

Lehecka, Tomas, Ekholm, Miranda, Håkans, Frida, Kortelainen, Emma, Lundström, Johanna, Tymiński, Mateusz och Österberg, Jessica 2022: Vilka faktorer predicerar behärsknigen av svenskans preteritumböjning? Om finska gymnasieelevers inläring av femton vardagliga verb. *Folkmålsstudier*, nr 60. S. 45–71. Doi: <https://doi.org/10.55293/fms.120812>

*Tomas Lehecka, Miranda Ekholm, Frida Håkans, Emma Kortelainen, Johanna Lundström, Mateusz Tymiński och Jessica Österberg*

## VILKA FAKTORER PREDICERAR BEHÄRSKNINGEN AV SVENSKANS PRETERITUMBÖJNING? OM FINSKA GYMNASIEELEVERS INLÄRNING AV FEMTON VARDAGLIGA VERB

*Predicting proficiency in Swedish past tense inflection: On the acquisition of fifteen common verbs by Finnish upper secondary students*

In this study, we examine how Finnish upper secondary students learning Swedish as a second language inflect Swedish verbs in the past tense. We analyze the answers by 351 students to an online quiz regarding the inflection of 15 common verbs using mixed effects logistic regression modeling. Our results indicate that the correctness of the answers in the verb inflection task can be predicted based on a combination of linguistic and extralinguistic variables. Specifically, with respect to linguistic variables, verb frequency and vowel shift (ablaut) are significant predictors of the correctness of the answers, whereas the traditional distinction between strong and weak verbs is not. With respect to extralinguistic variables, the correctness of the answers is significantly predicted by the students' gender, age/exposure and level of motivation. Furthermore, we demonstrate through error analysis that, initially, many students follow the same pattern for past tense formation; the verbs' infinitive form is interpreted as their stem and the past tense suffix *-de* is overgeneralized to most verbs. On a methodological level, we argue that multivariate analyses allow for a more detailed description of the acquisition order of language structures than what has been achieved by less sophisticated quantitative analyses in prior research, and we encourage other scholars to build on our work by expanding it both in terms of the number of verbs as well as the number of predictor variables.

### *1 Inledning<sup>1</sup>*

Den viktigaste ordklassen för svensk satsbildning är verb, både ur syntaktisk och ur semantisk synvinkel. Syntaktiskt utgör den finita verbfrasen, dvs. predikatsledet, det enda obligatoriska ledet i svenska satser. Semantiskt betecknar verb sakförhållanden eller aktioner som är avgörande för satsers betydelseinnehåll, och som bestämmer vilka deltagarroller de övriga satsleden kan uttrycka. Inte minst tidfäster verbets tempus den aktion eller det sakförhållande som satsen beskriver. Att använda rätt verb och rätt tempus är

---

<sup>1</sup> Vi vill tacka de två anonyma granskarna samt Folkmålsstudiers redaktörer för konstruktiva kommentarer till en tidigare version av denna artikel.

följaktligen en förutsättning för en välfungerande kommunikation. Med tanke på det ovan sagda är det något förvånande att forskningen i svenska som andraspråk hittills ägnat mer uppmärksamhet åt bisatsordföljd och speciesböjning (som inte är lika avgörande för kommunikativ framgång) än åt verb. Tidigare forskning om verb i inläraarsvenska har främst fokuserat på användningen av finita och infinita verbkonstruktioner (Bergman 1988, Muittari 2008) och behärsksningen av svenskans presensböjning (Paavilainen 2015). De övriga tempuskategorierna har inte studerats tidigare. Framförallt har inga tidigare studier använt sig av multivariata analyser för att undersöka hur inläraare tillägnar sig svenskans böjningsparadigm. I och med att andraspråksinläring är en ytterst mångfacetterad process (se t.ex. Ellis 1989, Doughty & Long 2003, Saville-Troike & Barto 2017) bör analysmetoderna inom andraspråksforskning återspegla denna komplexitet genom att inkludera statistiska modeller med flera oberoende variabler.

