id s0882).»)

6. Søkestreng i ASK: [pos = «adj»] \ <> \ \ :: language = «somali|vietnamesisk|tysk» & testlevel = «Språkprøven»

7. I ettertid av analysane har eg gått gjennom ei gruppe ord i datasettet på nytt. Det gjeld nokre få ord som ASK har tolka som språk-/nasjonalitetsadjektiv, men som også kan vere substantiv. Totalt 12 treff (7 'norsk') av dei totalt 7408 treffa er feilkategoriserte som adjektiv.

- førekomstar utan ein klar parallell i ordboka, er forsøkt tolka og retta opp etter målspråksnormen (førekomsten *tryggdelig* er tatt inn i typen *trygg*: «Jeg fikk panikk og bad Herren om vi kunne komme ned tryggdelig (tekst-id s0100).»

Utover den gjennomgangen som er skildra over, er det ikkje tatt omsyn til om bruken av adjektiva er grammatisk eller målspråkslik. Etter gjennomgangen sat eg igjen med 7405 førekomstar fordelt på 593 typar, mot opphavleg 7408 førekomstar fordelt på 692 typar.

5.3 Statistisk testing

I den statistiske testinga held eg meg til alfanivået $p < 0,05$, slik vanleg praksis er i andrespråksforskinga (Larson-Hall 2010: 389). Utvala er normalitetstesta med ein Shapiro Wilks-test (jf. Vedlegg 1) i tillegg til at eg har vurdert fordelingskurva visuelt i eit histogram (Figur 1, s. 160). Ved samanlikning av to grupper med normalfordelte data bruker eg ein t-test, og tilsvarande One-way ANOVA ved samanlikning av tre normalfordelte grupper, med Tukey HSD som post hoc-test. Der post hoc-testen finn ein statistisk⁸ skilnad mellom to grupper, blir Cohens *d* (Hedges 1981) brukt som effektmål. Når tre grupper med ikkje-normalfordelte data blir samanlikna, bruker eg Kruskal-Wallis og Mann Whitney U som post hoc-test. Alfanivået er Bonferroni-korrigert manuelt i desse tilfella (Field 2009: 69) og som effektmål bruker eg *r* (op.cit.: 227). Testane i studien er to-hala.

6 Resultat

I det følgande vil eg presentere resultatata av adjektivbruken i datasettet ved å gå gjennom forskingsspørsmåla suksessivt. Først går eg gjennom sentral- og spreingsmål for dei viktigaste variablane, med vekt på tokenfrekvens, og så viser eg resultatata av den statistiske testinga for variabelen tokenfrekvens.

6.1 Forskingsspørsmål 1

Det første forskingsspørsmålet dreier seg om kva frekvensmønster som kjem til syne i innlærartekstane sin bruk av adjektiv. Tabell 1 viser sentrale verdiar for tekstlengde som kan gi oss eit inntrykk av datasettet. Analysane her omfattar

8. I tråd med tilrådingane frå Larson-Hall (2010: 103–104) har eg valt å omtale resultat under alfanivået som *statistiske*, snarare enn *signifikante*.

alle dei 300 tekstane, og tala er på individnivå. Som vi kan sjå, er det stort spenn i tekstlengde. Skeivskapsmålet viser at fordelinga er lett positivt asymmetrisk ($N(300) = 1,26, p < 0.05$).

Tabell 1: Oversikt over tekstlengde i talet på ord i datasettet på individnivå

Gjennomsnitt	Median	St.avvik	Kortaste tekst	Lengste tekst	Skeivskap	Kurtose
258	246	70	110	595	1,3	3,1

For å få eit overblikk over mengda adjektiv som er brukt på gruppenivå i heile datasettet, går vi til Tabell 2. Han viser at det totalt er brukt 7405 adjektiv, og dei er fordelte på 593 ulike typar. Det høge talet på typar med berre éin førekomst, viser at datasettet er tungt i den eine enden.

Tabell 2: Oversikt over talet på adjektiv i datasettet på gruppenivå

	Løpeord	Adjektivtoken	Ulike typar	Typar med eitt treff
Totalt	77466	7405	593	244

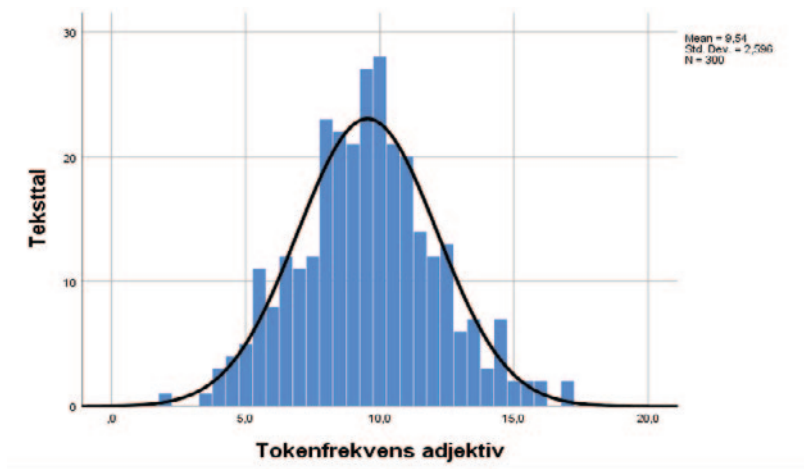
Tabell 3 viser til talet på og frekvensen av adjektiv (både token og type, rekna ut som talet på adjektiv dividert med talet på ord i teksten) i det samla materialet på individnivå.

Tabell 3: Oppsummering av frekvens og tal i datasettet på individnivå

Variabel	Gj.snitt	Median	St.avvik	Min	Maks	Skeivskap	Kurtose
Talet på adjektivtoken	24,7	23	9,7	5,0	68,0	1,0	1,8
Talet på adjektivtypar	16,3	15	6,2	5,0	42,0	1,0	1,6
Frekvensen av adjektivtoken	9,6	9,6	2,6	2,0	16,8	0,1	0,0

Oversikta syner at ingen av tekstane er utan adjektiv, og at spennet i talet på adjektivtoken er stort, mellom fem og 68. Gjennomsnittleg har dei 24,7 adjektiv, men standardavviket tilseier at variasjonen er relativt stor. Utvalet er igjen noko positivt assymmetrisk ($N(300) = 0,965, p < 0.05$). Adjektiva utgjer i snitt 9,6 % av løpeorda i tekstane. Vidare inneheld tekstane i snitt 16,3 ulike adjektivtypar kvar, med eit spenn på mellom fem og 42. Også når det gjeld mengda ad-

jektivtypar og frekvensen av adjektivtoken, er det ein del variasjon i datasettet. Av den grunn kan det vere nyttig å vise fordelinga av tekstane etter adjektivfrekvens med eit histogram (Figur 1). Det illustrerer korleis tekstane i stor grad følger kurva for normalfordeling når det gjeld tokensfrekvensen av adjektiv, med ein topp rundt 10,0. Den lågaste frekvensen er 2,0, og den med høgast frekvens inneheld 16,8 (jf. Tabell 3).



Figur 1: Histogram for fordelinga av tekstar etter tokenfrekvens

Samanfatta kan vi seie at tekstane i materialet er på i overkant av 250 ord gjennomsnittleg. Av desse orda er i snitt 25 adjektiv, og dei er fordelte på 16 ulike adjektivtypar. I frekvens utgjer adjektiva 9,6 % av den totale ordmengda. Spennet i verdiane på alle desse områda er likevel relativt stort.

6.2 Forskingsspørsmål 2

Det neste forskingsspørsmålet gjeld om innlærartekstane sin bruk av adjektiv skil seg frå kontrollkorpuset. I Tabell 4 ser vi ein oversikt over løpeord og adjektiv i dei to korpusa på gruppenivå. Sjølv om det er langt færre kontrolltekstar enn innlærartekstar, inneheld dei både fleire ulike adjektivtypar totalt og fleire typar med berre eitt treff enn innlærartekstane.

Tabell 4: Oversiktstal frå ASK på gruppenivå

	Løpeord totalt	Talet på adjektivtoken	Talet på adjektivtypar	Typar med eitt treff
ASK Språkprøven (N=300)	77466	7405	593	244
ASK kontrollkorpus (N=100)	27495	3068	614	345

I Tabell 5 kan vi sjå at tekstane i kontrollkorpuset i snitt er noko lengre enn innlærartekstane, men skilnaden er relativt liten. Om vi ser på førekomstane av adjektiv målt i tokenfrekvens, talet på token og typar, er skilnaden tydelegare – igjen med høgare verdiar i kontrollkorpuset. Ein t-test for uavhengige utval på variabelen tokenfrekvens slo ut statistisk mellom gruppene ($t(398) = -5,411$, $p < 0,0005$, $d = 0,60$).

Tabell 5: Oppsummering av datasettet på individnivå etter korpus

		Ord i teksten	Talet på adjektivtoken	Frekvensen av adjektivtoken	Talet på adjektivtypar
ASK Språkprøven (N=300)	Gj.snitt	258	24,7	9,6	16,3
	Median	246	23	9,6	15
	St.avvik	70	9,7	2,6	6,2
	Min	110	5	2	5
	Maks	595	68	16,8	42
ASK kontrollkorpus (N=100)	Gj.snitt	275	30,7	11,3	21,5
	Median	264	29,0	11,4	21,0
	St.avvik	71	11,1	3,5	7,5
	Min	78	5	2,7	6
	Maks	625	67	21,5	50

Oppsummert har altså førstespråksbrukarane ein noko høgare tokenfrekvens enn innlærarane i tekstane sine. Typetalet og talet på unike adjektiv er også høgare i S1-datasettet.

6.3 Forskingsspørsmål 3

Det tredje forskingsspørsmålet ser adjektivbruken i tekstane opp mot den plasseringa dei har fått på CEFR-skalaen. ASK opererer med mellom- og heilnivå, men for å få meir solide utval, har eg slått dei saman til heile nivå.⁹ Då står eg igjen med A2, B1 og B2. Dei to lågaste nivåa har omlag like mange tekstar kvar

9. A2 og A2/B1 blir rekna som A2. B1 og B1/B2 blir rekna som B1. B2 og B2/C1 blir rekna som B2.

(N=135 og N=144), mens B2-utvalet er lite (N=21). Tabell 6 viser sentral- og spreingsmål etter kva rammeverksnivå tekstane er tildelte.

Tabell 6: Oppsummering av datasettet på individnivå etter CEFR-nivå

	Ord i teksten	Talet på adjektivtoken	Frekvensen av adjektivtoken	Talet på adjektivtypar	
A2 (N=135)	Gj.snitt	225,7	21,0	9,4	13,6
	Median	219,0	21,0	9,4	13,0
	St.avvik	50,3	7,3	2,7	4,4
	Min	110	8,0	3,4	5,0
	Maks	393	44,0	15,8	26,0
B1 (N=144)	Gj.snitt	276,1	26,8	9,7	17,7
	Median	263,5	25,0	9,7	17,0
	St.avvik	63,2	9,9	2,5	6,1
	Min	177	5,0	2,0	5,0
	Maks	540	68,0	16,8	42,0
B2 (N=21)	Gj.snitt	345,1	33,8	9,8	23,4
	Median	332,0	34,0	9,9	23,0
	St.avvik	101,4	11,4	2,2	7,3
	Min	212	11,0	4,9	8,0
	Maks	595	58,0	16,1	36,0

Tekstlengda er kortast på A2-nivået og aukar tydeleg for kvart ferdigheitsnivå. Det same ser vi i talet på token og typar. Sentralmåla for tokenfrekvens viser den same auken mellom kvart av dei tre utvala, sjølv om skilnaden er liten mellom alle tre nivåa. Ein One-Way ANOVA-test viser heller ingen statistisk skilnad mellom gruppene for denne variabelen ($F_{2,297} = 0,516, p = 0,597$). Det ser altså ut til at innlærarane har omlag den same frekvensen av adjektiv i tekstane sine uavhengig av ferdigheitsnivå.

6.4 Forskingsspørsmål 4

Det neste forskningsspørsmålet dreier seg om utdanningsbakgrunnen til informantane. Av dei 300 informantane har om lag 1/3 rapportert at utdanninga deira ligg på vidaregåande nivå og om lag 1/3 har høgare utdanning. Den siste tredjedelen har anten gått grunnskule (46), har kryssa av for «Anna» (32) eller ikkje svart på spørsmålet (7). Dei informantane det manglar opplysningar om, er utelatne. Kategorien «Anna» er vanskeleg å tolke, og sidan kategorien kan påverke den statistiske analysen og dermed konklusjonen, vil heller denne ikkje bli tatt med vidare. Ein oversikt over sentral- og spreingsmål kjem fram i Tabell 7.

Tabell 7: Oppsummering av datasettet på individnivå etter utdanning

		Ord i teksten	Talet på adjektivtoken	Frekvensen av adjektivtoken	Talet på adjektivtyper
Grunnskule (N=46)	Gj.snitt	233,0	19,7	8,6	13,7
	Median	224,0	20,0	8,3	13,5
	St.avvik	63,2	7,9	3,0	5,6
	Min	110	5,0	2,0	5,0
	Maks	455	45,0	15,2	36,0
Vidaregåande (N=104)	Gj.snitt	263,8	25,3	9,6	16,3
	Median	248,0	24,0	9,7	16,0
	St.avvik	77,6	9,4	2,4	5,9
	Min	118	9,0	4,2	5,0
	Maks	595	58,0	15,8	34,0
Høgskule/ universitet (N=111)	Gj.snitt	269,6	25,9	9,5	16,8
	Median	253,0	24,0	9,3	16,0
	St.avvik	71,4	10,6	2,6	6,7
	Min	123	8,0	4,4	7,0
	Maks	540	68,0	16,8	42,0

Tekstlengda er stigande etter utdanningslengde, frå «Grunnskule» til «Vidaregåande» og så vidt frå «Vidaregåande» til «Høgskule/universitet». Det same er tilfellet for talet på token og typer. Likeeins er sentralmåla for frekvensen av adjektivtoken ganske like for gruppene med vidaregåande og høgare utdanning, og noko lågare i grunnskulegruppa. Ein One-Way ANOVA-test viser ikkje likevel nokon statistisk skilnad mellom utdanningsgruppene for tokenfrekvens ($F_{2,258} = 2,523, p = 0,082$). Oppsummert verkar det altså ikkje til å vere nokon samanheng i datasettet mellom utdanningsnivå og tokenfrekvensen av adjektiv.

6.5 Forskingsspørsmål 5

For å finne svar på om det er ein kvantitativ skilnad mellom S1-gruppene i bruken av adjektiv, deler eg datasettet i tre like store grupper (å N=100) etter kva førstespråksbakgrunn informantane har. I Tabell 8 blir sentral- og spreingsmål presentert på individnivå. Dei viser at innlærere med tysk som S1 (TY) skriv dei klart lengste tekstane og har høgast førekomst av adjektiv både i tal og frekvens. Tilsvarende har gruppa med somali (SO) som S1 dei lågaste verdiane, og dei vietnamesisktalande (VI) ligg i midten. Samtidig er spreinga i tekstlengde størst i gruppa med tysk som S1 og minst hos SO. Motsett er spreinga i tokenfrekvens minst i TY, mens VI og SO har litt større spreing.

Tabell 8: Oppsummering av datasettet etter S1 på individnivå

	Mål	Ord i teksten	Talet på adjektivtoken	Talet på adjektivtypar	Frekvensen av adjektivtoken
Somali (N=100)	Gj.snitt	228,5	20,0	12,9	8,8
	Median	220,5	19,0	13,0	8,4
	St.avvik	53,5	7,8	4,3	3,0
	Min	110	5	5	2,0
	Maks	410	49	25	15,8
Tysk (N=100)	Gj.snitt	291,3	30,5	20,7	10,5
	Median	270,5	29,0	20,0	10,1
	St.avvik	79,4	10,2	6,5	2,0
	Min	163	15	10	5,8
	Maks	595	68	42	16,8
Vietnamesisk (N=100)	Gj.snitt	254,9	23,6	15,2	9,3
	Median	240,0	23,0	15,0	9,3
	St.avvik	60,0	7,7	4,7	2,5
	Min	133	8	5	4,2
	Maks	420	44	26	16,8

Vidare vil eg sjå om skilnaden mellom S1-gruppene gir utslag statistisk. Ein Kruskal-Wallis H-test viser at skilnaden i tokenfrekvens mellom gruppene gir statistiske resultat ($H = 22,718$, $df=2$, $p < 0,0005^{10}$). Gruppene blir så samanlikna parvis med Mann Whitney U som post hoc-test ($\alpha = 0,017$). Post hoc-testane viser at skilnadene er statistiske både mellom SO (mdn = 8,4) og TY (mdn = 10,1) ($U = 3150,0$, $p < 0,0005$, $r = 0,32$) og mellom TY (mdn = 10,1) og VI (mdn = 9,3) ($U = 3603,0$, $p = 0,001$, $r = 0,24$), men ikkje mellom VI (mdn = 9,3) og SO (mdn = 8,4) ($U = 4427,5$, $p = 0,162$). Det er altså lite truleg at skilnaden i tokenfrekvens mellom TY på den eine sida og SO og VI på den andre, er tilfeldig. Effekten er middels mellom SO og TY og liten til middels mellom TY og VI.

Resultata frå forskings spørsmålet peikar i retning av ein skilnad mellom førstespråksgruppene, men det er likevel ikkje gitt at skilnaden ligg i S1-bakgrunn. Andre forskjellar som heng saman med språkbakgrunnen, kan ligge bak. Innanfor Jarvis sitt metodiske rammeverk for identifisering av transfer, er det tilrådd å kontrollere andre faktorar (Jarvis 2000: 260–261).¹¹ Eg vel derfor å

10. Verdiane i SPSS viser $p=0,000$, men eg rapporterer det som $p < 0,0005$ både her og seinare ettersom skilnaden i realiteten ikkje er null (Larson-Hall 2010: 279).

11. Studien kvalifiserer til krava i Jarvis (2000) sitt rammeverk når det gjeld intragruppe-homogenitet og intergruppe-heterogenitet i performansen ettersom eg bruker statistikk som kan vise at åtterden til ei gruppe er ulik nok ei anna gruppe – og lik nok internt – til å hevde at det observerte kan koplast til S1-bakgrunnen. Vidare gir eg under overskrifta «Andre påverknadsfaktorar» eit oversyn over nokre kontrastive forhold mellom dei aktuelle språka (Gujord & Ragnhildstveit 2018: 141).

sjå nærare på distribusjonen av språkbakgrunnar saman med høvesvis rammeverksnivå og utdanningsnivå.

6.6 Førstespråk og ferdigheitsnivå

Det blir generelt rekna som viktig å halde kontroll på ferdigheitsnivå i andre-språksstudiar. Om ferdigheitsnivået varierer mellom S1-gruppene, kan det vise seg i adjektivbruken deira. Om ferdigheitsnivået derimot er jamt, kan ein skilnad i adjektivbruken tyde på ein transfereffekt. Tabell 9 viser korleis tekstane i materialet fordeler seg på ulike nivå etter kva S1 informantane har. Ein langt høgare del av dei tysktalande innlærarane er vurderte til B-nivået enn i dei to andre språkgruppene. Hovudvekta av TY-tekstane ligg på B1-nivå, og fleire er på B2-nivå enn A2. Til samanlikning er ingen av tekstane frå VI og berre éin frå SO på B2-nivå, og fleirtalet av tekstane i desse gruppene ligg på A2-nivå.

Tabell 9: Krysstabell for talet på tekstar på gruppenivå etter S1 og CEFR-nivå

	Somali (N=100)	Tysk (N=100)	Vietnamesisk (N=100)	Total (N=300)
A2	69	11	55	135
B1	30	69	45	144
B2	1	20	0	21

Den skeive fordelinga av gruppene er ei potensiell feilkjelde i materialet. For å finne ut meir om nivåplasseringa i forhold til S1 og adjektivbruk, gjer eg derfor ein analyse med CEFR-nivå som kontrollvariabel. Frå Tabell 10 kan vi sjå at TY-utvalet skil seg ut med dei høgaste middelverdiene uavhengig av CEFR-nivå, mens SO-utvalet har dei lågaste. Variasjonen er noko større i SO og VI enn i TY.

Tabell 10: Frekvensen for adjektivtoken etter S1 og CEFR-nivå

		CEFR-nivå			
		A2	B1	B2	
Førstespråk	Somali	N=	69	30	1
		Gj.snitt	9,1	8,3	4,9
		St.avvik	3,0	2,7	
	Tysk	N=	11	69	20
		Gj.snitt	10,5	10,6	10,1
		St.avvik	1,1	2,1	2,0
	Vietnamesisk	N=	55	45	0
		Gj.snitt	9,5	9,2	
		St.avvik	2,5	2,5	

Den statistiske testinga vidare blir gjort utan B2-nivået. Når CEFR-nivåa blir haldne konstante, viser ein Kruskal-Wallis-test ikkje statistiske resultat mellom S1-gruppene på A2-nivå ($H = 3,855$, $df = 2$, $p = 0,146$), men derimot på B1-nivå ($H = 19,823$, $df = 2$, $p < 0,0005$). Ein Mann Whitney U-test viser at det er statistiske skilnader ($\alpha = 0,017$) på B1-nivå mellom både SO (mdn = 8,1) og TY (mdn = 10,5) ($U = 523,0$, $p < 0,0005$, $r = 0,39$) og TY (mdn = 10,5) – VI (mdn = 9,0) ($U = 987,5$, $p = 0,001$, $r = 0,31$), men ikkje mellom SO (mdn = 8,1) – VI (mdn = 9,0) ($U = 557,0$, $p = 0,202$). Resultata tyder altså på at skilnaden på B1-nivå mellom TY på den eine sida og SO og VI på den andre ikkje er tilfeldig, og i begge tilfella slår effekten ut som middels.

Eit usikkerheitsmoment ved CEFR-nivået er at heile nivå og mellomnivå er slått saman (jf. forsk.spm. 3). Som vi ser av Tabell 11, er fordelinga mellom S1-bakgrunnar ujamn, særleg innanfor det som blir rekna som B1-nivået (B1 og B1/B2), der testen viste statistiske skilnader. TY har totalt flest tekstar på B1-nivå, og ein stor del av dei er vurdert til mellomnivået B1/B2, i motsetnad til SO- og VI-tekstane, som for det meste er vurderte til heilnivået.

Tabell 11: Fordeling av CEFR-nivå med mellomnivå på førstespråksbakgrunn

	A2	A2/B1	B1	B1/B2	B2	B2/C1
Somali	14	55	27	3	1	0
Tysk	3	8	42	27	19	1
Vietnamesisk	3	52	44	1	0	0

For å sjekke om skilnaden på B1-nivå kunne skrive seg frå ei skeivfordeling av S1-bakgrunnane på B1 og B1/B2, gjorde eg ein analyse der eg fjerna mellomnivået og heldt det utanfor analysen av B1-nivået. Utvala var normalfordelte og relativt like i storleik: SO ($N=27$, $m=8,6$, $w=0,983$, $p=0,929$), TY ($N=42$, $m=10,6$, $w=0,985$, $p=0,863$) og VI ($N=44$, $m=9,1$, $w=0,961$, $p=0,143$). Ein One-way ANOVA viste statistiske resultat ($F_{2,110} = 6,478$, $p = 0,002$). Skilnaden er framleis statistisk mellom TY (mdn = 10,6) og SO (mdn = 8,3) ($p = 0,004$, $d = -0,88$) og TY (mdn = 10,6) og VI (mdn = 8,95) ($p = 0,019$, $d = -0,65$), men ikkje mellom SO (mdn = 8,3) og VI (mdn = 8,95) ($p = 0,640$). Effekten er stor mellom TY og SO og middels mellom SO og VI.

Utifrå dette kan vi konkludere med at skilnadene mellom den tyskspråklege gruppa og dei to andre mest truleg ikkje er eit tilfeldig resultat, og at det eksisterer ein reell skilnad mellom dei på B1-nivået. Vi kan ikkje seie det same om skilnaden mellom SO og VI.

6.7 Førstespråk og utdanningsnivå

Slik vi har sett på S1-gruppene i lys av CEFR-nivået til tekstane, vil eg òg undersøke utdanningsbakgrunnen deira. Tabell 12 viser ei fordeling av tekstane etter kva S1 og utdanningsbakgrunn informantane har. Vietnamesisktalende dominerer i gruppa for høgare utdanning, mens TY har flest på vgs-nivå og SO klart flest med grunnskulebakgrunn.

Tabell 12: Krysstabell for talet på tekstar på gruppenivå etter utdanning og S1

		Utdanning		
		Grunnskule	Vidaregåande	Høgskule/universitet
Førstespråk	Somali	31	35	28
	Tysk	3	40	33
	Vietnamesisk	12	29	50
Total		46	104	111

I Tabell 13 finn vi den høgaste middelværdien for adjektivfrekvens i gruppa som har TY som S1 og utdanning rapportert som «Grunnskule» (11,4). Den lågaste middelværdien har SO-utvalet i same utdanningskategori (8,2). Samtidig er det ein del variasjon i utvala, særleg i SO.

Tabell 13: Sentral- og spreingsmål for adjektivfrekvens etter S1 og utdanning på individnivå

			Utdanning		
			Grunnskule	Vidaregåande	Høgskule/universitet
Førstespråk	Somali	N=	31	35	28
		Gj.snitt	8,2	9,2	9,0
		St.avvik	3,1	3	2,8
	Tysk	N=	3	40	33
		Gj.snitt	11,4	9,7	10,5
		St.avvik	2	1,6	2
	Vietnamesisk	N=	12	29	50
		Gj.snitt	8,9	9,9	9,2
		St.avvik	2,7	2,4	2,6

VI-utvalet med grunnskulebakgrunn er ikkje normalfordelt ($w=0,0844$, $p=0,031$). Ein Kruskal-Wallis-test der utdanningsbakgrunnen blir halden konstant gir statistiske resultat mellom S1-gruppene i «Høgskule/universitet» ($H = 7,297$, $df = 2$, $p = 0,026$), men ikkje i «Grunnskule» ($H = 3,561$, $df = 2$, $p =$

0,169) og «Vidaregåande» ($H = 1,906$, $df = 2$, $p = 0,386$). Ein Mann Whitney U viser ikkje statistiske resultat i gruppa «Høgskule/universitet» ($\alpha = 0,017$) verken mellom SO (mdn = 8,55) og TY (mdn = 10,1) ($U = 304,5$, $p = 0,023$), mellom SO (mdn = 8,55) og VI (mdn = 9,2) ($U = 671,0$, $p = 0,763$), berre mellom TY (mdn = 10,1) og VI (mdn = 9,2) ($U = 566,0$, $p = 0,016$). Det er altså berre grunnlag for å seie at det er skilnad mellom høgt utdanna tysk- og vietnamesisktalande.

7 Drøfting

7.1 Frekvens og språkleg utviklingsnivå

Utforskinga av forskingsspørsmål 1 har gitt oss eit inntrykk av datasettet, og i forskingsspørsmål 3 såg vi etter samanheng mellom tekstane sitt CEFR-nivå og adjektivbruken. Sjølv om dei deskriptive tala viste ein knapp auke i tokenfrekvens for kvart nivå frå A2 til B2, slo ikkje desse ut ved statistisk testing. Mine data gir altså ikkje nokon tydeleg indikasjon på at det eksisterer ein samanheng mellom tokenfrekvensen av adjektiv og det språklege utviklingsnivået til innlærarar av norsk på dei aktuelle CEFR-nivåa. Nilsen (2018) ser ein eventuell auke i adjektivfrekvens frå A1 til B2 som eit teikn på auka kompleksitet i innlærarspråket. Ei slik utvikling kan vere i tråd med dei forventingane vi har til læringsgang generelt om aukande kompleksitet (Berggreen & Tenfjord 1999: 65). Kompleksiteten kan ligge i at nominalfrasen blir bygd ut ved at substantivkjernen blir moderert eller presisert (Faarlund, Lie og Vannebo 1997: 346) ved hjelp av adjektiv. Det kan òg vere kognitive krav knytte til bruken av adjektiv som gjer dei meir komplekse å lære og ta i bruk enn til dømes substantiv og verb, noko som tilseier at innlærarane får grep om adjektivbruken på eit seinare tidspunkt i læringsprosessen (Llach 2011: 5). I Axelsson (1994: 116) sin studie brukte informantar på eit høgare utviklingsnivå talmessig fleire token og typar av adjektiv enn dei på eit lågare nivå, men det har låg samanlikningsverdi så lenge vi ikkje kjenner frekvensen. Kotsinas (1982: 45) rangerer ikkje informantane etter ferdigheitsnivå, men oppgir storleiken på ordforrådet dei bruker i materialet og adjektivfrekvensen. Eg kan ikkje sjå at det er nokon klar samanheng mellom dei to. Nilsen (2018) finn ein svak auke i frekvensen frå A1–B2 og at auken ser ut til å vere tydelegast mellom A1 og A2. Det same kan ikkje seiast med bakgrunn i mitt materiale, men her er heller ikkje tekstar på det lågaste rammeverksnivået representerte. Ein mogleg konklusjon kan altså vere

at innlærerne i materialet mitt allereie har tileigna seg ein adjektivtokenfrekvens som liknar førstespråksbruken.

På same tid som vi ikkje kan sjå nokon klar samheng mellom frekvens og utviklingsnivå (målt i CEFR-nivå) i studien min, viser dataa stor grad av variasjon mellom innlærerne seg i mellom. Det same viser seg både i Kotsinas (1982) og Axelsson (1994) sine studiar. Systematikk og variasjon i andrespråklæringa er tilbakevendande tema innanfor andrespråksforskninga. Ei rekke faktorar kan skape variasjon (Berggreen & Tenfjord 1999: 26–27), og bakgrunnsvariablane i denne studien kan representere nokre av dei. Men i eit bruksbasert perspektiv knyt ein òg variasjon til faktorar ved innputt:

Such variability is a natural consequence of input factors such as exemplar type and token frequency, recency, context, salience, contingency, regularity, and reliability, along with the various other associative learning factors that affect the emergence of attractors in the problem space (Ellis & Wulff 2014: 87).

Den framtrekande variasjonen i materialet mitt stemmer med at bruksbaserte tilnærmingar ser variasjon som sjølve standarden ved språklæring og -bruk (Nistov et al. 2018: 120). I sin natur er språkbrukaren sin kompetanse eit dynamisk system som blir forma av erfaringane og dei rammene språkbruken til ei kvar tid skjer innanfor (ibid.), anten det er i S1 eller S2. Dette synet representerer ei viss relativisering samanlikna med meir tradisjonelle tilnærmingar.

I forskings spørsmål 2 jamførte eg innlærartekstane med kontrollkorpuset i ASK. Frekvensen i dei to datasetta skil seg statistisk frå kvarandre ved at innlærartekstane inneheld relativt færre adjektiv enn kontrolltekstane. Kotsinas (1982: 233) konkluderte i studien sin av munnleg språkbruk med at det var lite som skilte adjektivfrekvensen til informantane hennar frå S1-bruk, men det blei ikkje testa statistisk. Det er mogleg at modusen, altså at Kotsinas sitt materiale er munnleg og mitt skriftleg, kan ha noko å seie for at Kotsinas får eit anna resultat, for i mitt materiale er det altså ein skilnad.

Ei forklaring på ulikt frekvensmønster i S1-bruk og S2-bruk kan ha med den ulike mengda norsk innputt kvar av desse gruppene har møtt. Vaksne S1-brukarar har opplevd betydeleg meir eksponering for norske adjektiv og kan dra veksling på sterkare og fleire kognitive representasjonar enn innlærerne i korpuset venteleg kan. I oppbygginga av desse mentale representasjonane er frekvens ein sentral faktor. Skilnader i læringsprosessane som ligg bak S1-læring og S2-læring kan òg gjere at innlærerne i mindre grad har nyttegjort seg

av innputten dei har møtt på norsk. Til saman kan det føre til eit bruksmønster av adjektiv som skil seg frå det personar med norsk som S1 har.

Det kan vere verdt å påpeike at ein skilnad i rammevilkåra kan ha verka inn på tekstproduksjonen. S2-brukarane skreiv som del av ein språkprøve, mens S1-brukarane berre var med i eit forskingsprosjekt. Presset på S2-informantane var derfor større, og skrivinga hadde eit anna siktemål. Vilkåra var likevel dei same når det gjeld tidsramme, oppgåvetype og tilgang på hjelpemiddel.

7.2 Leksikalsk variasjon

Materialet mitt gir òg ei viss innsikt i den leksikalske variasjonen i informantane sine tekstar jamført med kontrollkorpuset. Ein måte å måle leksikalsk variasjon på, er å sjå på forholdet mellom typar og token. Slike mål er sårbare for tekstlengde, ettersom lengre tekstar normalt inneheld færre typar i forhold til tekstlengda enn kortare tekstar. Som vi kan sjå av Tabell 5, er tekstane noko lengre i kontrollkorpuset enn i innlærartekstane. På gruppenivå i Tabell 4 ser vi òg at trass i at innlærartvalet (N=300) er tre gongar større enn kontrollutvalet (N=100), finn vi flest ulike typar adjektiv og flest typar med berre éin førekomst i kontrolltekstane. For det tredje ser vi i Tabell 5 at kvar av tekstane i kontrollkorpuset i snitt inneheld langt fleire adjektivtypar enn innlærarkorpuset. Dette peikar i retning av større leksikalsk variasjon i tekstane til S1-brukarane.

Eit forhold som kan ha verka på variasjonen, er tematikken i tekstane, sidan informantane har svart på tematisk ulike oppgåver. Mens tekstane i kontrollkorpuset berre spenner over fire ulike oppgåver, spenner S2-tekstane over 32. Det kunne ut ifrå det vore rimeleg å vente større grad av leksikalsk spreiding i S2-gruppa, men det er altså ikkje tilfellet.

Axelsson (1994: 117) finn at informantane hennar i all hovudsak nyttar adjektiv som er relativt frekvente i svensk, og Kotsinas (1982: 211) observerer det same om ord generelt. Det kan stemme med at innlærartekstane samla ser ut til å ha mindre leksikalsk breidde enn S1-tekstane. Teoretisk kan ein innanfor det bruksbaserte perspektivet sjå leksikalsk variasjon som eit teikn på kor godt utbygd og konsolidert nettverket av mentale representasjonar er. Nettverket blir bygd opp gjennom eksponering for innputt. I tråd med bevaringseffekten treng innlærarane store mengder autentisk innputt for å tileigne seg mindre frekvente konstruksjonar. Den avgrensa språklege erfaringa gjer at dei derfor ikkje har dei same føresetnadene for leksikalsk breidde som vaksne S1-brukarar. I studien min tyder altså den leksikalske variasjonen i tekstane på skilnader mellom S1- og S2-bruk.

7.3 Tverrspråkleg påverknad

Eg har samanlikna korleis S2-innlærere med tre ulike S1-bakgrunnar bruker adjektiv i tekstane sine. Den tyskspråklege gruppa har høgast frekvens av adjektivtoken. Skilnadene er statistiske mellom TY og kvar av dei to andre gruppene, men ikkje mellom SO og VI. Også når CEFR-nivået blir halde konstant, viser det seg at TY bruker fleire adjektiv enn dei to andre gruppene på B1-nivå. Likeeins bruker TY fleire adjektiv enn VI på utdanningsnivået «Høgskule/universitet». Dette er den einaste skilnaden som slår ut med utdanning som kontrollvariabel.

Om skilnadene mellom gruppene kjem av tverrspråkleg påverknad, kan det ha samband med den typologiske og genetiske avstanden mellom S1 og norsk. Oppfatta nærleik mellom S1 og S2 gir større sjanse for transfer og motsett mindre sjanse ved oppfatta avstand. Både leksikalsk og morfologisk har adjektivkategorien i norsk og tysk mykje til felles. Det svært nære sambandet mellom norsk og tysk, og dei mange kognata, gir grunn til å forvente at TY oppfatar likskapar mellom språka, noko som igjen seinkar terskelen for å overføre S1-kunnskapen til S2 (Berggreen & Tenfjord 1999: 210). Tilsvarende kan den typologiske avstanden i handsaminga av adjektivkategorien og ein liten grad av felles ordforråd gjere terskelen høg for SO og VI for å overføre sin leksikalske kunnskap frå S1. Skilnadene i adjektivfrekvens mellom TY på den eine sida og SO og VI på den andre kan i lys av dette tolkast som utslag av høvesvis positiv og negativ lingvistisk transfer. Sjølv om rammeverksnivå og utdanning kan vere faktorar som verkar parallelt med S1 og fører til ulik adjektivbruk i datasettet, er det mogleg at skilnadene kan tilskrivast tverrspråkleg påverknad.

7.4 Utdanning og adjektivbruk

Analysen avdekkjer ikkje statistiske samband mellom utdanningsbakgrunn og adjektivbruk, trass i at det ut frå dei deskriptive tala ser ut til å vere eit skilje mellom grunnskulegruppa på den eine sida og dei med lengre skulegang på den andre.

Den tidlegare forskinga konkluderer ulikt om samanhengen mellom skulegang og adjektivlæring i skandinaviske språk. Axelsson (1994: 147) konstaterer kort at utdanning ser ut til å vere ein av faktorane som verkar inn på læringa. Kotsinas (1982: 225) skriv at den lange skulegangen til den eine informanten kanskje «varit henne til hjelp» i språklæringa generelt, men ikkje spesifikt kopla til adjektiv. Svendsen (1996: 129) meiner på si side at utdanningsbakgrunn i liten grad har nokon forklaringsverdi for resultatane. Studiane gir altså sprikande signal om effekten av utdanning på adjektivlæring, og min eigen studie gir hel-

ler ikkje grunnlag for å hevde nokon samanheng. Det er heller ikkje gitt at ein eventuell utdanningsrelatert skilnad skulle ligge i variabelen tokenfrekvens. Han kunne vere relatert til andre aspekt ved adjektivbruken, som leksikalsk variasjon.

8 Sluttord

Målet med undersøkinga har vore å skaffe eit større empirisk grunnlag for korleis adjektiv blir brukt og lært i norsk som andrespråk hos vaksne. Funna kan skaffe innsikter som betrar utgangspunktet for å stille meir spissa hypotesar i framtida. Drøftinga har vist at trass i stor individuell variasjon, tilsvarer ikkje andrespråksbruken av adjektiv i det aktuelle materialet førstespråksbruken i kontrollkorpuset verken når det gjeld tokenfrekvens eller leksikalsk variasjon. Vidare ser førstespråket til innlærarane ut til å henge saman med tokenfrekvensen av adjektiv, mens bruken ikkje varierer nemneverdig etter rammeverksnivå (A2–B2) og utdanningsbakgrunn. Funna gir meining å tolke i lys av bruksbaserte perspektiv.

Kontrollkorpuset er lite og tematisk avgrensa i tillegg til at vi veit lite om tekstforfattarane. Som kjelde til innputt kunne det vere aktuelt å bruke større norskspråkelege korpus, eventuelt justert i frekvens for adjektiv som er mykje brukt i læremateriell eller relevante kontekstar. Å kunne sjå fleire faktorar i samanheng er ein fordel når ein undersøker komplekse fenomen som språklæring. Derfor bør ein ha eit datagrunnlag som i større grad gjer det metodisk mogleg. Vidare ville det vore ønskeleg å ha meir kontroll på temaa i tekstane og utdanningsbakgrunnen til informantane enn det ASK har. Til slutt vil longitudinelle data kunne gi eit utfyllande bilde av variasjon og systematikk over tid hos enkeltindivid eller grupper. Eit design som legg vekt på individuelle erfaringar og rammevilkår for språklæring over tid, vil vere særleg aktuelt ut frå eit bruksbasert perspektiv.

Vedlegg 1: Resultat av normalitetstesting

Testa med Shapiro-Wilk. Alle testresultata gjeld variabelen tokenfrekvens.

Gruppe	Test-statistikk	P-verdi	Konklusjon	
FORSKINGSPØRSMÅL 2				
SAMANLIKNING MED S1-BRUK				
Innlærartekstane	w = 0,996	p = 0,668	Dataa er normalfordelte	
Kontrolltekstane	w = 0,993	p = 0,858	Dataa er normalfordelte	
FORSKINGSPØRSMÅL 3				
CEFR-NIVÅ				
A2	w = 0,990	p = 0,412	Dataa er normalfordelte	
B1	w = 0,995	p = 0,913	Dataa er normalfordelte	
B2	w = 0,929	p = 0,130	Dataa er normalfordelte	
FORSKINGSPØRSMÅL 4				
UTDANNINGSNIVÅ				
Grunnskule	w = 0,979	p = 0,569	Dataa er normalfordelte	
Vidaregåande	w = 0,991	p = 0,748	Dataa er normalfordelte	
Høgskule /universitet	w = 0,988	p = 0,411	Dataa er normalfordelte	
FORSKINGSPØRSMÅL 5				
PÅVERKNAD FRÅ S1				
SO	w = 0,985	p = 0,317	Dataa er normalfordelte	
TY	w = 0,971	p = 0,027	Dataa er ikkje normalfordelte	
VI	w = 0,988	p = 0,521	Dataa er normalfordelte	
PÅVERKNAD FRÅ S1 – CEFR-NIVÅ SOM KONTROLLVARIABEL				
SO	A2	w = 0,976	p = 0,218	Dataa er normalfordelte
SO	B1	w = 0,992	p = 0,997	Dataa er normalfordelte
TY	A2	w = 0,767	p = 0,003	Dataa er ikkje normalfordelte
TY	B1	w = 0,984	p = 0,504	Dataa er normalfordelte
VI	A2	w = 0,987	p = 0,797	Dataa er normalfordelte
VI	B1	w = 0,962	p = 0,141	Dataa er normalfordelte
PÅVERKNAD FRÅ S1 – UTDANNING SOM KONTROLLVARIABEL				
SO	Grunnskule	w = 0,981	p = 0,852	Dataa er normalfordelte
	Vidaregåande	w = 0,968	p = 0,397	Dataa er normalfordelte
	Høgskule /universitet	w = 0,970	p = 0,591	Dataa er normalfordelte
TY	Grunnskule	w = 0,900	p = 0,384	Dataa er normalfordelte
	Vidaregåande	w = 0,981	p = 0,719	Dataa er normalfordelte
	Høgskule /universitet	w = 0,952	p = 0,149	Dataa er normalfordelte
VI	Grunnskule	w = 0,844	p = 0,031	Dataa er ikkje normalfordelte
	Vidaregåande	w = 0,973	p = 0,642	Dataa er normalfordelte
	Høgskule /universitet	w = 0,974	p = 0,322	Dataa er normalfordelte

Litteraturliste

- Axelsson, Monica. 1989. «Vuxna SV2-inlärare lär sig adjektiv.» *Andra symposiet om svenska som andraspråk*, Göteborg.
- Axelsson, Monica. 1994. Noun phrase development in Swedish as a second language : a study of adult learners acquiring definiteness and the semantics and morphology of adjectives. Doktorgradsavhandling, Stockholm: Centrum för tvåspråkighet, Stockholms universitet.