Föreliggande studie utgör en pilotundersökning av andraspråksinlärares behärsksning av svenskans verbböjning genom multivariata analyser – med utgångspunkt i hur finskspråkiga gymnasieelever behärskar preteritumböjningen. Specifikt demonstrerar studien hur blandade logistiska regressionsmodeller kan utnyttjas för att granska vilka av verbens böjningsparadigm som är lättare respektive svårare för finska andraspråksinläraare att tillägna sig, och vilka egenskaper hos inläraarna själva som samvarierar med hur väl de behärskar dessa paradigmat. Materialet för studien utgörs av enkätsvar av 351 finska gymnasieelever som läser svenska som B1-språk (dvs. fr.o.m. årskurs 6). Enkäten bestod av en uppgift i verbböjning med 15 vardagliga svenska verb (t.ex. *studera*, *gå*, *läsa* och *stjåla*) samt av frågor om elevernas bakgrund. I tillägg till blandad logistisk regressionsanalys använder vi kvalitativ och kvantitativ felanalys för att undersöka de svar eleverna gav i verbböjningsuppgiften.

Studien avser att besvara följande forskningsfrågor:

- I. Vilka preteritumformer, med avseende på verbens grammatiska egenskaper, behärskar finska gymnasieelever bättre respektive sämre?
- II. Vilka utomspråkliga (dvs. inläraarrelaterade) variabler predicerar hur väl gymnasieelever behärskar svenskans preteritumböjning?
- III. Vilka feltyper förekommer mest frekvent när finska gymnasieelever bildar preteritumformer av svenska verb?

Eftersom vårt material endast gäller ett begränsat urval verb är syftet med studien inte att dra generella slutsatser om andraspråkstälarens inläring av svenskans preteritumböjning. Avsikten är i stället att skapa en metodologisk

grund och en praktisk utgångspunkt för fortsatt forskning om hur andraspråkstalare tillägnar sig svenskans verbmorfologi. Med tanke på den avgörande roll som verb har för kommunikativ framgång är det viktigt att verbmorfologi får en framstående plats inom forskningen i svenska som andraspråk. Att vi blir medvetna om vilka inom- och utomspråkliga faktorer som påverkar inläringen av svenskans verbböjning och att vi identifierar de vanliga mönster som andraspråksinlärare följer vid preteritumbildning kan dessutom hjälpa oss att ta fram mer effektiva undervisningsstrategier för svenskans verbmorfologi.

I det som följer presenterar vi först tidigare forskning om de utom- och inomspråkliga faktorer vid andraspråksinläring som ingår som förklarande variabler i våra analyser. Därefter ger vi en kort översikt av regressionsanalys, och i synnerhet av blandad logistisk regressionsanalys, som en metod för att analysera effekten av flera förklarande variabler samtidigt. Efter detta beskriver vi specifikt det material och de metoder som vi använt för denna studie. Artikelns kärna utgörs av det avsnitt där vi redogör för resultaten av regressions- och felanalyserna. Slutligen diskuterar vi resultaten i relation till tidigare forskningsresultat och vi lyfter även fram implikationer för språkundervisning och fortsatt forskning i svenska som andraspråk.

## *2 Tidigare forskning*

### *2.1 Faktorer som påverkar andraspråksinläring*

Enligt den klassiska modellen av Ellis (1989) finns det grovt taget tre typer av faktorer som påverkar andraspråksinläring: (i) kognitiva, (ii) sociokulturella och (iii) inomspråkliga. De kognitiva faktorerna inbegriper fem variabler som bidrar till individuella inläringsskillnader: ålder, begåvning, motivation, personlighet och kognitiv stil<sup>2</sup>. De sociokulturella faktorerna omfattar bl.a. inlärningskontext, exponeringstid, andraspråkets ställning i samhället, men också t.ex. inlärarens ålder, kön och socialklass. Inomspråkliga faktorer inbegriper i sin tur målspråkets grammatiska egenskaper såsom fonologisk/morfologisk/syntaktisk struktur, typologisk markering och frekvensen av konstruktioner som lärs in. Att göra en entydig distinktion mellan kognitiva, sociokulturella och inomspråkliga faktorer är ofta svårt i och med att en och samma variabel kan betraktas ur olika perspektiv – t.ex. kan ålder betraktas både som en kognitiv och som en social faktor. I idealfallet skulle studier inom andraspråksforskning alltid beakta alla de ovannämnda

---

<sup>2</sup> *Kognitiv stil* syftar på ens sätt att uppfatta, bearbeta, lagra och använda information.

faktorerna och variablerna, men av praktiska skäl (bl.a. på grund av tidsbegränsning) är detta naturligtvis sällan en realistisk målsättning. Det är dock viktigt att studierna tar så många variabler som möjligt i beaktande.