- Berggreen, Harald & Kari Tenfjord. 1999. *Andrespråkslæring*. Oslo: Ad notam Gyldendal.
- Bybee, Joan. 2008. Usage-based grammar and second language acquisition. I Robinson, Peter J. & Nick C. Ellis (red.): *Handbook of cognitive linguistics and second language acquisition*. New York: Routledge, s. 216–236.
- Carlsen, Cecilie (red.). 2012. *Norsk profil: det felles europeiske rammeverket spesifisert for norsk: et første steg*. Oslo: Novus forlag.
- Ellis, Nick C. 2006. Cognitive perspectives on SLA: The associative-cognitive CREED. Themes in SLA Research: *AILA Review* 19, 100–121.
- Ellis, Nick C. & Stefanie Wulff. 2014. «Usage-based approaches to SLA.» I VanPatten, Bill & Jessica Williams: *Theories in second language acquisition: an introduction*. Florence, KY: Routledge, 75–93.
- Europarådet 2001. *Common European Framework of Reference for Languages: Learning, teaching and assessment*. Cambridge: Cambridge University Press.
- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Field, Andy. 2009. *Discovering Statistics using SPSS (and Sex and Drugs and Rock 'n' roll)* (3. utg.). Los Angeles: SAGE.
- Golden, Anne. 2014. *Ordforråd, ordbruk og ordlæring*. 4. utg. utg. Oslo: Gyldendal akademisk.
- Golden, Anne, Scott Jarvis & Kari Tenfjord (red.). 2017. *Crosslinguistic influence and distinctive patterns of language learning : findings and insights from a learner corpus*. Bristol: Multilingual Matters.
- Gujord, Ann-Kristin Helland & Silje Ragnhildstveit. 2018. Tverrspråkleg påverknad – Status og trendar i norsk andrespråksforskning. I Gujord, Ann-Kristin H. & Gunhild Randen (red.): *Norsk som andrespråk – perspektiver på læring og utvikling*. Oslo: Cappelen Damm Akademisk, 133–164.
- Gujord, Ann-Kristin Helland. 2017. «Hos en voksen er tanken helt full av masse ting». Personlege narrativar i Norsk andrespråkskorpus (ASK) – ei kjelde til innsikt i læringsprosessen? *Acta Didactica Norge* 11-1, 7–22.
- Hedges, L. V. 1981. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6, 107–128
- Husby, Olaf. 2017. *Innvandrerens morsmål: En ressursbok for lærere*. Bergen: Fagbokforlaget.
- Janik, Marta Olga. 2015. *Adjektivbøyning i norsk – kongruenstilegnelse hos polskspråklige norskinnlærere*. Doktorgradsavhandling, Institutt for skandinaviske studier, Adam Mickiewicz-universitetet i Poznan.

- Jarvis, Scott. 2000. Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language learning* 50-2, 245–309.
- Jarvis, Scott & Aneta Pavlenko. 2008. *Crosslinguistic Influence in Language and Cognition*. New York og London: Routledge.
- Johannessen, Janne Bondi; Hagen, Kristin, Lynum André & Nøklestad, Anders. 2012. OBT+stat. A combined rule-based and statistical tagger. I Andersen, Gisle (red.): *Exploring Newspaper Language. Corpus compilation and research based on the Norwegian Newspaper Corpus*. John Benjamins Publishing Company, 51–65.
- Kotsinas, Ulla-Britt. 1982. *Svenska svårt: Några invandrades svenska talspråk*. MINS, Stockholms universitet.
- Larson-Hall, Jenifer. 2010. *A guide to doing statistics in second language research using SPSS*. Second language acquisition research series: theoretical and methodological issues. New York/London: Routledge.
- Llach, María Pilar Agustín. 2011. *Lexical Errors and Accuracy in Foreign Language Writing*. Bristol/Buffalo/Toronto: Multilingual Matters.
- Meurer, Paul. 2012. *ASK – Norsk andrespråkskorpus*. Bergen: Uni Computing. Lokalisert 2019 på <http://clarino.uib.no/ask/ask>
- Mitchell, Rosamond, Florence Myles & Emma Marsden. 2013. *Second language learning theories*. 3. utg. London/New York: Routledge.
- Nilsen, Gølin Christine Kaurin. 2018. Substantivfrasen i norsk som andrespråk: Fra A1 til B2. ILN, UiO; 2018-10-11 – 2018-10-12.
- Nistov, Ingvild, Hana Gustafsson, Teresa Cadierno. 2018. Bruksbaserte tilnærminger til andrespråksspråklæring. I Gujord, Ann-Kristin H. & Gunhild Randen (red): *Norsk som andrespråk – perspektiver på læring og utvikling*. Oslo: Cappelen Damm Akademisk, 107–132.
- Rosén, Victoria. 1999. *Vietnamesisk: en kontrastiv og typologisk introduksjon*. Trondheim: Tapir.
- Saeed, John Ibrahim. 1993. *Somali reference grammar*. 2. utg. Kensington, Md: Dunwoody Press.
- Sveen, Andreas, Hanne Gram Simonsen & Rolf Theil. 2000. *Innføring i lingvistikk*. 2. utg. Oslo: Universitetsforlaget.
- Svendsen, Bente Ailin. 1996. *Adjektiv i norsk som andrespråk: En frekvensbasert leksikalsk-semantic tværnsnittstudie av adjektivtilegnelse hos voksne innlærere*. Hovedoppgåve, Institutt for nordistikk og litteraturvitenskap, Universitetet i Oslo.

- Tekstlaboratoriet. 2010. *Oslo-Bergen-taggeren*. Universitetet i Oslo. Lokalisert 2019 på <http://www.tekstlab.uio.no/obt-ny/index.html>
- Tenfjord, Kari, Jon Erik Hagen & Hilde Johansen. 2009. Norsk andrespråkskorpus (ASK) – design og metodiske forutsetninger. NOA. Norsk som andrespråk 25-1, 52-81.
- van de Craats, Ineke, Jeanne Kurvers & Martha Young-Scholten. 2005. Research on low-educated second language and literacy acquisition. *Proceedings of the Inaugural Symposium*. Tilburg, Utrecht.

Summary

This article aims to explore empirically some aspects of the acquisition and use of adjectives in adult learners of Norwegian as a second language. It is a quantitative cross-sectional study of the adjective use of texts written by 300 learners in a learner corpus of Norwegian (Norsk andrespråkskorpus, ASK). I have compared the learner data to first language texts. In addition, I have examined the connection between the use of adjectives and some background variables. Guided by a usage-based approach, frequency was chosen as the measure of adjective use. Firstly, adjectives are used less frequently and less varied in the learner data than in the first language data. Secondly, there are group differences based on three different first language backgrounds (German, Somali, and Vietnamese) that indicate cross-linguistic influence. Proficiency level and educational background, on the other hand, do not seem affect the frequency of adjectives. Individual variation is widespread throughout the data, a finding that is common in second language research and corresponds well with the theoretical framework of the current study.

Anna-Marie Kjøde Olsen
Universitetslektor ved Norskurska
Institutt for lingvistiske, litterære og estetiske studier
Universitetet i Bergen
Postboks 7805
N-5020 Bergen
anna-marie.olsen@uib.no



Syntaktisk kompleksitet på tvers av rammeverksnivå

Ann-Kristin Helland Gujord

Universitetet i Bergen

Artikkelen presenterer ein studie av samanhengen mellom syntaktisk kompleksitet og ferdigheitsnivå basert på 120 tekstar fordelte på fire ferdigheitsnivå (A2–C1) etter *Det felles europeiske rammeverket for læring, undervisning og vurdering av språk* (Utdanningsdirektoratet 2011). Utgangspunktet for analysen er CAF-tradisjonen (Complexity, Accuracy, Fluency); særleg funn i tre tidlegare norske studiar av syntaktisk kompleksitet og korrektheit (Moe 2013, Berggreen og Sørland 2016, Sørensen 2017). Ved hjelp av ulike statistiske metodar finn eg at det er ein signifikant *positiv* samanheng mellom setningslengd og rammeverksnivå, fraseutbygging og rammeverksnivå og mellom korrektheit og rammeverksnivå, og at det er ein signifikant *negativ* samanheng mellom rammeverksnivå og høvesvis sidestilt koordinasjon og underordning. Eit viktig funn er at modellen først og fremst estimerer sannsynet for plassering på nivå B1 og B2. Eit unntak er målet på korrektheit, som også skil godt mellom dei øvste nivåa (B2 og C1).

Nøkkelord: syntaktisk kompleksitet, CAF-tradisjonen, ferdigheitsnivå (CEFR), skriftlege andrespråkstekstar, ordinal logistisk regresjon

Innleiing

Denne artikkelen presenterer ein studie av samanhengen mellom syntaktiske trekk i skriftlege innlærartekstar og ferdigheitsnivået tekstane er plasserte på etter *Det felles europeiske rammeverket for språk* (Utdanningsdirektoratet 2011, heretter Rammeverket). Studien er forankra i forskinga på kompleksitet innanfor CAF-tradisjonen, det vil seia, forskinga på tre dimensjonar ved andrespråkutvikling og -prestande; *kompleksitet* (‘complexity’), *korrektheit* (‘accuracy’) og *flyt* (‘fluency’). I denne studien er det aspekt ved den syntaktiske

kompleksiteten i tekstane som blir undersøkt, med eit blick også på syntaktisk korrektheit. Det har vore forska mykje på syntaktiske trekk i andrespråkstekstar i norsk innlærarspråk; i følge Golden og Hvistendahl (2010) er syntaks det trekket som oftast er blitt studert i studiar av språknivået i den norske forskinga på andrespråksskriving. Ein del av denne forskinga er oppsummert i Jensen (2018a), og Eide (2015) samlar fleire arbeid på syntaks i norsk innlærarspråk innanfor ei generativ tilnærming. I all hovudsak undersøker desse studiane enkeltfenomen (td. subjekt/verbal-inversjon, subjektstvang, negasjon, finittheit); så langt finst det berre tre norske studiar som kan relaterast til CAF-tradisjonen (Moe 2013, Berggreen & Sørland 2016, Sørensen 2017). Dette står i motsetnad til den internasjonale forskinga på andrespråksskriving ('L2 writing') der CAF-inspirerte studiar utgjer ein stor del av forskinga (jf. td. Norris & Ortega 2009, Ortega 2003, 2015, Housen, Kuiken & Vedder 2012).

Føremålet i denne studien er først og fremst å skaffa fram meir empirisk basert kunnskap om syntaktisk kompleksitet i norsk innlærarspråk. Det gjer eg ved å undersøkje samanhengen mellom mål på syntaktisk kompleksitet og korrektheit og ferdighetsnivå basert på 120 tekstar på fire ulike ferdighetsnivå i Rammeverket (A2–C1). Tekstane er svar på ein offisiell prøve i norsk for vaksne innvandrarar og er anten henta ut frå Norsk andrespråkskorpus (ASK) eller arkivet til Norsk språkttest (Folkeuniversitetet). Det er også eit mål å bidra inn i metodeutviklinga innanfor den norske kvantitative andrespråksforskinga. I følge Jensen har mange av studiane av syntaks i norsk innlærarspråk «svakheter i sine kvantitative tilnærmingar» (2018a: 254). Det er ikkje særskilt for syntaksforskinga. Dei seinare åra har det både i det internasjonale og nasjonale fagfeltet vore retta større merksemd mot bruk av kvantitative metodar (td. Jensen 2018b, Plonsky & Ghanbar 2018). I denne studien brukar eg ein multivariat statistisk metode, ordinal logistisk regresjon, for å utforska samanhengen mellom fleire uavhengige variablar (syntaktiske kompleksitetsmål, korrektheitsmål) og ein avhengig variabel (ferdighetsnivå). Eg presenterer først CAF-tradisjonen med fokus på forskinga på syntaktisk kompleksitet, særleg tre norske studiar. Eg gjer så greie for metode og data før eg presenterer resultatata av analysane. Siste del av artikkelen er ein diskusjon av funna.

2 Syntaktisk kompleksitet i andrespråkstekstar

2.1 CAF-tradisjonen

Eit viktig utgangspunktet for denne studien er forskinga på kompleksitet innanfor CAF-tradisjonen, der tre dimensjonar ved andrespråksutvikling og -premanse blir utforska: kompleksitet, korrektheit, flyt. Denne tradisjonen har røter tilbake til andrespråksforskinga som gjekk føre seg innanfor to område på 1970-talet (Housen et al. 2012, Michel 2017). Inspirert av forskinga på førstespråksutvikling, som mellom anna etablerte gjennomsnittleg ytringslengde ('MLU – mean length of utterances', td. Bellugi & Brown 1964) som mål på talespråksutvikling hos barn, og gjennomsnittleg setningslengde ('T-unit', Hunt 1965) som mål på skriveutvikling hos skulebarn, vart det sett i gang eit arbeid for å identifisera mål for utvikling av *andrespråks*ferdigheiter. Larsen-Freeman (td. 1978) var blant dei første som forsøkte å utvikla ein indeks som kunne måla globale og ikkje språkspesifikke andrespråksferdigheiter, slik at det ville vera mogleg å samanlikna studiar på tvers av populasjonar og språk. Ho fann mellom anna at mengd feilfrie heilsetningar var eit mål som skilde bra mellom utviklingsnivå, og som fanga opp aspekt ved både kompleksitet og korrektheit (Berggreen & Sørland 2016). Om lag på same tida, innanfor den meir pedagogisk orienterte andrespråksforskinga, utvikla ein måtar å karakterisera flyt, og ikkje berre korrektheit på, for å kunne fanga korleis *kommunikative* ferdigheiter utvikla seg på andrespråket (jf. Housen et al. 2012, Michel 2017). Fram mot 1990-talet vaks distinksjonane mellom dei tre dimensjonane kompleksitet, korrektheit og flyt tydeleg fram og vart samanstilte for første gong av Skehan (td. 1998), som eit rammeverk for å fanga inn og vurdera sentrale komponentar ved andrespråksutvikling og andrespråksferdigheiter. Den overordna, teoretiske forståinga av dei tre CAF-konstrakta som etablerte seg på denne tida, er framleis eit felles utgangspunkt for forskarar som arbeider innanfor CAF-tradisjonen, som Housen et al. (2012) beskriv slik:

[...] complexity is commonly characterized as the ability to use a wide and varied range of sophisticated structures and vocabulary in the L2, accuracy as the ability to produce target-like and error-free language, and fluency as the ability to produce the L2 with native-like rapidity, pausing, hesitation, or reformulation. (Housen et al. 2012: 2).

Det er ei generell semje blant andrespråksforskarar at CAF-triaden er ein fornuftig og valid måte å undersøkje og beskriva andrespråkspremanse og and-

respråksutvikling på (Michel 2017). I dette ligg det også ein felles tanke om at desse tre dimensjonane kan karakterisera performanse (som er) på ulike steg i utviklinga. Innlærarar med meir avanserte ferdigheiter på andrespråket, eller innlærarar som har kome lenger i utviklinga av andrespråket, vil i samanlikning med innlærarar med meir grunnleggande andrespråksferdigheiter, eller i samanlikning med seg sjølv på eit tidlegare steg i utviklinga, ha fleire komplekse og korrekte grammatiske strukturar i språket sitt, ha eit meir avansert ordforråd og kunne snakka og skriva andrespråket med større flyt (Michel 2017). Samstundes finst det også kontroversar og uavklarte spørsmål innanfor CAF-forskinga. Housen et al. (2012) peikar ut fem problemområde: (1) Sjølv om CAF-studiane spring ut frå ei generell forståing av kva aspekt ved språkbruk og utvikling dei tre dimensjonane fangar opp, er det eit gjennomgåande problem i forskinga at mange studiar ikkje gir meir presise definisjonar på dimensjonen(ane) dei studerer. Manglande operasjonalisering av konstrakta gjer det vanskeleg å samanlikna funn på tvers av studiar. Kompleksitet er den dimensjonen som er mest vagt og inkonsistent definert i litteraturen (Housen et al. 2012). (2) Ei anna utfordring er å forklara korleis dei tre dimensjonane er kopla til kognitive, lingvistiske og psykolingvistiske prosessar. Dei er forsøkt forklart ved ulike underliggjande endringsprosessar i språkutvikling. Til dømes vil internalisering av nye mønster i andrespråket gjera at språket vil bli meir *komplekst* i betydinga meir utvikla (for eksempel meir bruk av underordning og eit vidare og meir nysansert ordforråd), andrespråket blir meir *korrekt* når innlæraren sin kunnskap blir restrukturert, og andrespråket vil bli produsert med større *flyt* når innlæraren utviklar prosedurale kunnskapar i andrespråket (Housen et al. 2012). (3) Ei tredje utfordring er at det slett ikkje er klart korleis dei tre dimensjonane, og deira tilhøyrande subkomponentar er relaterte til kvarandre i andrespråksproduksjonen og i sjølve utviklingsprosessen, til trass for at det i litteraturen blir understreka at CAF-dimensjonane ikkje er eindimensjonale. Ei rekkje forskarar har i nyare tid hevda at dimensjonane interagerer på komplekse måtar, og at interaksjonen mellom dei somme tider er støttande og somme tider konkurrerende (jf. til dømes Bulté & Housen 2018, Larsen-Freeman 2006, Verspoor, Lowie, Chan & Vahtrick 2017). Desse studiane har også ofte eit større fokus på individuelle utviklingsprofilar enn det som før har vore vanleg i CAF-forskinga (td. Vyatkina, Hirschmann & Golcher 2015). (4) Den fjerde utfordringa har samanheng med den manglande definisjonen og operasjonaliseringa av CAF-konstrakta: Det er ein enorm variasjon i kompleksitetsmåla som blir brukte i studiar. Bulté og Housen (2012) rapporterte om ikkje mindre enn 40 ulike kompleksitetsmåla basert på ein metaanalyse av 40 empiriske studiar mellom 2005 og

2008. Dette representerer ei stor utfordring med omsyn til samanlikning av studiar, reliabilitet og validitet (Housen et al. 2012). (5) Til slutt; ein har lite kunnskap om kva for eksterne faktorar som påverkar utviklinga av CAF-dimensjonane. På dette området skjer det ein del forskning, til dømes studiar av effekt av undervisning, L1-bakgrunn, sjanger og modalitet (jf. Ortega 2015).

2.2 Syntaktisk kompleksitet

I denne studien er det i hovudsak kompleksitetsdimensjonen av CAF-triaden som blir undersøkt, til ein viss grad også korrektheitsdimensjonen. Det finst ikkje éin akseptert definisjon av kompleksitet, men termen kan romma ulike betydingar.¹ Eg tek utgangspunkt i Bulté & Housen (2012) sin taksonomi over kompleksitetsomgrepet og deira definisjonar og avgrensingar av ulike former for kompleksitet. Eit viktig poeng hos Bulté og Housen er at kompleksitetskonstrakta må avklarast og definerast på fleire nivå. Den teoretiske forståinga av konstrakta som skal undersøkjast må vera avklarte, i tillegg til korleis kompleksiteten kan observerast i språket og korleis han kan målast. Eit viktig skilje går mellom relativ eller kognitiv kompleksitet og absolutt kompleksitet (sjå også Berggreen & Sørland 2016). Førstnemnde tar utgangspunkt i språkbrukaren og definerer kompleksitet som «the mental ease or difficulty with which linguistic items are learned, processed or verbalized» (Bulté & Housen 2012: 23). Absolutt kompleksitet derimot, som er utgangspunktet i denne studien, er definert i høve til språkssystemet «som enheter som er mer omfattende (antall delkomponenter) og med mer sammensatte relasjoner til andre enheter, eller leksikalsk» (Berggreen & Sørland 2016: 33). Denne studien er avgrensa til den grammatiske komponenten av språkssystemet, ikkje den leksikalske, og det er *syntaktiske* aspekt ved den grammatiske kompleksiteten som er i fokus, det vil seia, syntaktisk kompleksitet (i motsetnad til morfologisk kompleksitet). Den syntaktiske kompleksiteten i dei skriftlege andrespråkstekstane blir observert gjennom setningslengd, koordinasjon, underordning, utbygging av substantivfrasar, korrektheit og gjennom variasjon i syntaktisk funksjon og val av subjunksjon. Den syntaktiske kompleksiteten er operasjonalisert gjennom ti målbare variablar, der åtte av dei er på skalanivå (jf. del 2).

I følge Ortega (2012) blir kompleksitet utforska innanfor tre område med ulike forskingsføremål. Kompleksitet blir nytta til å bestemma ferdigheitsnivå

1. Sjå til dømes Bulté og Housen (2012) og Pallotti (2015) for grundige diskusjonar om kompleksitetsomgrepet.

i L2 ved å undersøkje i kva grad kompleksiteten i andrespråket aukar i takt med at innlæraren når høgare ferdigheitsnivå. Ein annan type forskning på kompleksitet er knytt til forskinga på oppgåvebasert språkundervisning ('task-based language teaching'). Slike studiar har som mål å beskriva performansen for å undersøkje om og korleis kompleksiteten i andrespråket varierer systematisk med dei kognitive krava oppgåvene stiller innlæraren overfor. Det tredje føremålet ved å måla kompleksitet i innlærarspråk er å identifisera referansepunkt for nivå i utviklinga av andrespråket. Dersom ein antar at språket blir meir kompleks etter kvart, er det gunstig å identifisera ein kompleksitetsindeks som kan brukast for å skilja mellom utviklingsnivå. Forskinga innanfor alle desse områda har i all hovudsak skjedd basert på mål på syntaktisk og leksikalsk kompleksitet (Bulté & Housen 2014, De Clercq & Housen 2019).

I denne studien er det samanhengen mellom syntaktisk kompleksitet og ferdigheitsnivå som er i fokus, og studien sorterer slik først og fremst under det første området Ortega (2012) skisserer. Metastudier av forskinga på syntaktisk kompleksitet (Norris & Ortega 2009, Wolfe-Quintero et al. 1998) slår fast at syntaktisk kompleksitet korrelerer relativt sterkt med ferdigheitsnivå, og skisserer ei utvikling der den syntaktiske kompleksiteten i innlærarspråket utviklar seg frå koordinasjon til underordning og til utbygging av frasekompleksitet. Basert på døme hos Kuiken og Vedder (2019: 195) kan stega illustrerast slik på norsk:

- Steg 0: Eg har ein son. Han er 12 år gammal ('I have a son. He is 12 years old').
- Steg 1: Eg har ein son og han er 12 år gammal. ('I have a son and he is 12 years old').
- Steg 2: Eg har ein som er 12 år gammal. ('I have a son who is 12 years old.')
- Steg 3: Min 12-år gamle son. ('My 12-year-old son.')

Kuiken og Vedder presiserer at desse stega ofte overlappar i utviklinga, at utvikling ikkje er lineær innanfor alle områda, men at nokre av dei syntaktiske måla i somme studiar har vist ei u-forma utviklingskurve. Kuiken og Vedder (2019) viser også til studiar som dokumenterer individuelle utviklingsprofilar som ikkje samsvarar med tendensane på gruppenivå.

Dei mest nytta måla på syntaktisk kompleksitet har vore lengdebaserte mål, som har hatt den metodiske utfordringa ved seg ved at dei er hybridmål, det vil seia mål som κløper over og kvantifiserer noe som er svært komposisjo-

nelt» (Berggreen & Sørland 2016: 34). Til dømes vil lengdemål for heilsetningar også innehalda informasjon om fraselengd og tal på leddsetningar. Litteraturen tilrår difor at ein i tillegg bør ha andre syntaktiske mål som seier noko om andre dimensjonar av den syntaktiske kompleksiteten (Norris & Ortega 2009), for eksempel målingar av frasekompleksitet (Berggreen & Sørland 2016, Bulté & Housen 2012).

2.3 *Forskning på syntaktisk kompleksitet i norske innlærardata*

Det finst tre tidlegare studiar av syntaktisk kompleksitet i norsk innlærarspråk.² Basert på data frå Norsk andrespråkskorpus (ASK) undersøkte Moe (2013) om syntaksen i skriftlege andrespråkstekstar på tre nivå skilde seg i samsvar med Rammeverket si beskriving av ferdigheiter i syntaks på nivå A2, B1 og B2. Berggreen og Sørland (2016) studerte utvikling av syntaktisk kompleksitet over tid med butid som avhengig variabel i eit skriftleg materiale. Tekstane er skrivne over tre år av elevar som var på 5. og 6. trinn ved første innsamlingspunkt og hadde budd 1–2 år i Noreg (Berggreen & Sørland 2016). Sørensen (2017) er ei undersøking av syntaktisk kompleksitet i skriftlege tekstar på ulike ferdigheitsnivå etter Rammeverket skrivne av vaksne innlærarar. Ho undersøkte skriftlege tekstar produserte som del av eit prøvesvar (Norskprøven). Moe (2013), Berggreen og Sørland (2016) og Sørensen (2017) analyserer og operasjonaliserer kompleksitet delvis på same måte, to av studiane ser også på korrektheit. Tabell 1 gir ei oversikt over data, variablar og metodisk tilnærming i desse studiane.

2. Kompleksitetsomgrepet vart nok først nytta i hovudoppgåva til Mangerud (1988), der ho mellom anna såg på underordning i ein studie av mellomspråka til fire elevar med tyrkisk som førstespråk.

Tabell 1. Data, variablar og metodisk tilnærming i tre tidlegare norske L2-studiar av kompleksitet.

	Moe (2013)	Berggreen og Sørland (2016)	Sørensen (2017) ³
Data	Tverrsnittsdata, skriftlege tekstar frå ASK, A2 (247), B1 (567), B2 (375)	Lengdesnittsdata over 3 år, 221 skriftlege tekstar, 19 elevar på mellomtrinnet	Tverrsnittsdata, skriftlege tekstar frå ASK, A2 (15), B1 (15), B2 (15)
CAF-dimensjon	Kompleksitet (C) Korrektheit (A)	Kompleksitet (C)	Kompleksitet (C) Korrektheit (A)
CAF-mål	1. Heilsetningslengd 2. Underordning 3. Variasjon i leddsetningstype 4. Inversjonsfeil 5. Feil plassering av adverbial i leddsetningar	1. Heilsetningslengd 2. Underordning 3. Koordinasjon 4. Substantivfraseomfang	1. Heilsetningslengd 2. Underordning 3. Koordinasjon 4. Substantivfraseomfang 5. Koordinasjon m/elidering 6. Syntaktiske ukorrekte heilsetningar
Metode	Anova, post hoc-testing med parvise t-testar (mål 4 og 5 er ikkje signifikantesta).	Anova, post hoc-testing med parvise t-testar med justert signifikansnivå for å motverka gruppevis feilrate ⁵ .	Kuskal-Wallis, post hoc-testing med Mann-Whitney U-test.

I litteraturen er lengdemål rekna som gode generelle mål på kompleksitetsutvikling innanfor det grammatiske området og på lingvistiske ferdigheiter (td. Bulté & Housen 2012, Norris & Ortega 2009). Funna i dei norske studiane viser også at det mest brukte lengdemålet i CAF-studiar, *heilsetningslengd* (løpeord pr. heilsetning), stig jamt med aukande nivå/eksponeringstid. Medan Moe (2013) finn gjennomgåande signifikante resultat⁵, er denne skilnaden berre signifikant mellom første- og andreårstekstar i Berggreen og Sørland (2016) sin studie, og Sørensen (2017) finn berre at A2 har signifikant kortare setningar enn dei andre nivåa. Litteraturen viser til litt motstridande resultat når det gjeld *koordinasjon* som kompleksitetsmål (Kuiken & Vedder 2019). Somme har funne at koordinasjonsgraden aukar for så å falla med stigande ferdigheitsnivå, somme finn ingen signifikante endringar, medan somme har dokumentert ei utviklingskurve for koordinasjon som liknar ein u opp ned («omega-shaped»), Kuiken & Vedder 2019: 195). Dei to norske studiane som inkluderer koordinasjon (koordinerte heilsetningar pr. heilsetning), får også ulike resultat: Berggreen og Sørland observerer ein auke gjennom alle tre åra, og ein signifikant auke mellom første- og tredjeårstekstar, medan Sørensen finn ein auke i koor-

-
- Sørensen (2017) studerte både kortsvarsoppgåva og langsvarsoppgåva som kandidatane hadde svara på under prøven. Eg forheld meg berre til resultatata frå langsvarsoppgåva, då både Moe (2013) og eg berre har hatt tilgang til langsvarstekstane.
 - Sjå Jensen (2020) for ei forklaring på faren for type I-feil (falske positive funn) som oppstår, dersom ein gjennomfører fleire testar samstundes utan å senka signifikansnivået.
 - Merk at Moe (2013) samanliknar A1 mot B1 og B1 mot B2 (og ikkje A2 mot B2).

dinasjonsgraden frå lågast nivå til mellomnivå og deretter eit fall på høgare nivå (berre B2 som skil seg signifikant frå dei andre nivåa).

Dei fleste studiane av *underordning* viser også ulike resultat, sjølv om fleirtalet finn ein auke frå begynnarnivå til mellomnivå og eit fall frå mellomnivå til høgare nivå (Kuiken & Vedder 2019). Det er også ein del diskusjon om korleis ein best kan måla underordning i studiar av grammatisk kompleksitet (jf. Bulté & Housen 2012). Alle dei tre norske studiane måler grad av underordning på same måte og i tråd med den tradisjonelle måten underordning blir målt på, det vil si antall leddsetningar pr. heilsetning (Norris & Ortega 2009). Alle observerer at underordningsgraden stig med aukande nivå/eksponeringstid, men det er berre Moe som finn gjennomgåande signifikante skilnader. Skilnaden i grad av underordning er berre signifikant mellom første- og tredjeårstekstar i Berggreen og Sørland sin studie, og Sørensen finn at berre A2 og B2 skil seg signifikant frå kvarandre.

Litteraturen peikar på utfordringar i operasjonaliseringa av målingar av *frasekompleksitet*, og at det difor er vanskeleg å samanlikna funn på tvers av studiar (Kuiken & Vedder 2019). Berggreen, Sørland og Sørensen måler alle fraseutbygging på same måte, ved å telja talet på ord i den maksimale substantivfrasa i kvar tekst, og observerer same tendens til at substantivfrasane blir lengre og meir komplekse ved aukande eksponering og nivå. Igjen er det berre første- og tredjeårstekstar i studien til Berggreen og Sørland, og A2- og B1-tekstane i Sørensen sin studie, som viser signifikante skilnader.

Det knyter seg også diskusjonar til målingar av *korrektheit*, til dømes knytt til kriterium for å identifisera avvik, og om ein bør skilja mellom avvik som er «more or less deviant or erroneous» (Bulté & Housen 2012: 4). Moe og Sørensen får signifikante resultat som peikar mot at grad av syntaktiske feil er eit mål som skil godt mellom nivå. Hos Sørensen minkar mengda syntaktisk ukorrekte heilsetningar signifikant mellom alle tre nivåa. Dette samsvarar med Eriksson og Carlsen (2013) sin deskriptive studie av fordelinga av grammatiske feilkodar i ASK på rammeverksnivå (tal på feilkodar delt på tal på ord på det enkelte rammeverksnivå, s. 254). Resultata antyd at talet på syntaktiske feil fell med stigande rammeverksnivå, sjølv om resultatet må tolkast med forsiktighet sidan analysen berre byggjer på gruppeverdiar utan at det er teke omsyn til den individuelle variasjonen. Moe (2013) finn også at variasjonen i type leddsetning aukar signifikant med stigande nivå. Basert på korrespondanseanalyse finn Carlsen (2013) at tekstar på dei høgare nivåa har fleire ulike typar bindeord,

fleire av dei subjunksjonar, enn tekstar på dei lågare nivåa, og at skiljet ser ut til å gå mellom B1+⁶.

Basert på den internasjonale litteraturen og funn i analysar av norske innlærarspråk (Moe 2013, Berggreen og Sørland 2016, Sørensen 2017) formulerer eg tre forskingsspørsmål, som til saman kan kasta lys over det overordna spørsmålet om i kva grad andrespråkstekstar vurderte til å vera på ulike ferdigheitsnivå, skil seg systematisk frå kvarandre i syntaktisk kompleksitet og korrektheit:

1. Aukar setningslengda, underordningsgraden, substantivfrasetbygginga, grad av korrektheit og type av subjunksjonar mellom rammeverksnivåa A2–C1?
2. Søkk koordinasjonsgraden mellom rammeverksnivåa A2–C1?
3. Skil tekstar på ulike rammeverksnivå seg med omsyn til frekvens av leddsetningstype?

Samanlikning av gruppeverdiar, som så langt har vore nytta i dei norske studiane, kan ikkje seia noko om og i kva grad kompleksitetsmåla er relaterte til kvarandre. Eit tilleggsmål i denne studien er difor å analysera syntaktisk kompleksitet ved hjelp av ein ny statistisk metode.

3 Metode

3.1 Data og datahandsaming

Tekstane til 113 av informantane er elisiterte frå *ASK Norsk andrespråskorpus* (Meurer 2012)⁷. Det er få tekstar på C1-nivå i ASK, og for å få like mange tekstar på alle nivåa (30 per nivå), fekk eg sju C1-tekstar frå arkivet frå Norsk språkttest (Folkeuniversitetet).⁸ Alle tekstane er vurderte til å vera på eitt av fire ulike ferdigheitsnivå i norsk (A2–C1) etter Rammeverket. Alle er svar på ein del av ein offisiell prøve i norsk for vaksne innvandrarakar (*Språkprøven i norsk for vaksne innvandrarakar* eller *Test i norsk høgare nivå*, heretter kalla Språkprøven og Bergenstesten). Språkprøve-svara er anten forteljande (ofte personlege nar-

6. Carlsen studerer tekstar frå A2–C1 og nyttar mellomnivå: A2, A2+, B1, B1+, B2, B2+ og C1.

7. Vedlegg 1 gir oversikt over pid-nummeret til tekstane eg har henta frå ASK, som kan søkjast opp i sin heilskap.

8. Tekstane frå Norsk språkttest er av same type språk- og persondata som Norsk språkttest tidlegare har levert til ASK.

rativar) eller utgreiande, medan tekstane skrivne til Bergenstesten ofte også er argumenterande. Tekstane i ASK handlar også om ulike tema.⁹ Det er også skilnader i lengde på tekstane på ulike nivå. Gjennomsnittsverdien av tal på ord i tekst stig mellom nivåa (A2 198.2, B1 277.1, B2 437.5, C1 487.5). Spennet mellom tekst med minst og flest ord er endrar seg også mellom nivåa (A2 121–301, B1 173–448, B2 222–860, C1 355–803). ASK er ikkje syntaktisk tagga, og alle dei 120 tekstane har vore manuelt analyserte av to kodarar.

I analysen vart tal på løpeord i tekstane talde opp, og tekstane vart inndelte i heilsetningar, som svarar til det som blir omtala som ‘t-units’¹⁰ i den internasjonale litteraturen, dvs. heilsetningar med eventuelt underordna setningar (Berggreen & Sørland 2016: 37). I kvar enkelt tekst er alle koordinerte heilsetningar, leddsetningar med tilhøyrande syntaktisk funksjon, maksimal substantivfrase, subjunksjon og syntaktiske avvik registrerte. I tillegg til brot på normert norsk leddstilling omfattar syntaktiske avvik i denne studien også brot på kravet i norsk om at setningar må ha eit finitt verb. Årsaka er at meistring av skiljet mellom finitte og infinitte verbformer i norsk, krev innsikt i det syntaktiske fenomenet finittheit (Eide 2009, Gujord 2017, Hagen 2001/2013). I tabell 2 er framgangsmåten eksemplifisert og kommentert:

Tabell 2. Uavhengige variablar i analysen.

Analysekategori	Beskriving
Løpeord	Løpeord som ikkje inngår i ei heilsetning, er stroke (td. spørjepronomenet i <i>Noen barn er veldig uhøflig til gamle mennesker. Hvorfor? Foreldrene må snakke med sine barn om alt, s0729</i>). Sitat frå oppgåveteksten eller tekstvedlegg er også stroke. Forkortingar blei talde som eitt løpeord.
Heilsetningar	Teiknsetting og bruken av stor og liten bokstav er ignorert i inndelinga i heilsetningar. Fire heilsetningar er identifiserte i dette utdraget: (1) <i>Da jeg var liten ville jeg bli en sykepleier.</i> (2) <i>Jeg tenkte at jeg skulle bli fint i en hvit uniform.</i> (3) <i>i hjemlandet mitt må syke pleier ta på seg hvit uniform.</i> (4) <i>Men faren min ville at jeg skulle studere økonomi</i> (s0853).
Koordinasjon i heilsetningar	I registreringa er det skilt mellom koordinerte heilsetningar med subjekt (<i>De har arbeidet mye og nå fortjener de en spesiell belønning, h7058</i>) og koordinasjon utan subjekt (<i>Vi drakk kaff og satt sammen, s0582</i>).
Leddsetningar	(1) Nominale, td. <i>Men noen mener at morsmålsundervisningen må ut av skolen, og at ressursene bør brukes til norskopplæring</i> (h5251) (2) Relative, td. <i>Det er mange verdier på verden som er viktig for alle mennesker for peis og harmoni på global samfunn</i> (s0336) (3) Adverbiale, td. <i>Jeg vet at det verst er å ikke kjenne noen fordi man bli depriment</i> (s0112). Inndeling etter Hagen (2002)

9. I utveljinga av tekstane prøvde eg å avgrensa temavariasjonen, men enda likevel opp med 32 ulike tematikkar.

10. Det var Hunt som først introduserte denne analyseeininga, som den minste syntaktiske eininga, og som han definerte slik: «one main clause with all subordinate clauses attached to it» (1965: 20).

Maksimal substantivfrase	Variablane er baserte på Berggreen og Sørland (2016: 44) sine avgrensingar ¹² : <ul style="list-style-type: none"> - Finitte leddsetningar er ikkje talte med, noko som gir 10 og ikkje 14 løpeord i <i>tiltak for å forhindre mennesker fra å forlate de yrkene som Norge trenger desperat</i> (h0614). - Samansatte kjerner tel som eitt løpeord: td. er det fire løpeord i <i>natur og kultur til andre land</i> (s0103). - Alle ord ved koordinasjon i adledd medrekna: td. 11 løpeord i <i>Ett av de mest kompliserte og diskuterte temaene her i landet</i> (h9903) - Ord i etterstilte lokalisering eller temporale adledd medrekna: td. <i>her i landet</i> i setninga over.
Subjunksjon	Type subjunksjon registrert og tilfelle av strokne subjunksjonar, td. <i>at</i> og <i>som</i> i <i>Det er ofte du hører på bussen for eksempel et slikt samtal</i> (h0223) og <i>Det ønsker jeg nå er det å forstå, å kommunisere med nordmenn</i> (s0853)
Syntaktiske avvik	Tre grove kategoriar syntaktiske avvik: (1) Eitt ledd er plassert feil i heilsetninga, td. <i>Ungdommer sjeldn har tid til det</i> (s0352), <i>Det er noen situasjoner du kan ikke gå gjennom alene</i> (s0127). (2) Eitt obligatorisk ledd (inkl. finitt verb) manglar i heilsetninga, td. <i>Dette er veldig viktig når_ er mørkt ute</i> (s0107), <i>Så jeg flytte til Norge</i> (s0853). (3) Det er eitt ledd (inkl. finitt verb) for mykje i heilsetninga, td. <i>Når begynte vi å lure på</i> (h0200), <i>Også bare å bo og leve koster det mer og mer</i> (h0295). I tillegg er val av ukorrekt subjunksjon registrert, td. <i>I dag er mobiltelefon en ting hva nesten alle mennesker har med seg</i> (h0291).

Basert på dette analyse- og systematiseringsarbeidet har eg utarbeidd ti variablar, der åtte av dei er på skalanivå (rationivå), det vil seia at verdiane uttrykker eit forhold mellom to tal, der det eine er dividert med det andre. Tabell 3 gir ei oversikt over desse variablane, som er dei uavhengige variablane i studien (rammeverksnivå er avhengig variabel).

Tabell 3. Uavhengige variablar i analysen.

Namn på variabel	Informasjon i variabel	Utrekningsmåte for variabel
Løpeord pr. heilsetn.	Heilsetningslengd	Tal på løpeord dividert med tal på heilsetningar
Koordinasjon pr. heilsetn.	Grad av sidestilt koordinasjon	Tal på koordinerte heilsetningar dividert med tal på heilsetningar totalt
Eliderte koord. pr. heilsetn.	Grad av sidestilt koordinasjon med subjektsutlating	Tal på eliderte koordinerte heilsetningar dividert med tal på heilsetningar totalt
Leddsetn. pr. heilsetn.	Grad av underordning	Tal på leddsetningar dividert med tal på heilsetningar
Korrekte setn. pr. heilsetn.	Grad av meistring av syntaksen i målspråket	Tal på heilsetningar utan syntaktiske avvik dividert med tal på heilsetningar totalt
Maks tal på løpeord i NP	Grad av fraseutbygging	Tal på løpeord i den lengste substantivfrasen i kvar enkelt tekst
Nom. ledds. pr. leddsetn.	Mengd leddsetningar med nominal funksjon i heilsetninga	Tal på nominale leddsetningar dividert med tal på leddsetningar totalt
Rel. ledds. pr. leddsetn.	Mengd leddsetningar med relativ funksjon i heilsetninga	Tal på relative leddsetningar dividert med tal på leddsetningar totalt
Adv. ledds. pr. leddsetn.	Mengd leddsetningar med adverbial funksjon i heilsetninga	Tal på adverbiale leddsetningar dividert med tal på leddsetningar totalt
Type subj. pr. leddsetn.	Variasjon i bruk av subjunksjonar	Tal på ulike typar subjunksjonar

11. All informasjon om avgrensingar kjem ikkje fram i Berggreen og Sørland (2016). Eg har difor hatt kontakt på e-post for å avklara korleis dei handterte ulike tilfelle slik at eg kan samanlikna mine funn med deira.

3.2 Statistiske metodar

Eg har nytta to ulike typar statistisk analyse, og har brukt programvara Stata SE 16. For å kunne samanlikna mine funn med dei andre norske studiane, har eg nytta same metodikk som dei: signifikanstesting av gruppeskilnader der kompleksitetsvariablane, dei uavhengige variablane, er analyserte kvar for seg. Fordi dataa med unntak av ein variabel (rel. ledds. pr. leddsetn.) ikkje er normalfordelte, (*Shapiro Wilk W-test*), brukar eg ikkje-parametriske metodar. Eg nyttar *Kruskal–Wallis* (rangeringstest) først for å finna ut om det finst signifikante skilnader i materialet, og deretter brukar eg *Dunn-test* som post-hoc test for å finna ut kva for grupper som skil seg signifikant frå kvarandre. Eg har nytta regresjonsanalyse for å undersøkje samanhengene mellom rammeverksnivå og kompleksitetsvariablane. Oftast skjer modelleringa basert på lineær regresjon, der ein antar at forholdet mellom y og x , avhengig og uavhengig variabel, kan beskrivast som ei rett linje i eit koordinatorsystem, og der x vil auka og minka proporsjonalt med y . Denne typen regresjonsanalyse føreset at den avhengige variabelen er kontinuerleg. Den avhengige variabelen i denne studien, rammeverksnivå, er nominal. Han er fleirdelt og ordinal fordi det finst fire moglege utfall, A2–C1, som er rangerte i forhold til kvarandre. Ved ein logistisk regresjon kjenner ein ikkje dei underliggende verdiane på y , men ein antar at dei underliggende verdiane på den avhengige variabelen liknar ein kontinuerleg variabel, men i analysen blir det tatt omsyn til at avstanden mellom kategoriane på den avhengige variabelen kan vera ulik (Liao 1994). Ein logistisk regresjon gir oss sannsynet for at ei hending inntreffer eller ikkje inntreffer; i dette tilfellet at ein tekst får ferdigheitsnivå a (td. A2), og ikkje ferdigheitsnivå b, c eller d. Ein føresetnad for å kunne gjera logistisk regresjon er at antakinga om proporsjonale odds (også kalla føresetnaden om parallelle linjer). I dette ligg det at ein føreset at effekten av den uavhengige variabelen er den same mellom dei ulike kategoriane på den avhengige variabelen (Liao 1994). Eg nyttar *ordinal* logistisk regresjon fordi det er meir enn to moglege utfall på den avhengig variabelen (fire rammeverksnivå).

Det er krevjande å fortolka og presentera resultatata frå logistisk regresjon (Tuft 2000), særleg storleiken på samanhengane i modellen. Ein kan gå ut frå odds ratio, som viser kor mange gongar oddsen blir endra når verdien på den uavhengige variabelen stig med ei eining (Tuft 2000). Ei anna moglegheit som er lettare å forstå, og som i mange tilfelle vil vera meir nyttig, er å illustrera samanhengane i modellen basert på dei faktiske tala. Dette kan gjerast ved å rekna ut sannsynet for gitte kombinasjonar av verdiar på dei uavhengige variablane basert på den estimerte logiten (log odds) som regresjonsanalysen gir

(Tuft 2000, UCLA Statistical Consulting Group 2020). Ein kan altså rekna ut sannsynlegheita for plassering av tekstar på nivå basert på ulike verdiar på dei uavhengige variablane, og slik få sannsynet for rammeverks plassering innanfor ulike scenario. På den måten vil storleiken på samanhengane i modellen vera tydelegare, og effektane av dei uavhengige variablane på den avhengige variabelen enklare å sjå.

I litteraturen er R^2 og/eller justert R^2 rekna som dei fundamentale måla for prediksjonsgraden til ein regresjonsmodell (Plonsky & Ghanbar 2018)¹². Dette statistikk målet er eit uttrykk for forklaringskrafta til modellen, og gir eit mål på kor stor prosentdel av variansen i avhengig variabel som blir forklart av dei uavhengige variablane. I denne studien vil eg ikkje leggja stor vekt på dette målet. Det er fordi eg ikkje primært er ute etter å finna ut kva for trekk sensorar legg vekt på når dei vurderer tekstar etter Rammeverket, og difor ikkje nærmar meg datasettet for å identifisera den beste modellen som kan predikera rammeverksnivå; i så fall ville eg ha inkludert andre variablar i studien og ikkje berre syntaktiske (til dømes leksikalske, morfologiske og meir tekstlege variablar). Føremålet i denne studien er å avdekkja samanhengen mellom utvalde mål på syntaktisk kompleksitet og korrektheit og ferdigheitsnivå.

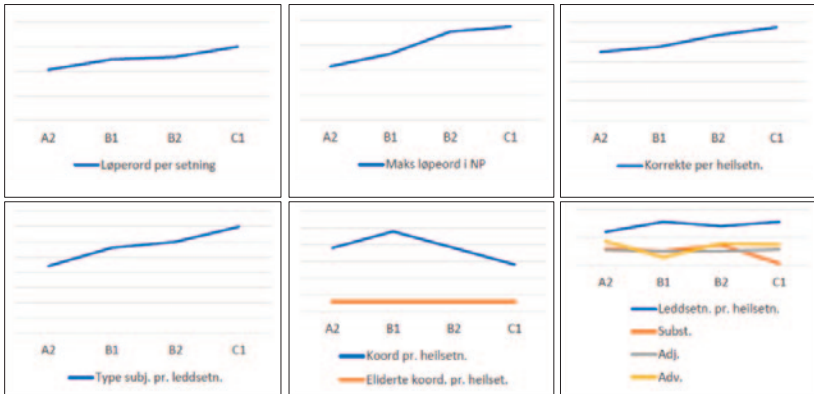
4 Resultat

4.1 Analyse av dei enkelte kompleksitetsmåla på tvers av rammeverksnivå

Trendane i datasettet er visuelt framstilte i enkle linjediagram i figur 1 basert på gjennomsnitta (m i tabell 4). Den grafiske framstillinga kan gi oss eit første inntrykk av korleis kompleksitetsmåla utviklar seg på tvers av rammeverksnivå. Ein kan sjå at verdiane på fire av variablane stig jamt med aukande rammeverksnivå (løpeord pr. heilsetning., maks tal på løpeord i NP, korrekte setn. pr. heilsetning, type subj. pr. leddsetning.), at variabelen som måler koordinasjon per heilsetning viser stigning frå A2 til B1 og deretter eit fall, at fleire har ein kurve som går opp og ned mellom nivå og at somme ikkje viser endring mellom nivå.

12. Forskjellen på R^2 og justert R^2 er at sistnemnde tek høgde for storleiken på utvalet og talet på forklaringsvariablar, og er difor også eit meir konservativt mål (Plonsky & Ghanbar 2018: 715).

Figur 1. Grafisk framstilling over gjennomsnittlige verdier på kompleksitetsmåla per nivå.



Tabell 4 gir sentral- og variasjonsmål for alle dei ti variablane. Variablar der verdier mellom grupper er signifikant forskjellige, er utheva i feit skrift. Dei andre kompleksitetsmåla er ikkje signifikant ulike mellom nokon av nivåa.

Tabell 4. Verdier for mål av syntaktisk kompleksitet og korrektheit: gjennomsnitt (*M*), standardavvik (*SD*) og 95%-konfidensintervall (*KI*).

	A2			B1		B2		C1	
	<i>M</i> (<i>SD</i>)	95% <i>KI</i>	<i>M</i> (<i>SD</i>)	95% <i>KI</i>	<i>M</i> (<i>SD</i>)	95% <i>KI</i>	<i>M</i> (<i>SD</i>)	95% <i>KI</i>	
Løpeord pr. heilsetn.	10.3 (2.1)	9.4, 11.1	12.4 (3.5)	11.1, 13.6	12.9 (1.8)	12.2, 13.6	15.0 (3.4)	13.7, 16.3	
Koordinasjon pr. heilsetn.	.19 (.13)	.15, .24	.24 (.11)	.20, .27	.19 (.08)	.16, .22	.14 (.06)	.12, .16	
Eliderte koord. pr. heilsetn.	.03 (.04)	.01, .04	.03 (.09)	.00, .07	.03 (.03)	.02, .05	.03 (.03)	.01, .04	
Leddsetn. pr. heilsetn.	.60 (.15)	.51, .70	.78 (.43)	.61, .93	.70 (.25)	.60, .79	.78 (.41)	.63, .93	
Korrekte setn. pr. heilsetning	.70 (.15)	.65, .76	.75 (.15)	.69, .80	.87 (.08)	.84, .90	.95 (.05)	.93, .97	
Maks tal på løpeord i NP	4.3 (1.9)	3.5, 5.0	5.3 (1.7)	4.7, 6.0	7.1 (2.2)	6.2, 7.9	7.5 (2.0)	6.8, 8.2	
Nom. ledds. pr. leddsetn.	.30 (.12)	.26, .35	.25 (.12)	.21, .30	.36 (.12)	.31, .41	.28 (.11)	.24, .32	
Rel. ledds. pr. leddsetn.	.27 (.13)	.22, .32	.25 (.10)	.21, .28	.25 (.13)	.22, .31	.28 (.10)	.21, .29	
Adv. ledds. pr. leddsetn.	.43 (.13)	.39, .48	.50 (.14)	.45, .56	.39 (.18)	.31, .44	.47 (.18)	.40, .54	
Type subjunk. pr. leddsetn.	4.4 (1.6)	3.8, 5.1	5.6 (1.6)	5.1, 6.2	6.0 (2.2)	5.5, 7.1	7.0 (1.8)	6.0, 7.4	

Den statistiske analysen viser at A2 skil seg signifikant frå både B1 ($p = 0,0500$), B2 ($p = 0,0001$) og C1 ($p = 0,0000$) i *setningslengde* målt ved løpeord per heilsetning, og at skilnaden mellom B1 og C1 også er signifikant ($p =$

0,0003). Signifikanstestinga av *koordinasjon* i heilsetningar viser at C1 skil seg signifikant frå B1 ($p = 0,0004$), og så vidt frå B2 ($p = 0,0536$), men ikkje frå A2 ($p = 0,1393$). Analysen av *korrekte heilsetningar per heilsetning* viser at A2 er signifikant ulik B2 ($p = 0,0002$) og C1 ($p = 0,0000$), at B1 er signifikant ulik B2 ($p = 0,0088$) og C1 ($p = 0,0000$) og at B2 og C1 er signifikant ulike ($p = 0,0045$). Dei to lågaste nivåa (A2 og B1) er altså ikkje signifikant ulike i grad av syntaktisk korrektheit. Analysen av *utbyggingsgrad av substantivfrasar målt ved maksimalt tal på løpeord* viser at dei to lågaste nivåa (A2 og B1) og dei to høgste nivåa (B2 og C1) ikkje er signifikant ulike. A2 skil seg derimot signifikant frå B2 ($p = 0,0000$) og C1 ($p = 0,0000$) og B1 skil seg også signifikant frå B2 ($p = 0,0162$) og C1 ($p = 0,0006$). Det er berre B2-gruppa som skil seg ut med omsyn til syntaktisk funksjon hos leddsetningane: B2 har ein signifikant *mindre del* leddsetningar med adverbial funksjon samanlikna med B1 ($p = 0,0382$). Den statistiske analysen viser at A2-tekstane har signifikant færre ulike typar subjunksjonar per leddsetning enn informantar på B2 ($p = 0,0012$) og C1 ($p = 0,0000$), og er så vidt signifikant mot B1 ($p = 0,0555$).

4.2 Regresjonsanalyse: Samanhengen mellom rammeverksnivå og ti forklaringsvariablar

Regresjonsanalysen viser at ein modell med berre fire av dei ti uavhengige variablane fyller føresetnadane for logistisk regresjon. Denne modellen finst i vedlegg 5 og har vore testa for linearitet (Brant test, $p = 0,526$), multikollinearitet ($VIF = 2,20$) og spesifikasjonsfeil (linktest, $p = 0,579$). Det er *heilsetningslengde* (løpeord pr. heilsetn.), *underordningsgrad* (leddsetn. pr. heilsetn.), *substantivfraseutbygging* (maks tal på løpeord i NP) og *type subjunksjon pr. leddsetn.* som har samanheng med rammeverksnivå. Alle variablane utan (variabelen som måler) *underordningsgrad* korrelerer positivt med rammeverksnivå. Denne modellen er resultat av ei utsiling av variablar som bryt med føresetnadane for proporsjonale odds. I slike tilfelle tilrår litteraturen at ein i staden bør bruka ein annan type analyse for logistisk regresjon, *generalized ordinal logistic regression* (UCLA Statistical Consulting Group 2020). Resultatet av den analysen er vist i vedlegg 6, og modellen har vore testa for multikollinearitet ($VIF = 2,10$) og spesifikasjonsfeil (linktest, $p = 0,045$ (A2), 0,028 (B1), 0,318 (B2)). Denne modellen viser at sju av dei ti variablane bidreg signifikant til å forklara plassering av tekstane på rammeverksnivå. Også i denne analysen har underordning negativ samanheng med rammeverksnivå, og det har også variabelen som måler koordinasjon. Dei to variablane som ikkje slår ut signifikant, er *rel. leddsetn. pr. leddsetn.* og *adv. leddsetn. pr. leddsetn.* Ein

kan også sjå at forklaringsverdien til modellen med sju variablar er høgare (56,21 %, vedlegg 6) enn modellen som er resultatet av den første regresjonsanalysen (33,50 %, vedlegg 5). For å få eit inntrykk av samanhengen mellom variablane i modellen og storleiken på effekten av dei uavhengige variablane på den avhengige (rammeverksnivå), har eg estimert nivåfordelinga basert på regresjonen vist i vedlegg 6. Estimeringa av den enkelte variabelen byggjer på verdiane på første, andre og tredje kvartil, og ved at verdien på dei andre variablane er sett på sitt gjennomsnitt. Ved å velja ut tre verdiar får eg ei estimert sannsynsfordeling basert på tre ulike scenario. Av tabell 5 ser me for eksempel at det er mest sannsynleg (70 %) at tekstar med 14,3 ord per heilsetning har B2-nivå, og at det ikkje er særleg sannsynleg at tekstar med same gjennomsnittlege heilsetningslengde er på A2-nivå, (3 %):

Tabell 5. Estimert sannsynleg fordeling på rammeverksnivå basert på tre scenario med ulike verdiar på uavhengig variabel.

Variabel	Scenario	Verdi	Estimat			
			A2	B1	B2	C1
Løpeord pr. heilsetn.	1	10,5	0,23	0,71	0,05	0,00
	2	12,1	0,09	0,57	0,33	0,02
	3	14,3	0,03	0,11	0,70	0,16
Leddsetn. pr. heilsetning	1	0,50	0,07	0,32	0,56	0,05
	2	0,65	0,08	0,48	0,41	0,02
	3	0,86	0,09	0,68	0,22	0,01
Koordinasjon pr. heilsetn.	1	12,50	0,11	0,45	0,41	0,03
	2	17,95	0,09	0,54	0,36	0,02
	3	24,72	0,07	0,63	0,29	0,01
Maks tal på løpeord i NP	1	4	0,19	0,58	0,21	0,02
	2	5	0,07	0,53	0,38	0,02
	3	7	0,05	0,49	0,45	0,01
Korrekte setn. pr. heilsetning	1	0,86	0,14	0,76	0,09	0,00
	2	0,92	0,06	0,34	0,55	0,05
	3	0,97	0,03	0,06	0,65	0,25
Nom. ledds. pr. leddsetn.	1	0,25	0,10	0,62	0,26	0,02
	2	0,29	0,09	0,57	0,32	0,02
	3	0,37	0,07	0,43	0,48	0,01
Type subjunk. pr. leddsetn.	1	33,3	0,13	0,44	0,41	0,02
	2	50,0	0,07	0,60	0,32	0,02
	3	56,4	0,05	0,64	0,28	0,02

Dersom det var slik at variablane skilde perfekt mellom nivå, ville kvar variabel i tabell 5 hatt låge verdiar oppe til venstre og høgare verdiar nede til høgre i tilfelle der prediksjonen var ein positiv korrelasjon, og motsett mønster i tilfelle ein predikerte ein negativ korrelasjon mellom variabelen og rammeverksnivå. Generelt er observasjonen derimot at variablane i størst grad fangar inn skilna-

der på tekstar på B1 og B2-nivå, og skil i liten grad mellom dei to øvste nivåa. For *løpeord pr. heilsetn.* ser me at det skjer eit tydeleg skifte mellom 12,1 og 14,3 i favor av B2-nivået. Sannsynet for B2 aukar også med tal på løpeord i NP, men her ligg tyngdepunktet på B1. Ein generell tendens er også at sannsynet for A2-nivå fell i takt med at verdien på dei uavhengige variablane stig med unntak av ein variabel: *leddsetn. pr. heilsetn.* Fleire leddsetningar reduserer først og fremst sjansen for B2, og tilsvarande aukar sannsynet for B1. Ein underordningsgrad på 0,5, dvs. at annakvar heilsetning i teksten inkluderer ei leddsetning, ser ut til å føra til størst sannsyn for nivå B2. Variabelen *koordinasjon pr. heilsetning* følgjer om lag same mønster som underordningsvariabelen. Variabelen som måler korrektheit er den som i størst grad er i stand til å skilje mellom nivå. Den fordeler seg slik ein burde, dersom antakinga er at ein skårar høgare på rammeverksnivå kvar gong den uavhengige variabelen aukar med ei eining. Bruken av nominale leddsetningar ser ut til å vera assosiert med B1-nivå opp til eit visst nivå. Variabelen *type subjunk. pr. leddsetn.* ser ut til å fanga opp lite. Tyngdepunktet ligg på B1 og B2, men den prosentvise skilnaden er liten, men fallande mot B1-nivå.

6 Diskusjon

6.1 Samanlikning av funn på tvers av studiar

Eg samanliknar først funna som svarar på forskings spørsmål 1. Som i dei tidlegare norske studiane er *setningslengde* (målt ved tal på løpeord per heilsetning) og *fraseutbygging* (målt ved talet på løpeord i maksimal NP-frase) mål som skil mellom tekstar skrivne av andrespråksinnlærarar som er på ulike ferdighetsnivå. Sannsynet for nivå A2 fell ein god del med aukande heilsetningslengd, men endringar på denne variabelen har størst betydning for sannsynet for at ein tekst er på B1 eller B2-nivå, og same trend er gjeldande for estimeringa av nivåfordeling etter variabelen *substantivfraseutbygging*. I motsetnad til dei norske studiane finn eg ikkje at grad av *underordning* aukar signifikant mellom nivåa. Det er ei stor auke i gjennomsnittleg underordningsgrad frå A2 til B1 også i mine data (frå 0,60 til 0,78, jf. tabell 4), men analysen av gjennomsnittsverdiar viser ingen signifikante skilnader mellom gruppene. Regresjonsanalysen viser eit anna resultat: Det er signifikant negativ samanheng mellom underordning og rammeverksnivå. Sannsynet for B2 fell med stigande underordningsgrad, medan sannsynet for dei to lågaste nivåa stig med aukande underordning. Scenarioet med høgaste underordningsgrad gir 68 % sannsyn for at ein tekst er

på B1-nivå. Dette samsvarar med funn i fleire internasjonale studiar som finn at underordningsgraden aukar fram til mellomnivå og avtar mot høgare nivå¹³ (jf. td. Kuiken & Vedder 2019). Det kan bety at innlærarar som har nådd eit mellomnivå, har nådd «toppen» med omsyn til underordningsgrad. Sjølv om eg har målt *korrektheitsdimensjonen* ved å sjå på mengd korrekte heilsetningar, og ikkje mengd ukorrekte slik Moe (2013) og Sørensen (2017) gjer, finn eg om lag same resultat: Korrektheit er eit mål som skil godt mellom tekstar på ulike rammeverksnivå, og i motsetnad til kompleksitetsmåla ein variabel som også skil mellom dei høgaste nivåa. Dette samsvarar med Eriksson og Carlsen (2013) som observerer at talet på syntaktiske feil på gruppenivå fell frå A2 til C1 (basert på fordelinga av grammatiske feilkodar i ASK, jf. omtalen av studien i del 2). Forskingsspørsmål 1 inkluderer også *type subjunksjon*. Analysane viser at variasjon i bruken aukar, men berre frå A2 til B1, eit funn som skil seg frå Carlsen (2013). Det kan ha samanheng med at Carlsen sin studie av bindeord omfattar ord av fleire typar enn berre subjunksjonar.

Graden av sidestilt koordinasjon fell signifikant frå B1 og B2 til C1 (forskingsspørsmål 2), som betyr at tekstar på C1-nivå og tekstar på A2-nivå er like i grad av koordinasjon. Dette samsvarar med Sørensen (2017) som også finn at tekstane i materialet på det høgste nivået, B2, har signifikant lågare grad av underordning samanlikna med B1-tekstar, men ikkje samanlikna med tekstar på det lågaste nivået, A2. På dette punktet skil altså Berggreen og Sørland (2016) seg ut ved å 1) finna at tekstar på lågaste og høgaste utviklingstrinn skil seg signifikant frå kvarandre, og 2) at skilnaden inneber ein auke i koordinasjonsgraden frå første til tredje målepunkt. Medan Sørensen og eg altså finn ein kurve for koordinasjonsgrad som liknar ein u opp ned, som også er dokumentert i den internasjonale litteraturen («omega-shaped», Kuiken & Vedder 2019: 195), der koordinasjonsgraden går opp for så å gå ned att, finn Berggreen og Sørland altså ei jamn stigning. Berggreen og Sørland forklarar den høge graden av koordinasjon med alderen på elevane, og at elevane skriv om lag slik dei snakkar: «Det virker som en rimelig tanke at dette trekket ved tekstene avspegler ‘on-line production circumstances’» (2016: 59). Det er mogleg at dette har samanheng også med sjanger, då dei fleste tekstane på høgare nivå i mitt og Sørensen sitt materiale ikkje er forteljingar, men meir argumenterande tekstar. I slike tekstar er additive forbindingar mindre vanleg. Det er også ein skilnad i kontekst mellom studiane. Tekstinnsamlinga i Berggreen og Sørland sin studie

13. I dei internasjonale studiane er ikkje ferdigheitsnivå målt eller kontrollert alltid med det same verktøyet. Det vil seia at CEFR ikkje er nytta i alle studiane.

vart gjennomført som del av den ordinære klasseromsundervisninga og administrert av læraren. Tekstane i Sørensen og min studie er skrivne som del av eit prøvesvar, i ein formell vurderingssituasjon som er av stor betydning for kandidaten.

Den grupperbaserte statistiske analysen og regresjonsanalysen av syntaktisk funksjon (forskingsspørsmål 3) er noko vanskeleg å tolka, men hovudkonklusjonen er at variabelen ikkje slår ut som betydingsfull for rammeverksnivå i denne studien.

6.2 Samanhengen mellom syntaktisk kompleksitet og ferdigheitsnivå

I samanfatningar av forskinga på CAF-triaden blir det konkludert med at det er korrelasjon av betydning mellom syntaktisk kompleksitet og ferdigheitsnivå (td. Kuiken & Vedder 2019, Norris & Ortega 2009, Wolfe-Quintero et al. 1998), som også kan observerast for fleire i denne studien, men ikkje for alle av variablane. Det er grunnlag for å konkludera med at det er signifikant samheng mellom fire av måla på syntaktisk kompleksitet (setningslengde, koordinasjon, underordning, fraseutbygging) og ferdigheitsnivået tekstane er plasserte på etter Rammeverket. I tillegg er det samheng mellom syntaktisk korrektheit og ferdigheitsnivå, det vil seia, ein av feilkategoriane, *feil val av subjunksjon*, handlar først og fremst om eit avvikande *leksikalsk* val. Denne feilen utgjør berre ein liten av del av dei andre avvikstypene (37 av 669 avvik totalt, dvs. 6 %), som er syntaktiske brot av tre hovudtypar (jf. Metode).

Sjølv om studien gir støtte til korrelasjonen mellom syntaktisk kompleksitet og ferdigheitsnivå, er det eit viktig funn at denne samanhengen ikkje er gjennomgåande. Progresjonen i kompleksifiseringsprosessen ser ut til å avta mot B2 då regresjonsmodellen først og fremst estimerer sannsynet for B1 eller B2-nivå. Variablane i denne studien ser altså i liten grad ut til å fanga opp skilnader mellom dei øvste nivåa, B2 og C1. Paquot (2019) konkluderer om lag på same måte basert på ein studie av 98 tekstar frå VESPA-korpuset¹⁴ skrivne av franske L2-innlærarar av engelsk. Ho undersøkte korleis tradisjonelle mål på syntaktisk (inkl. setningslengde, koordinasjon, fraseutbygging, underordning) og leksikalsk kompleksitet er i stand til å beskriva dei høgare nivåa i Rammeverket (B2, C1 og C2), til samanlikning med mål som fangar inn kompleksitet i fraseologiske einingar (ordkombinasjonar). Paquot finn at måla på «phraseological complexity», som ho introduserer som ein ny type kompleksitet, *i motsetnad* til dei syntaktiske og leksikalske kompleksifiseringsmåla skil mellom innlæ-

14. VESPA= Varieties of English for Specific Purposes database

rarperformanse på dei tre høgaste nivåa. Ho meiner dette funnet kan tyda på at «essential aspects of language development from upper-intermediate to very advanced proficiency level may be situated in the phraseological dimension» (2019: 121). I denne studien finst det eitt unntak frå regelen om manglande samanheng mellom syntaktisk kompleksitet og nivå. Dette gjeld målet for korrektheit som skil godt mellom B2 og C1: Sannsynet for C1 aukar frå 5 % til 25 % når mengda korrekte heilsetningar aukar frå 92 % til 97 %. Analysen i denne studien seier derimot ingenting om *type* syntaktiske feil, men vedlegg 3 viser at dei tre syntaktiske avvikstypane er ganske likt (relativt) fordelt mellom nivåa. Det kan tyda på at skilnaden i korrektheit først og fremst handlar om *mengd* (sjå vedlegg 4 for innsyn i tal på førekomstar totalt fordelt på alle typar syntaktiske feil), og ikkje *type*; til og med C1-tekstar i dette materialet har avvik som i litteraturen er rekna som elementære, til dømes utelating av verbal og subjekt. Ein studie av Thewissen (2013) peikar i retning av at *type* feil ikkje skil mellom skriftleg performanse på ulike ferdigheitsnivå. Ho studerer korleis 46 feiltypar¹⁵ utviklar seg mellom rammeverksnivåa A2–C2 i 223 tekstar frå eit engelsk innlærarkorpus (ICLE¹⁶). Thewissen finn at berre to av feiltypane viser eit gradvis signifikant fall med stigande rammeverksnivå, og at 16 av feiltypane, inkludert forvekslinga mellom finitte og infinitte verbformer, ikkje viste progresjon frå B1 til C2. Basert på studie av *feiltypar*, konkluderer altså Thewissen på ein annan måte om samanhengen mellom korrektheit og nivå, enn det eg kan gjera i denne studien som berre ser på *feilmengd*.

Til slutt, i ein diskusjon om samanhengen mellom kompleksifiseringsmål og ferdigheitsnivå på andrespråket, kan ein naturlegvis ikkje sjå vekk frå at ein kunne ha fått eit anna resultat dersom ferdigheitsnivået hadde vore fastsett basert på ein annan ferdigheitsskala og andre vurderingskriterium.

6.3 Utvikling av syntaktisk kompleksitet

Som nemnd i bakgrunnsdelen, skisserer metastudier av forskinga på syntaktisk kompleksitet (Norris & Ortega 2009, Wolfe-Quintero et al. 1998) ei utvikling der den syntaktiske kompleksiteten i innlærarspråket utviklar seg frå koordinasjon til underordning og til utbygging av frasekompleksitet. I denne studien er det vanskeleg å gi støtte til ideen om at underordning er noko som tek over for koordinasjon, som ein måte å gjera språket meir komplekst på. Årsaka er at

15. Thewissen såg på mange ulike typar feil, grammatiske, leksikalske, ortografiske og feil knytte til tekstkohesjon.

16. ICLE: The International Corpus of Learner English.

både koordinasjonsgraden og underordningsgraden viser litt parallelle utviklingar på tvers av rammeverksnivå ved at dei først aukar frå A2 til B1, og deretter søkk og/eller stabiliserer seg. I denne studien er det berre eit relativt enkelt mål på frasekompleksitet. Sjølv om grunnlaget for å konkludera om kompleksifisering på frasenivå utviklar seg etter koordinasjon og underordning difor er noko svakt, er det i det minste slik at variabelen som måler fraseutbygging i substantivfrasar, 1) korrelerer positivt med rammeverksnivå og 2) i større grad enn måla på koordinasjon og underordning estimerer B2-nivået. Denne studien viser også at enkelte trekk ved andrespråket ikkje utviklar seg lineært, som til dømes koordinasjonsgraden som først stig og så fell (jf. Kuiken & Vedder 2019).

Fleire av studiane i CAF-tradisjonen er longitudinelle, og det er ikkje heilt uproblematisk å samanlikna funn frå slike studiar med tverrsnittsnitt-studiar som denne, eller Moe (2013) og Sørensen (2017). Det er ikkje uvanleg å studera språkutvikling basert på kvasi-longitudinelle studiar, og metoden er utbreidd innanfor den korpusbaserte andrespråksforskinga, som til dømes Thewissen (2013) som har som føremål å studera «second language accuracy developmental trajectories» (2013: 77). Paquot (2019) er også ein tverrsnittstudie, som gir grunnlag for å stilla interessante spørsmål om utvikling. Hennar funn kan tyda på at syntaksen i andrespråkutvikling fell på plass før meir eller mindre faste uttrykk og ordkombinasjonar (fraseologiske einingar). Paquot såg ikkje på morfologisk kompleksitet, og eit interessant spørsmål er i kva grad mål på morfologisk kompleksitet har samband med dei øvre rammeverksnivåa; Slabakova (2006) meiner til dømes at morfologien i eit andrespråk blir tileigna etter syntaktiske og semantiske trekk («The Bottleneck hypothesis»). Dersom det stemmer, skulle ein kunna forventa at ein liknande studie som denne, basert på morfologiske og ikkje syntaktiske variablar, i større grad ville ha estimert B2 og C1-nivåa. I følgje Bulté og Housen (2014) er det viktig å sjå korrelasjonen mellom skriftleg performanse og nivåplassering i lys av kva ein veit om korleis andrespråk utviklar seg over tid. Vurdering av andrespråksferdigheiter byggjer på kunnskap om korleis andrespråk utviklar seg, og dersom verktøya skal vera gode og kunna skilja mellom nivå, må dei ta utgangspunkt i trekk som er sensitive for utvikling. I ein studie av skriftlege tekstar frå 45 vaksne L2-innlærarar som deltok på eit intensivt kurs i engelsk, finn dei også at sambanden mellom kompleksitetsmål (syntaktiske og leksikalske) som viser utvikling over tid, og vurdering av ferdigheitsnivå, slett ikkje alltid samsvarar. Ut frå det problematiserer dei subjektive vurderingar basert på holistiske vurderingskriterium.

6.4 Metodiske refleksjonar

I denne studien har det vore eit mål å visa fram ein metodikk som gjer det mogleg å 1) kunna gjera ein type analyse som kan fanga betydinga av fleire trekk samstundes og 2) som viser fram konsekvensen av resultatet av ein kompleks analyse for det eg vil finna ut av: samanhengen syntaktisk kompleksitet og vurdering av skriftleg andrespråkskompetansen på ferdigheitsnivå. Eg meiner studien viser at det er tenleg og nødvendig å bruka multivariat statistikk. For eksempel viser det seg at fleire av måla som ifølgje Moe (2013) har signifikant samanheng med rammeverksnivå, ikkje slår ut når fleire av variablane blir analyserte saman. Me såg også at signifikansanalysane ikkje fanga opp at underordning slår signifikant negativt ut når variabelen inngår i ein modell med fleire andre syntaktiske variablar. I tillegg gir ein regresjonsanalyse i større grad eit bilete av samanhengen fordi vi også får informasjon om kor stor betyding dei enkelte variablane har for plassering på rammeverksnivå. Når det er sagt, er det ein svakheit ved denne studien at han til liks med dei tidlegare norske studiane, og dei aller fleste internasjonale, ikkje undersøker betydinga av eksterne faktorar som teksttype og sjanger, og at han heller ikkje undersøker potensielle påverknad av L1-bakgrunn (jf. Ortega 2015). Dei individuelle profilane er også skjult i ein kvantitativ studie som denne. Ei anna avgrensing ved studien er også at den kviler på globale mål. Så sjølv om analysen viser at dei sju syntaktiske variablane som slår signifikant ut i regresjonsanalysen (vedlegg 6) forklarar 50 % av nivåtildelinga, er det noko avgrensa med innsikt me får om kva denne kompleksifiseringa inneber fordi analysen kviler på globale og ikkje språk-spesifikke mål. For eksempel veit me ikkje kva for element eller type modifisering som gjer at substantivfrasane blir lengre frå A2 til B2 (med unntak av vedlegg 3 og 4 som gir innsyn i type feil per nivå på gruppenivå). Bruken av globale mål i CAF-tradisjonen er også kritisert (td. Baten & Håkansson 2015).

7 Avslutning

Trass avgrensingane ved studien har han dokumentert at det er ein samanheng mellom syntaktisk kompleksitet og korrektheit og rammeverksnivå. Samstundes viser studien også at måla først og fremst beskriv skriftleg andrespråkskompetanse på nedre og øvre mellomnivå. Unntaket frå dette er korrektheit som tilsynelatande spelar ei rolle for om ein tekst er vurdert til å vera på høgt mellomnivå (B2) eller avansert nivå (C1). Det manglande samsvaret mellom mål på syntaktisk kompleksitet og meir avanserte andrespråksferdigheter pei-

kar i retning av at det trengst liknande studiar som kan undersøkje samanhengen mellom andre språklege trekk og ferdigheitsnivå. Dette kan vera fraseologiske studiar, som antyda av Paquot (2019), eller studier som er innretta mot morfologiske problemstillingar.

Litteraturliste

- Baten, Kristof & Gisela Håkansson. 2015. The development of subordinate clauses in German and Swedish as L2s: A theoretical and methodological comparison. *Studies in Second Language Acquisition* 37, 517–547.
- Berggreen, Harald & Kjartan Sørland. 2016. Syntaktisk kompleksitet i et skriftlig innlærermateriale. *NOA. Norsk som andrespråk* 32-1/2, 31–75.
- Bellugi, Ursula & Roger Brown (red.). 1964. The acquisition of language. *Monographs of the Society for Research in Child Development* 29-1, 1–192.
- Bulté, Bram & Alex Housen. 2012. Defining and operationalising L2 complexity. I Housen, Alex, Folkert Kuiken & Ineke Vedder (red.): *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins, 21–46.
- Bulté, Bram & Alex Housen. 2014. Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing* 26, 42–65.
- Bulté, Bram & Alex Housen. 2018. Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International journal of applied linguistics* 28, 147–164.
- Carlsen, Cecilie. 2013. Tekstsammenheng og bindeledd på tvers av rammeverksnivåer. I Carlsen, Cecilie (red.): *Norsk profil. Det felles europeiske rammeverket spesifisert for norsk. Et første steg*. Oslo: Novus, 223–243.
- De Clercq, Bastien de & Alex Housen. 2019. The development of morphological complexity: A cross-linguistic study of L2 French and English. *Second language research* 35-1, 71–97.
- Eide, Kristin Melum. 2009. Finiteness: The haves and the have-nots. I Alexiadou, Artemis, Jorge Hankamer, Thomas McFadden, Justin Nuger & Florian Schäfer (red.): *Advances in comparative Germanic syntax. Linguistik Aktuell/Linguistic Today* 141, 357–390. Philadelphia: John Benjamins.
- Eide, Kristin Melum. 2015. *Norsk andrespråkssyntaks*. Oslo: Novus Forlag.
- Eriksson, Trinelise og Cecilie Carlsen. 2013. Rammeverket, feilmengde og feilmønster. I Carlsen, Cecilie (red.): *Norsk profil. Det felles europeiske rammeverket spesifisert for norsk. Et første steg*. Oslo: Novus, 245–267.

- Golden, Anne & Rita Hvistendahl. 2010. Skriveforskning i et andrespråksperspektiv *NOA. Norsk som andrespråk* 26-2, 36–66.
- Gujord, Ann-Kristin H. 2017. Eit empirisk tilskot til utforskinga av finittheit- og tempusdistinksjonar i norsk som andrespråk. *NOA. Norsk som andrespråk* 33-2, 88–116.
- Hagen, Jon Erik. [2001]2013. Finittkategoriens kritiske karakter i norsk som andrespråk. I Nordanger, Marte, Silje Ragnhildstveit & Aasne Vikøy (red.): *Fra grammatikk til språkpolitikk. Utdrag fra Jon Erik Hagens forfatterskap*. Oslo: Novus forlag, 187–201.
- Hagen, Jon Erik. 2002. *Norsk grammatikk for andrespråklærere*. Oslo: Ad Notam Gyldendal.
- Housen, Alex, Folkert Kuiken & Ineke Vedder (red.). 2012. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.
- Hunt, Kellogg W. 1965. *Grammatical structures written at three grade levels*. Urbana, IL: The National Council of Teachers of English.
- Jensen, Bård Uri. 2018a. Syntaks i norsk innlærerspråk: empiriske funn. I Gujord, Ann-Kristin H. & Gunhild Randen (red.): *Norsk som andrespråk – perspektiver på læring og utvikling*. Oslo: Cappelen Damm Akademisk, 235–260.
- Jensen, Bård Uri. 2018b. Er resultatet gyldig? Noen utfordringer ved bruk av kvantitative metoder i andrespråksforskningen. I Gujord, Ann-Kristin H. & Gunhild Randen (red.): *Norsk som andrespråk – perspektiver på læring og utvikling*. Oslo: Cappelen Damm Akademisk, 449–466.
- Jensen, Bård Uri. 2020. Hypotesetesting versus utforskende statistikk i språkforskning. *NOA. Norsk som andrespråk* 36-1, 39–71.
- Kuiken, Folkert & Ineke Vedder. 2019. Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. *International journal of applied linguistics* 29, 192–210.
- Larsen-Freeman, Diane. 1978. An ESL Index of Development. *TESOL Quarterly* 12-4, 439–448
- Larsen-Freeman, Diane. 2006. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics* 27, 590–619.
- Liao, Tim Futing. 1994. *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. Thousand Oaks, CA: Sage Publications, Inc.

- Mangerud, Turid. 1988. *Analyse av tyrkiske grunnskoleelevers norske mellom-språk*. Hovedoppgave. Universitetet i Oslo.
- Meurer, Paul. 2012. ASK - Norsk andrespråskorpus. Bergen: Uni Computing. Lokalisert 2020 på <http://clarino.uib.no/ask/ask>
- Michel, Marije. 2017. Complexity, accuracy and fluency in L2 production. I Loewen, Shawn & Masatoshi Sato (red.): *The Handbook of Instructed Second Language Acquisition*. New York: Routledge, 50–68.
- Moe, Eli. 2013. Syntaktisk kompleksitet og rammeverksnivå. I Carlsen, Cecilie (red.): *Norsk profil. Det felles europeiske rammeverket spesifisert for norsk. Et første steg*. Oslo: Novus, 137–159.
- Norris, John M. & Lourdes Ortega. 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics* 30-4, 555–578.
- Ortega, Lourdes. 2003. Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing. *Applied Linguistics* 24-4, 492–518.
- Ortega, Lourdes. 2012. Interlanguage complexity: A construct in search of theoretical renewal. I Kortmann, Bernd & Benedikt Szmrecsanyi (red.): *Linguistic complexity: Second language acquisition, indigenization, contact*. Berlin: Mouton de Gruyter, 127–155.
- Ortega, Lourdes. 2015. Syntactic complexity in L2 writing: Progress and expansion. *Journal of second language writing* 29-1-2, 82–94.
- Pallotti, Gabriele. 2015. A simple view of linguistic complexity. *Second Language Research* 31-1, 117–134.
- Plonksy, Luke & Hessameddin Ghanbar. 2018. Multiple Regression in L2 Research: A Methodological Synthesis and Guide to Interpreting R2 Values. *The Modern language journal* 102-4, 713–731.
- Paquot, Magali. 2019. The phraseological dimension in interlanguage complexity research. *Second language research* 35-1, 121–145.
- Slabakova, Roumyana. 2006. Is There a Critical Period for Semantics? *Second Language Research* 22-3, 302–338.
- Skehan, Peter. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Sørensen, Ida Marie M. 2017. *Utvikling av syntaktisk kompleksitet i tekster skrevet av vaksne andrespråksinnlærere*. Masteroppgåve ved Universitetet i Bergen. <http://bora.uib.no/handle/1956/17854>

Thewissen, Jennifer. 2013. Capturing L2 Accuracy Developmental Patterns: Insights From an Error-Tagged EFL Learner Corpus. *The modern language journal* 97-S1, 77–101.

Tufte, Per Arne. 2000. *En intuitiv innføring i logistisk regresjon*. Oslo: Statens institutt for forbruksforskning.

Utdanningsdirektoratet. 2011. *Det felles europeiske rammeverket for språk: læring, undervisning, vurdering*. Oslo: Utdanningsdirektoratet.

Verspoor, Marjolijn, Lowie Wander, Hui Ping Chan & Louisa Vahtrick. 2017. Linguistic complexity in second language development: variability and variation at advanced stages. *Recherches en didactique des langues et des cultures* 14-1.

Vyatkina, Nina, Hagen Hirschmann & Felix Golcher. 2015. Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. *Journal of second language writing* 29-1-2, 28–50.

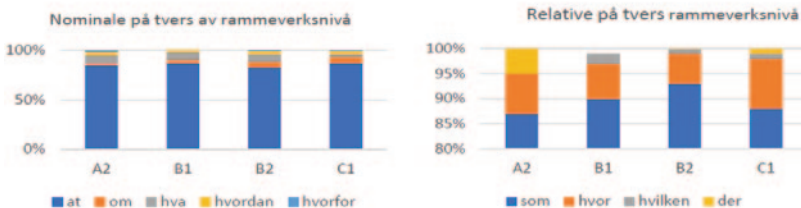
Wolfe-Quintero, Kate, Shunji Inagaki & Hae-Young Kim. 1998. *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawaii.

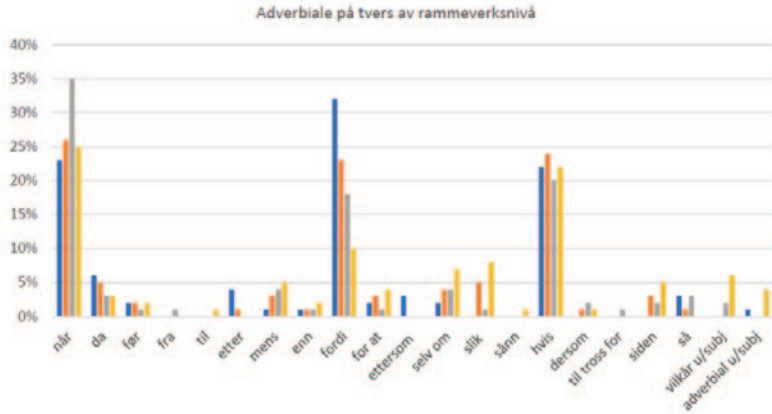
UCLA Statistical Consulting Group (8. august, 2020). Ordered logistics regression. Henta frå <https://stats.idre.ucla.edu/stata/dae/ordered-logistic-regression/>

Vedlegg 1: pid-nummera til dei 123 ASK-tekstane som er med i undersøkinga

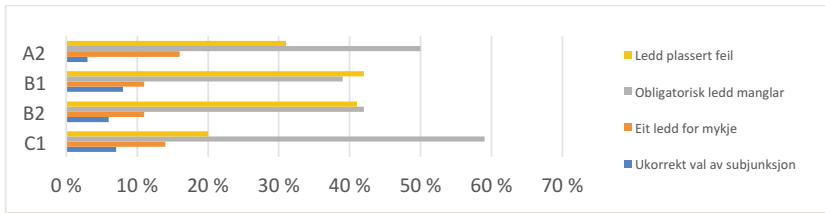
s0002	s0493	s0643	h0283	s0231	s0791	h0028	h0313	h0614	h0009	h0436	h0603
s0079	s0507	s0657	h0393	s0311	s0808	h0031	h0331	h0655	h0156	h0461	h0638
s0097	s0528	s0681	h0651	s0352	s0846	h0075	h0397	h0662	h0157	h0475	h0731
s0103	s0546	s0702	h0690	s0356	s0848	h0136	h0450	h0678	h0165	h0481	h100
s0104	s0552	s0708	h0691	s0542	s0853	h0168	h0466	h0684	h0189	h0528	h218
s0107	s0554	s0710	s0003	s0611	s0867	h0200	h0470	h0706	h0210	h0549	h5251
s0335	s0570	s0714	s0042	s0647	s0897	h0223	h0504	s0168	h0212	h0559	h7011
s0336	s0582	s0729	s0062	s0666	s0994	h0259	h0513	s0522	h0219	h0567	h7032
s0440	s0596	s0873	s0112	s0731	s1003	h0291	h0541	s0633	h0310	h0579	h7058
s0467	s0642	s0987	s0127	s0767	s1004	h0295	h0575	s0935	h0370	h0595	h9903

Vedlegg 2: Subjunksjonar i bruk fordelt på rammeverksnivå og syntaktisk funksjon (relative verdier)





Vedlegg 3: Fordeling av fire hovudtypar avvik etter rammeverksnivå (relative verdjar)



Vedlegg 4: Fordeling av syntaktiske avvikstypar etter rammeverksnivå (absolutte verdjar)

	Ledd plassert feil i heilsetning (HS) og leddsetning (LS)								Obligatorisk ledd manglar i setninga								Overmarkering			sum
	HS oversene	HS underniv.	HS adverbial	HS - anna ledd	LS underniv.	LS adverbial	LS ledd	LS - anna ledd	Verbal	Subjekt	Objekt	Preposisjon	Partikkel verbale	Infinitivsam erke	Subjunksjo n	Anna	Partikkel verbale	Subjekt	Anna	
A2	26	5	3	2	17	18	1	18	27	10	5	46	1	8	1	29	8	0	225	
B1	29	2	18	7	9	26	0	7	10	10	26	22	1	8	1	14	4	5	199	
B2	16	0	16	7	7	21	1	2	5	7	28	15	2	10	1	13	5	0	156	
C1	2	1	6	0	0	2	0	4	11	1	6	4	2	4	1	7	1	0	52	

Vedlegg 5: Regresjon – output frå Stata med fire avhengige variablar som ikkje bryt med føresetnaden for logistisk regresjon:

```
Ordered logistic regression
```

Number of obs	=	120
LR chi2(4)	=	111.47
Prob > chi2	=	0.0000
Pseudo R2	=	0.3350

Log likelihood = -110.61849

	CEFR2	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
TINDEKSløpeordTeiningar_LOG		8847.528	14452.97	5.56	0.000	369.025 217425.9
MAKSFRASEANTALLORD_LOG		8.700125	5.579819	3.37	0.001	2.475187 30.580939
VARSUBJUNKINDEKSantallulike		1.645971	.1756986	4.67	0.000	1.335245 2.029006
LEDDINDEKSleddsTeiningar		.0191736	.0183966	-4.12	0.000	.0029241 .125722
/cut1		24.27449	3.470048			17.47332 31.07566
/cut2		26.63576	3.648194			19.48543 33.78609
/cut3		28.71976	3.793374			21.28488 36.15464

Vedlegg 6: Regresjon – output frå Stata med gologit-kommando:

```
Generalized Ordered Logit Estimates
```

Number of obs	=	120
LR chi2(21)	=	187.00
Prob > chi2	=	0.0000
Pseudo R2	=	0.5621

Log likelihood = -72.85341

	CEFR2	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
A2						
TINDEKSløpeordTeiningar_LOG		2528.666	7285.719	2.72	0.007	8.920047 716829.5
LEDDINDEKSleddsTeiningar		.4410831	.6768179	-0.53	0.594	.0217971 8.925707
MAKSFRASEANTALLORD_LOG		15.16659	15.7345	2.62	0.009	1.985241 115.8677
KOORINDEKKoordTeiningar2		1.052109	.0313232	1.71	0.088	.9924738 1.115328
KORRINDEKSTotaltkorrekteT_SQ		1.608005	.6683782	3.44	0.001	.465.8109 5.55e+09
VARLEDDINDEKSSUBSTTypesubs		12.26459	41.86491	0.73	0.463	.0152439 9867.573
VARLEDDINDEKSADVERBIALType2		1.044916	.0326755	1.41	0.160	.9827961 1.110962
_cons		3.64e-17	3.40e-16	-4.05	0.000	3.99e-25 1.31e-09
B1						
TINDEKSløpeordTeiningar_LOG		3484770	1.49e+07	3.52	0.000	793.6627 1.53e+10
LEDDINDEKSleddsTeiningar		.0098705	.0205425	-2.22	0.026	.000167 .5832572
MAKSFRASEANTALLORD_LOG		6.430238	9.953148	1.20	0.229	.3095145 133.5897
KOORINDEKKoordTeiningar2		.9517107	.0564773	-0.83	0.404	.847212 1.069099
KORRINDEKSTotaltkorrekteT_SQ		1.05e+18	1.13e+19	3.84	0.000	6.76e+08 1.62e+27
VARLEDDINDEKSSUBSTTypesubs		2119.078	11205.03	1.45	0.148	.0668736 6.71e+07
VARLEDDINDEKSADVERBIALType2		.9772522	.0360856	-0.62	0.533	.9090245 1.050601
_cons		1.20e-33	2.00e-32	-4.54	0.000	7.36e-48 1.96e-19
B2						
TINDEKSløpeordTeiningar_LOG		2854431	1.44e+07	2.95	0.003	148.9018 5.47e+10
LEDDINDEKSleddsTeiningar		.0094613	.0240118	-1.84	0.066	.0000654 1.368381
MAKSFRASEANTALLORD_LOG		.5160496	.792047	-0.43	0.666	.0254826 10.45054
KOORINDEKKoordTeiningar2		.9170627	.0456393	-1.74	0.082	.8318354 1.011022
KORRINDEKSTotaltkorrekteT_SQ		1.36e+19	1.68e+20	3.56	0.000	3.94e+08 4.68e+29
VARLEDDINDEKSSUBSTTypesubs		.0455193	.2609506	-0.54	0.590	6.00e-07 3450.758
VARLEDDINDEKSADVERBIALType2		1.003284	.0395356	0.08	0.934	.9287129 1.083844
_cons		7.60e-33	1.50e-31	-3.76	0.000	1.35e-49 4.27e-16

Note: _cons estimates baseline odds.

Ann-Kristin Helland Gujord
 Førsteamanuensis i norsk som andrespråk
 Institutt for lingvistiske, litterære og estetiske studium,
 Universitetet i Bergen
 Postbok 8805, NO 5020 Bergen
 ann-kristin.gujord@uib.no



Ummæle av *Språket som vart norsk*

Sverre Stausland Johnsen

Universitetet i Oslo

1 Innleiing

På hausten i 2019 kom ei ny målsogobok på marknaden: *Språket som vart norsk. Språkhistorie frå urnordisk til 1800-talet*. Ivar Berg, hågsynar i norderlendisk målkunna på NTNU, hev ritat henne, og ho er utkomi på Universitetsforlaget.

Det er munga målsogobøker fyre norsk å få kaupt no, og det er rimelegt å spyryja um det er torv til endå ei. Til det segjer Berg detta (s. 5):

[Me har] ikkje lenger [...] høvelege lærebøker; anten er dei for omfattande, eller dei har for lite eldre språkhistorie og særleg gamalnorsk (etter mitt syn).

Berg kjem ikkje inn på kvat bøker som hev desse lyti, men det er liklegt at han hugsar på Indrebø (2001), som er nokot romstor fyre nyemningar i dag, Torp og Vikør (2014), som hoppar yver det gamalnorska målstiget i utgreidingi, og Nesse (2013), som Berg sjølv hev dømt å ikkje vera høvande på allskulastiget (2013: 297).

Det stend ikkje i boki kven ho er etlad, men bak på umværet stend det at ho er «veleigna som pensumlitteratur ved universitet og høgskular». Det er ei helder tunn bok der utgreidingi er på vel 120 sidor, og det kann ikkje vera ev i at ho er etlad deim som ikkje hev voret burti onnor målsoga en det dei hev noyet på framhaldsskulen. Di må boki vera tiltakande på 1000-stiget, dei kallar, på allskulen.

Boki er deild i seks stykke. Fyrsta stykket (s. 15–25) er ei framlegging av kvat målsoga og målsogogransking er. Andra stykket (s. 27–39) dryfter up-

* Takk til Thorgeir Holm fyre gjennomlesnad og ordskitte.

phavet åt målet vårt og den frumnorderlendske tidbolken. Tridja stykket (s. 41–78) er um gamalnorsk, og er det lengsta stykket i boki. Fjorda stykket (s. 79–107) er um millomnorsk, og tek fyre seg korso gamalnorsk brigdest til dei nynorske målføri. Femta stykket (s. 109–123) heiter «Inn i nytida» og ser på danatidi og skiftet frå norskt til danskt bokmål i Norig. Setta og sidsta stykket (s. 125–134) er um namnet «norsk» og tokknaden um eit norskt mål. Bak i boki er det eit vedlegg på fjora sidor um målkunna, og heilt attarste boktilvisingar og eit upptal med uppslagsord.

2 Ålment

Sidan boki er so stutt, er ho høveleg på skeid der den gamla målsogo berre er ein lut av det næmingarne skal igjenom. Men på skular der eit innføringsskeid gjev myket rom åt den gamla målsogo, vert denne boki for grunn. Det er likt til at boki vart ritad fyre norderlendsklaupet på NTNU, der den gamla målsogo ikkje synest femna meir en ein tridjung elder fjordung av innføringsskeidet *Språkhistorie og talemål*. Men i dag synest dét vera vanen på flestalle skularne i landet, og på den visi hev boki funnet sin rette stad.

Ulikt Torp og Vikør (2014), der den «indre» og den «ytre» målsogo vert utlagda i kvar si bokhalva, vel Berg å leggja deim fram i lag innanfyre kvar tidbolk. Det er so me kjenner det or *Norsk målsoga* (Indrebø 2001), og ei sovori avdeildng er meir lagleg når ein skal lesa seg igjenom sogo.

Men som han kann sjå av namnet på boki, stadnar utgreidingi ved veldeskiftet i 1814. Berg segjer ikkje nokot um kvi boki stadnar der, og når ho er so stutt som ho alt er, burde det vera endafamt å leggja til eit stykke som fører henne fram til i dag. Eit nytt stykke um målsogo etter 1814 kann byggja på ei horg med kjeldor fyre bokmåli og målstriden i denne tidbolken, og når det gjeld talamálet, hev me no boki *Talemål etter 1800* (Sandøy 2015) å tyfta utgreidingi på.

Ugreido med ei målsogobok som stadnar i 1814, er at fyrelesaren fyre eit målsogoskeid då må hava ei onnor bok attved som tek fyre seg tidi etter 1814. Men flestalla målsogobøker tek fyre seg heila målsogo, og dimed endar det med at næmingarne sit med tvo bøker um den gamla målsogo. Skal ein sleppa frå eit slikt uføre, må ein finna ei målsogobok som tek til i 1814. Etter det eg veit, finst det berre éi sovori bok å få no, og det er Jahr (2019). Berg mælte um fyrsteutgjevi av den boki i 2016, og kom til at ho var «høveleg i undervisningssamanheng på lægre nivå» (Berg 2016a: 115). Det kann synast som boki *Språket som vart norsk* radt er ritad fyre eit skeid der Jahr (2019) vert lesi. Det

kann godt vera tilfellet på NTNU, men på andre skular kann synararne vilja bera seg undan *Språket som vart norsk*, um dei samstundes lyt binda seg til Jahr (2019).

Ein som vil fåst med å greida ut um gamalnorsk fyre næmingarne, kjem i det skorfestet at gamalnorsk samstundes er eit framandmål og det gamla heimamålet. Det ålmenna talamålet i Norig i dag er so myket avnorskat (Stausland Johnsen 2019) at næmingarne ikkje lenger hev eit godt grunnlag å standa på når dei vil taka til seg gamalnorsk. Men skular ein næmingarne i gamalnorsk som det var græsk, kann det vera vandt å kveikja hugen fyre det som endå er målet åt ættarfederne deira. I detta stykket hev Berg funnet ein god midveg. Han læst ikkje som gamalnorsk er nokot næmingarne kann taka til seg tvert, men gjerer endå ikkje målet meir framandt en det treng vera, og syner sambandi millom dei gamalnorske og nynorske ordskaði.

Men når det gjeld fyrelogorna i boki, er midvegen ikkje funnen. Her midar Berg både for hågt og lågt på einn gong. I jadaren er det i alt fjortan spurningar åt lesararne, men dei er jamt so letta at det ikkje trongst å gjeva upp svari på enden av kvart stykke. Samstundes er lesestykker i boki for vanskalege. Det er ingi umsetjingar og berre fåa ordtydingar og merknader. Eg trur til dømes ikkje at ein nyemning utan hjelp kjem til å skyna at gamalnorsk *morna* er ei bøyging av ordet *morginn*, elder kvat millomnorsk *trjodskædhes* skal vera for eit ord.

Boki stend på norsk. Målføringi er i nokon mun etter gamal gjerd, men samstundes nær det norsk-danska bokmålet på mange håttar – i alle høve frå min synstad. Men sume stader i boki tykkjer eg det er uturvande mange framandord som næmingarne snaudt kann skyna av sjølvom seg. Merke på det er det når me les «akkomodasjon», «divergens», «konvergens», «tradera» og «diferensiasjon». I fleste falli lyt Berg tyda deim ut med vanlege ord, kor som er, og kunde han ikkje då likso godt haldet seg til dei vanlege ordi i staden? Når det til dømes stend «eintal», «fleirtal» og «dualis» i ei og sama setningi (s. 67), må ein spyryja seg kvifyre det ikkje kunde stadet «total» i staden fyre «dualis».

I grunnbøker som denne er det oftaste fått med tilvisingar til bokheimen. Det er synd, for lesarar vil gjerna vita kvar ritaren hev heimtat tilfanget og utgreidingarne sina. Eg hev sjeldan seet nokon grunn til at tilvisingar skal vanta i grunnbøkerna, annat en eit tykke um det elles skulde gjera lesingi tyngre. Berg hev sjølv funnet åt andre ritarar fyre detta. I det nemnda ummælet av boki hans Jahr skriv Berg detta (2016a: 114):

Men eit problem som melder seg for undervisningsføremål, er mangelen på fortlopande referansar “for å lette lesinga” (s. 153; kanskje heller “skrivninga”?).

Men same skorten på tilvisingar er det i denne boki òg, endå um Berg stend på at det er «fleire referansar enn ein ofte ser i lærebøker» (s. 6). Det er kann henda nokot fleire, men det er lika fullt ovlega fåa, og det var mange stader i boki eg skulde likat å vita kvat kjeldo var. Det vert ei sjølvmotsegn når me er so stride med at næmingarne våre skal føra upp kjeldorna sina, samstundes som det vantar kjeldor i dei bøkerna næmingarne ser i.

Prentingi og setjingi av boki er god, og ho er væn å sjå til. Men eg skulde helst seet at den inste jadaren var større, so at ikkje ordi vart standande so langt inn. Ein må draga brædi vel ut, skal han sjå dei inste ordi. Sume stader er det lika vel nokra veilor. På s. 92 er det tvifeld skåskrift i groprodo, på s. 93–94 ser krullen på bokstaven <ç> annarleides ut når han stend einsaman en når han stend i bindslet <ç̄>. På sama sido stend tviprikk <: > i staden fyre lengdarteiknet <: >. På s. 97 stend tavlo langt ut i jadaren, og på s. 136 stend korkje bokstaven /ŋ/ elder skurden nedst på sido med oragravster (vektorgrafikk).

3 Mistak

Med desse ålmenne ordi sagde, er det tidi fyre å ganga grannare inn på einskilde drag og utgreidingar. Mest mætande er det me kann kalla røynelege mistak. Det er ikkje fleire av deim en at eg skal nemna alle eg gådde her, og vonlega kann dei rettast upp i næsta utgjevi.

På s. 28 vert det frumnordværska nemnefallsskapet av ordet *fader* atterskapat til **fader*, men det lyt ha voret **fadēr* (Tremblay 2003: 79).

På næsta sido fortel Berg at «[u]like austgermanske språk vart tala av folkegrupper som [...] til dømes langobardar [...]». Men langbardarne («langskjeggarne») var eit *väst-nordværskt*, ikkje *aust-nordværskt*, folk, og talamålet deira var nærskyldt det høgtysdka (Sonderegger 2003). Det sama stend i dei norske målsogobøkerna (Seip 1955: 5; Indrebø 2001: 33). Som sogogranskaren Ludwig Schmidt ritade alt i 1885 (s. 74):

Dass die Langobarden den Westgermanen, und zwar dem hochdeutschen Stamme angehörten, kann heutzutage Niemand mehr bestreiten.

I samanhaldingi med austnorderlendsk på s. 34 fær me høyra at «[u]rnordisk **wantur* vart [...] gamaldansk *vante*». Gamaldansk **wante* (nemnefallsskapet **wante* er ikkje kjent, berre underfallsskapet *wanta*, sjå Brøndum-Nielsen 1935: 203) er ein *an-stomn*, og kann di ikkje koma beint frå **wantur*, som er ein *u-stomn*.

Berg tidfester den gamalnorske tidbolken til åri 750–1350, og ritar på s. 36 at «før gamalnorsk tid fall kort /æ/ saman med /e/, slik at *i*-omljod av kort /a/ gjev /e/ i gamalnorsk». Samanfallet millom *æ* og *e* skal med andre ord ha hendt fyre 750. Men *æ* og *e* vert enno haldne greidt frå einannan i nordvestlendske og trøndske handrit på 12-talet, 500 år etter at den gamalnorske tidbolken tok til (Benediktsson 1964: 102–104).

På s. 37 kjem Berg med den sama gåløyso me finn i manga andra bøker um gamalnorsk, som er at «det aldri står *a* før ein *u* i neste staving, for då vert *a* alltid endra til *ø*». Men det er ikkje ljodbrigde av *a* framanfyre det bundna kjenneordet *-nu(m)* elder varaordet *-ðu* (Stausland Johnsen 2019: 92–93). Det er jamvel mange døme på detta i dei gamalnorske lesestykki i boki, der me millom annat finn <hjartanu> og <hafðu> (s. 74–75). Ein merksam næming kjem til å spyrja seg: Er so det *aldri* stend *a* fyre ein *u* i næsta stavingi, korso kann det heita *hafðu* då?

I bolken *Avljod* på s. 38 hev Berg rett i at ein lyt «halda omljod og avljod klårt frå kvarandre», men i sama utgreidingi blandar han deim i hop sjølv. I både dømi på ljodsprang i tavla 2.3 held Berg rotljodskiftet i *krupu* – *kropit* og *bugr* – *bogi* fyre å vera ljodsprang. Men skiftet *u* – *o* kjem av *a*-ljodbrigdet (Noreen 1923: 54). I ljodsprangrekkjo stend *u* og *o* på sama stiget, sidan både kjem frå **u* i frumnordværsk (1923: 143). Ein må elles spyrja seg kvifyre *a*-ljodbrigdet ikkje er nemnt i denne boki.

Når me kjem til den nynorske tidbolken, ritar Berg at målet i Aust-Telamork hev dei utjamnade ords-kapi /vot:o/ (norsk ljodskrift *våttå*) og /vuk:w/ (*vukku*) fyre *vita* og *vika* (s. 88). Men soverne ords-kap er ukjende i detta målvaldet. Han må heilt til Nord-Øyterdalen, åder ein høyrer *våttå* og *vukku*.

På næsta sidu held Berg fram at «me finn spor etter [jarning] i stadnamn [...] som [...] *Gåpåru* (gno. *Gaparud*) i Trøgstad». Her læt det som det gamalnorska namnet på garden er kjent, men det er det ikkje, so her skulde det stadet **Gaparud* med ei stjerna. Berg berrlegg ikkje kvat han byggjer på her, men det er likt til å vera Rygh (1897: 11), som ritar at *Gåpåru* «[k]unde tænkes at have lydt Gaparud og at have faaet sin nuv[ærende] Udtaleform ved ‘Ligedannelse’». Men detta kann ikkje vera rett, for ein vokster *a*–*a* > *å*–*å* er heilt ukjend i denne landsluten. Det upphavlega namnet på garden må di ha voret **Goparuð* (Hars-son 2010: 203; Hoel 2019: 147).

På s. 93 fær me *vita* at nynorsk /baʎ:/ kjem frå gamalnorsk «*ball* (AKK.SG)». Men det finst ikkje nokot namnord *ball* på gamalnorsk; skapet er *boll*. Ulikt det nynorska ordet *boll* ‘eista’, kann *ball* vera eit gamalt lån frå dansk elder tydk.

Nokot dilikt kjem upp på s. 96, der det stend at det gamalnorska skapet *þokk* hev gjevet nynorsk *takk*. Men det kann ikkje ha voret nokon ljodrett vokster *þokk* > *takk*, for gamalnorsk *o* vert aldri til *a* i denne stodo. Ordskapet *takk* må anten hava sjølvljodet frå eit annat bøyingssskap (jf. eig.et. *þakkar*, nem.mt. *þakkir*), elder, som er meir trulegt, ordet er lånt frå dansk.

4 Ulaglege døme

Det er langt fleire stader i boki der dei framlagde dømi er nokot ulaglege, og der utgreidingarne vert helder misvisande. Det er ikkje rom å lyfta fram alle desse dømi her, so eg skal berre leggja fram sume av deim.

I sama utgreidingi um ljodspranget nemnd ovanfyre (s. 38) legg Berg fram ordet *bugt* ‘buktt’ til døme på stiget *u* i gamalnorsk. Men detta ordet er eit lånord frå lågtydsk (de Vries 1962: 63) – det heimavaksna skapet er *bót* (som i *hombot*, *kalvbot* og *knesbot*). Sedvanen er å halda seg til nedervde ord når ein gjev døme på ljodsprang og ljodbrigde i eit mål.

På s. 87 ritar Berg at framburden av det gamalnorska ordet *sótt* i «moderne norsk» er /sot:/ (norsk ljodskrift *sått*). Men gamalnorsk *ó* vert oftaste hoven til /u/, og den fyresagde framburden, både på norsk og norsk-dansk, er med di /sut:/ (norsk ljodskrift *sott*), sjå m.a. Aasen (1873: 730), NO (10: 907), Alnæs (1925: 222), Berulfsen (1969: 299) og Vanvik (1985: 276). Segnaden *sått* er rett nok myket vanleg i dag, men den hev fyrst og fremst opphav i ei bokstavrett framsegging hjå deim som berre kjenner ordet or skrifti og ikkje or heimamålet sitt, og som difyre trur at ordet *sott* må rima på ord som *dott*, *knott*, *skrott*, *spott* og *vott*.

På s. 89 fær me høyra at «jamvektsmål på Austlandet [...] kan ha /vi:ta/». Berg segjer ikkje kvat han legg i «mål på Austlandet», men um me held oss til den vanlega tydingi av *austlendsk* (Jahr 1990: 24; Skjekkeland 1997: 20, 276; Papazian og Helleland 2005: 84–85; Mæhlum og Røyneland 2012: 44–45, 51), då er det ingi austlenske målføre der det heiter /vi:ta/. Dei einaste jamtyn-gdmáli med ordskapet /vi:ta/ er fulla nordvest i Telamork og øvst i Setesdalen. Detta målvaldet ligg på yvergangen millom midlendsk og vestnorsk, og med di tener dei ikkje vel til framsyning av «jamvektsmål på Austlandet».

På sama sido held Berg fram at «me finn spor etter [jamning] i stadnamn heilt ned til [...] Østfold». Men utjamning er ei velkjend og utbreidd ovring i talamálet i Austfold, so det er ingen grunn til å leita fram «spor i stadnamn». I Indre Austfold heiter det til dømes *fålå*, *bråná*, *nåva*, *rōko* og *suļu* frå gamalnorsk *folá*, *bruna*, *hnefa*, *reku* og *svølu* (Hoff 1946).

I dryftingi av brigde i medljodsamband i vestnorsk ritar Berg at «eit område i ytre Hordland har [...] *mm* > *bm*, som inf. *kobma*» (s. 92). Men segnordet *koma* hev ikkje ein gamal lang *mm*, so det er eit litet laglegt døme. Ein gløgg næming kjem til å spyrya kvifyre detta målføret hev *kobma* når ordet på gamalnorsk er *koma*. Her vøre det betre å finna eit ord med retteleg gamal *mm* til døme.

Nokot dilikt finn me i bolken um brigde i dei gamalnorske lange sjølvljodi på s. 96–97, der dømet me fær på tviljodingi av gamal lang /*ø*:/, er skapat /*øyl*/ i setesdalsk. Men detta ordet hadde ikkje gamal lang *ø* på gamalnorsk; ordet var *øl*. Det er uvisst kvifyre detta ordet hev fenget ei sovori tviljoding i setesdalsk (Skomedal 1971: 298), og difyre er det ikkje eit godt døme på denne voksteren. Det er mange ord med gamal *ø* som kunde ha stadet i staden.

I same bolken ritar Berg at det gamla stutta sjølvljodet «/u/ vert framleis uttala om lag slik [han] var i gamalnorsk». Det er myket misvisande, for gamalnorsk /u/ hev ålment fenget ein framskoten framburd, og er med di ikkje lenger eit innljod. Den gamle segnaden [u] er i dag undantaket, og me finn honom oftaste framanfyre nasaljod (Aasen 1864: 13).

Og sidst skal eg lyfta fram eit døme or ordfylgjo. Her dreg Berg fram at «syntaksen på mange vis har stått på gamalnorsk stadium gjennom mellomnorsk» (s. 101). Men det einaste dømet han legg fram på det, er at «me [finn] framleis brot på V2-regelen med verbet fyrst i forteljande setningar». Men den ordfylgjo er kjend frå nynorske målføre med (Mørck 2016: 339), so det var kann henda ikkje det besta grunnlaget fyre eit slikt påstand.

5 Usemja

I ei innføringsbok på detta stiget byr ein halda seg til det som er visst og uumstridt. Nokre stader gjev Berg seg lika fullt inn på gruningar, og i minsto tvo stader er eg usamd med den grunen han legg fram. Fyrsta fallet er det når Berg læt på at «det er tydeleg at det latinske alfabetet har vore den viktigaste modellen» fyre runorekkjo (s. 31). I røyndi er detta myket umstridt, sjå dryftingi og tilvisingar hjå Düwel (2008: 175–181) og Barnes (2012: 9–15).

Den andre staden er i dryftingi av eit domsbrev frå Veøy i Romsdal i 1439, der varaordet i 1.et. er stavat ⟨*æjk*⟩ (s. 101). Berg dømtar at «uttalen var *eig*», og der stavegjerd med *-k* er «frå skrifttradisjonen». Jamvel um det godt kann vera so (det kann me aldri få greida på), er det vandt å sjå korso Berg kann vera so viss um det. Når me veit at grannamåli på Nordmøre hev haldet på *-k* i detta

varaordet fram til våre dagar (Indrebø 2001: 119), er det fulla ikkje so orventes at det enno stod ein *-k* der i romsdalsk fyre 550 år sidan.

6 Norsk og nynorsk

I innføringsbøker um den norska målsogo var det sedvanen å segja «nynorsk» fyre tidbolken og målstiget etter gamalnorsk og millomnorsk (Hægstad 1924; Indrebø 1947). Då talade dei um nynorskt talamål, dei nynorske målføri og nynorskt bokmål. Men dei som ritade på norsk-dansk, tok til å segja «moderne norsk» i staden. Myket av grunnen til det låg fulla i ynsket å kunna få med det norsk-danska talamålet og bokmålet i umgripet. Kvat ugreida denne nya gjerdi hev ført til, ser me i nya målsogobøker. Der kann dei på eina sido umgripsfesta «moderne norsk» til målet som er framvakset or gamalnorsken, men samstundes føra upp ordskap or det norsk-danska bokmålet til døme på «moderne norsk», endå desse ordskapi hev upphavet sitt i dansk, ikkje gamal norsk (sjå t.d. Torp og Vikør 2014: 17, 90). Det er slikt som gjerer at næmingar i dag trur det norsk-danska bokmålet ættar frå gamalnorsken (Stausland Johnsen 2019: 124).

Dimed er det godt å sjå at Berg tek upp atter det gamla umgripet «nynorsk» fyre målstiget etter millomnorsk (s. 23–24). Men når det kjem til stykket, er det lika vel «moderne norsk» som gjeng atter i boki i staden. Grunnen til det er at «‘nynorsk’ [...] må ikkje forvekslast med det moderne skriftspråket nynorsk. For å unngå samanblanding kan me kalla den siste perioden ‘moderne norsk’» (s. 24). Men ei slik samanblanding ovrar seg berre um ein sjølv ikkje er skilleg i ordleggingi si. Er so han treng å skilja ut skriftmålet frå målføri, du vil segja, er det berre å segja *nynorskt skriftmål* og *nynorske målføre*, og det skal vera uråd å blanda deim i hop då (sjå t.d. Stausland Johnsen 2019). Når eg les målsoga fyre næmingarne, segjer eg berre «nynorsk» um målstiget, og eg hev aldri røynt at næmingarne hev vanskar med å venja seg til det.

Merke på at Berg legg sitt til at umgripet «nynorsk» vert uskilleg, kjem fram sume stader i boki:

I eitt ord finst [*u-oml*joden] framleis i nynorsk, nemleg *barn.SG – born.PL*, men nokre dialektar har den same vekslinga i fleire ord, som *vavn – vovn* og *land – lond* (s. 38).

Her skapar Berg ein motsetnad millom «nynorsk» og «dialektar». Men målføri er nynorske òg, so den einaste visi å skyna denne setningi på er at Berg i detta høvet vil at «nynorsk» berre skal vera namnet på det nynorska bokmålet. Med

di kann lesaren lett verta undren, sidan me fyrr hev leset at «nynorsk» er namnet på målstiget, og at me ikkje må mistaka «nynorsk» fyre namnet på skriftmålet.

Eit annat døme kjem på s. 62:

I nynorsk kan alle inkjekjønnsord bøyast på same måte som *hus*, men tre ord har frameis valfritt svak bøying: *auga*, *øyra* og *hjarta*.

Det er langt fleire en desse try inkjekynsordi med linn bøyging i dei nynorske målføri (Aasen 1864: 154), so her lyt me på nytt lesa «det nynorska bokmålet» i namnet «nynorsk».

Bak i boki gjeng Berg endå eit stig lenger. Her fær me høyra at «[i] moderne norsk er [kasus] berre relevant for pronomen» (s. 137), jamvel um me fyrr på s. 99 fekk vita at «[i] ein del dialektar finn me dessutan eigne dativformer av substantiv i bunden form og av nokre determinativ». Her kann det låta som målføri ikkje høyrer under «moderne norsk» helder. Lagt saman kann det synast som «norsk», sama um det vert kallat *nynorsk* eller *moderne norsk*, fyrst og fremst skal vera namnet på bokmålet, og so er målføri våre nokot annat som kjem attåt. Det er ei vanleg folkaleg syn at bokmålet er det rettelega målet og at talamålet kjem i andra rekkjo, men ei vitskapleg innføringsbok byr ikkje nora ei slik syn.

I setta hovudstykket i boki kjem Berg inn på sogo åt namnet *norsk*. I gamalnorsk fanst ikkje detta ordskapet, men dei hadde ordet *norrønn*, som tydde ‘norsk; or Norig’. Berg ritar so at «på 1300-talet dukkar *svensk* og *dansk* opp som språknamn, og står då i motsetjing til norrønt» (s. 126). Han segjer ikkje kvat han byggjer på her, men det er likt til å vera hans eget stykke frå 2016, der han tek eit myket større atterhald (2016b: 41):

[T]his identification of Swedish and Danish *should imply* that *norróna* now took on the specific meaning ‘West Norse’, i.e. the language of Norway and the Atlantic islands (mi utmerking).

I dei bokstykki Berg viser til, målber ikkje ordi *svensk* og *dansk* nokot motstykke til *norrønt*. Men Berg dømatar at sjølva tilværet av ordi *svensk* og *dansk* i dei landi er merke på at dei laut hava eit annat ord fyre talamålet i Norig. Det er uvisst. Korso kann me vita at svearne på den tidi ikkje totte nordmennerne talade «svensk» elder «dansk» au? I alle høve er detta altfor uvisst til at eit slikt atterhaldslaut påstand høyrer heima i ei innføringsbok.

Berg kann fortelja at «[s]iste gongen dette språknamnet [d.e. *norróna*] er bruka i norske kjelder, er i 1436», og at «fyrste gong namnet ‘norsk’ om språket» kjem upp, er i 1486 (s. 126–127). Men her læt det som me talar um tvau heilt ulike namn, og myket av ordsogo kjem burt når me ikkje fær vita at ordet *norsk* etter alt å døma er sama ordet som *norrøn*, berre at etterfestet *-sk* er pålagt: *norrónn* > *noren* + *-sk* > *nornsk* > *norsk* (Hægstad 1910; 1915).

Det finst nokre millomalderbrev frå åri fyre kyrkjoboti i 1537 der det stend at det vart haldna messor i Norig på «norsk» og «dansk». Berg tek det til merke på at dei ikkje skilde so grant millom desse måli på den tidi, sidan det i både høvi gjeld messor på «morsmålet» (s. 127). Kvat Berg legg i «morsmålet» her, er uvisst. Er det norsk elder dansk? I alle høve er det vandt å sjå kvifyre det ikkje spyrst um messor på både norsk og dansk her. Er det uråd at sume prestarne var norske og heldt messor på norsk, og sume var danar som heldt deim på dansk (jf. Holm 2020: 140)?

7 Prenteveilor

Suma prenteveilor er det i boki, mest i dei gamalnorske ordi. Her er dei eg gádde. Gamalnorsk: *dyrð* f. *dýrð* (s. 58, 72), *fót* f. *fótr* (s. 60), *vikur* f. *vikur* (s. 62), *lyð* f. *lýð* (s. 72), *syndi* f. *sýndi* (s. 72), *síða* f. *síða* (s. 75), *firir* f. *fyrir* (s. 75–76), *lǫligum* f. *lǫgligum* (s. 76). På s. 72 stend *Dávíðs* og *Daviðs* um einannan. Nynorsk: *lageleg* f. *lagleg* (s. 16), *villa* f. *vilja* (s. 67), *tiende* f. *tiend* (s. 79). I tilvisingarne bak i boki er *Hå-* sett fyre *Hø-*.

På s. 68 må det vera ein glepp, på gjerd, når Berg skal greida ut korso varaordet *ér* vart *þér*. Som kjent er, kjem *þ-* frå segnordendingi *-ð* når ho stod framfyre varaordet: *komið ér* > *komi þér* (Noreen 1923: 311). Den utgreidingi er det uråd å få tak på når Berg legg det fram soleides: «*ságuð þér* ‘såg de’ > *ságuðer*». Her hev varaordet skapet *þér* med *þ-* alt til å byrja med – det skulde fulla stadet *ér*? Elles er det ug Reidt fyre næmingen å skyna kvifyre det etterhengda varaordet *ér* misser sjølvljodlengdi si på s. 68 (*ságuðer*) når varaordet *mér* ikkje gjerer det på s. 67 (*ságumér*).

8 Utgang

Boki *Språket som vart norsk* kann vandalaust havast på grunnstiget i norsk målsoga på allskularne, men då krevst det at fyrelesaren finn ei onnor bok fyre målsogo etter 1814. Skal denne boki vera besta valet på eit grunnskeid, burde ho førast fram til notidi. Næsta utgjevi må òg retta upp dei mistaki som er upptalde

i detta ummælet (sjå bolck 3), og helst skifta ut dei ulaglege og misvisande dømi nemnde i bolck 4 her. Ho vert òg hævare fyre næmingarne um lesestykki fær fleire ordtydingar og merknader med seg, og um Berg berrlegg dei kjeldorna han hev nøytt.

Tilvisingar

- Alnæs, Ivar. 1925. *Norsk uttaleordbok*. 2. utg. Utgitt av Bymålslaget. Oslo: H. Aschehoug & Co. (W. Nygaard).
- Barnes, Michael P. 2012. *Runes. A handbook*. Woodbridge: Boydell Press.
- Benediktsson, Hreinn. 1964. Old Norse short *e*: One phoneme or two? *Arkiv för nordisk filologi* 79, 63–104.
- Berg, Ivar. 2013. Ummæle av Agnete Nesse *Innføring i norsk språkhistorie*. *Norsk Lingvistisk Tidsskrift* 31(2), 290–297.
- . 2016a. Ummæle av Ernst Håkon Jahr *Language planning as a sociolinguistic experiment. The case of modern Norwegian og Språkplanlegging og språkstrid. Utsyn over norsk språkhistorie etter 1814*. *Norsk Lingvistisk Tidsskrift* 34(1), 107–116.
- . 2016b. The making of the Scandinavian languages. *Metalinguistic perspectives on Germanic languages. European case studies from past to present*. Gijsbert Rutten og Kristine Horner styrde. Historical sociolinguistics 4. Oxford: Peter Lang, 35–55.
- Berulfsen, Bjarne. 1969. *Norsk uttaleordbok*. Utgitt av Bymålslaget. Oslo: H. Aschehoug & Co. (W. Nygaard).
- Brøndum-Nielsen, Johs. 1935. *Gammeldansk Grammatik i sproghistorisk Fremstilling*. Band 3: *Substantivernes Deklination*. København: J. H. Schultz.
- Düwel, Klaus. 2008. *Runenkunde*. 4. utg. Sammlung Metzler 72. Stuttgart: J.B. Metzler.
- Harsson, Margit. 2010. *Leksikon over norske rud-namn frå mellomalderen*. Oslo: Novus.
- Hoel, Kåre. 2019. *Bustadnavn i Østfold*. Band 18: *Trøgstad*. Tom Schmidt styrde. Oslo: Novus.
- Hoff, Ingeborg. 1946. *Skjetvemålet. Utsyn over lydvoeksteren i målet i Skiptvet i Østfold i jamføring med andre østfoldske mål*. Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo. II. Historisk-filosofisk klasse. 1946 1. Oslo: Jacob Dybwad.
- Holm, Thorgeir. 2020. Målbrotet på 1500-talet. *Norsk Årbok*, 118–146.

- Hægstad, Marius. 1910. Upphavet til ordet „norsk“. *Maal og Minne*, 51–52.
- . 1915. Smaating or maalsoga. *Maal og Minne*, 167–169.
- . 1924. *Norsk målsoga for skule og heim*. 4. utg. Norske folkeskrifter 22. Oslo: Noregs ungdomslag og student-mållaget.
- Indrebø, Gustav. 1947. *Norsk målsoge*. 3. utg. Bergen: A.S Lunde & Co.
- . 2001. *Norsk målsoga*. 2. utg. Bergen: Norsk bokreidingsforlag.
- Jahr, Ernst Håkon (styr.). 1990. *Den store dialektboka*. Oslo: Novus.
- . 2019. *Språkplanlegging og språkstrid. Utsyn over norsk språkhistorie etter 1814*. 2. utg. Oslo: Novus.
- Mæhlum, Brit og Unn Røyneland. 2012. *Det norske dialektlandskapet. Innføring i studiet av dialekter*. Cappelen Damm Akademisk.
- Mørck, Endre. 2016. Syntaks. *Norsk språkhistorie*. Band 1: *Mønster*. Helge Sandøy styrde. Oslo: Novus, 317–445.
- Nesse, Agnete. 2013. *Innføring i norsk språkhistorie*. Cappelen Damm Akademisk.
- NO = *Norsk ordbok*. 1966–2016. *Ordbok over det norske folkemålet og det nynorske skriftmålet*. Alf Hellevik styrde. 12 band. Oslo: Det norske samlaget.
- Noreen, Adolf. 1923. *Altnordische grammatik*. Band 1: *Altisländische und altnorwegische grammatik (laut- und flexionslehre) unter berücksichtigung des urnordischen*. 4. utg. Sammlung kurzer Grammatiken germanischer Dialekte 4. Halle (Saale): Max Niemeyer.
- Papazian, Eric og Botolv Helleland. 2005. *Norsk talemål. Lokal og sosial variasjon*. Kristiansand: Høyskoleforlaget.
- Rygh, O. 1897. *Norske Gaardnavne. Oplysninger samlede til Brug ved Matrikelens Revision*. Band 1: *Gaardnavne i Smaalenenes Amt*. Kristiania: Cammermeyers Boghandel.
- Sandøy, Helge (styr.). 2015. *Talemål etter 1800. Norsk i jamføring med andre nordiske språk*. Oslo: Novus.
- Schmidt, Ludwig. 1885. *Zur Geschichte der Langobarden*. Leipzig: Gustav Fock.
- Seip, Didrik Arup. 1955. *Norsk språkhistorie til omkring 1370*. 2. utg. Oslo: H. Aschehoug & Co. (W. Nygaard).
- Skjekkeland, Martin. 1997. *Dei norske dialektane. Tradisjonelle særdrag i jamføring med skriftmåla*. Kristiansand: Høyskoleforlaget.
- Skomedal, Trygve. 1971. Vokalsystemet i Sætedalsmålet. *Mål og namn. Studiar i nordisk mål- og namnegransking*. Hallvard Magerøy og Kjell Venås styrde. Oslo: Universitetsforlaget, 291–306.

- Sonderegger, Stefan. 2003. *Althochdeutsche Sprache und Literatur. Eine Einführung in das älteste Deutsch. Darstellung und Grammatik*. 3. utg. de Gruyter Studienbuch. Berlin: Walter de Gruyter.
- Stausland Johnsen, Sverre. 2019. Grunnleggjande spørsmål om språkending og norsk språkhistorie. Ein omtale av *Norsk språkhistorie I – Mønster*. *Norsk lingvistisk tidsskrift* 37(1), 77–138.
- Torp, Arne og Lars S. Vikør. 2014. *Hovuddrag i norsk språkhistorie*. 4. utg. Oslo: Gyldendal.
- Tremblay, Xavier. 2003. *La déclinaison des noms de parenté indo-européens en -ter-*. Innsbrucker Beiträge zur Sprachwissenschaft 106. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck. Abteilung Sprachwissenschaft.
- Vanvik, Arne. 1985. *Norsk uttaleordbok. A Norwegian pronouncing dictionary*. Fonetisk institutt. Universitetet i Oslo.
- de Vries, Jan. 1962. *Altnordisches etymologisches Wörterbuch*. 2. utg. Leiden: E. J. Brill.
- Aasen, Ivar. 1864. *Norsk Grammatik*. Omarbeidet Udgave af „Det norske Folkesprogs Grammatik“. Christiania: P. T. Mallings Forlagsboghandel.
- . 1873. *Norsk Ordbog med dansk Forklaring*. Omarbeidet og forøget Udgave af en ældre „Ordbog over det norske Folkesprog“. Christiania: P. T. Mallings Boghandel.

Sverre Stausland Johnsen
Universitetet i Oslo