I föreliggande studie granskar vi effekten av de utomspråkliga variablerna *kön*, *motivation* och *ålder/exponeringstid*, samt av de inomspråkliga variablerna *verbfrekvens* och *morfologisk form* (den sistnämnda operationaliserar vi som två separata variabler i vår analys; se avsnitt 4.1). Vi väljer att fokusera på dessa variabler av två orsaker. För det första kan de operationaliseras på ett relativt entydigt sätt i en webbenkätstudie (t.ex. är *ålder* och *kön* bekanta begrepp för alla, medan *kognitiv stil* inte är det). För det andra finns det starkt empiriskt stöd i tidigare forskning, som främst utgörs av univariata analyser (se referenser i följande avsnitt), för dessa variablers inverkan på andraspråksinläringen. Därmed lämpar sig dessa variabler bra för vårt syfte att åskådliggöra hur multivariata analyser kan användas för att studera effekten av flera ”viktiga” variabler samtidigt. Härnäst ger vi en kort översikt över tidigare forskning som rör de tre utomspråkliga variablerna och därefter redogör vi för tidigare forskning om verb i inläraresvenska som fokuserat på de inomspråkliga variablerna.

## 2.2 Utomspråkliga variabler: *kön*, *motivation*, *ålder/exponeringstid*

Inom internationell andraspråksforskning har man gett könsskillnaderna i andraspråksinläringen mycket uppmärksamhet. Generellt tenderar kvinnor att prestera bättre än män i uppgifter som mäter språkfärdigheter (Boyle 1987, Burstall 1975, Davies 2004). I en storskalig studie med över 27 000 andraspråksinlärare av nederländska (med sammanlagt 49 olika modersmål) visar Van der Slik, Van Hout och Schepens (2015) att det finns en tydlig könseffekt för inlärares tal- och skrivproduktion; de kvinnliga inläraarna presterar bättre än de manliga oberoende av ursprungsland eller modersmål. Däremot finns det enligt Van der Slik, Van Hout och Schepens (2015) ingen signifikant könsskillnad gällande hörförståelse. Den allmänna uppfattningen inom andraspråksforskningen är att könsskillnaderna i andraspråksinläring beror på rådande samhällsstrukturer snarare än på biologiska skillnader (Saville-Troike & Barto 2017).

Motivation är en annan utomspråklig variabel som fått mycket uppmärksamhet inom andraspråksforskningen. Csizér och Dörnyei (2005) beskriver motivation som en psykologisk drivkraft bakom beteenden med vilka en person vill minska klyftan mellan sitt faktiska jag och sitt ideala jag. Bilden av det ideala jaget kan för någon inkludera goda kunskaper i ett andraspråk, vilket också motiverar personen att arbeta för detta mål. I en

finländsk kontext har Marton och Kovács (2019) undersökt vad som motiverar finskspråkiga elever att lära sig svenska på en enspråkigt finsk respektive tvåspråkig ort. De visar att elever på den tvåspråkiga orten framförallt motiveras av svenskans integrativa värde (dvs. tanken på att kunna ta del av den svenskspråkiga gemenskapen) och av att se sig själv som en person med mångsidiga språkkunskaper. Elever på den enspråkigt finska orten motiveras däremot snarare av svenskans instrumentella värde, där språkkunskaper ses som en fördel i arbetslivet. Motivation är således beroende både av yttre omständigheter, såsom den ifrågavarande språkmiljön, och av individers personlighetsdrag. Vissa är inherent mer intresserade av främmande språk än andra och söker sig gärna till situationer där det talas olika språk. Enligt Moody (1988) och Ehrman (1990) korrelerar språkintresse framförallt med preferens för intuitivt tänkande, som kännetecknas av en god associations- och föreställningsförmåga.

Den tredje utomspråkliga variabeln, ålder, har framförallt behandlats i samband med diskussioner om den s.k. kritiska perioden för språkinläring (t.ex. Hyltenstam & Abrahamsson 2003, DeKeyser 2013). I föreliggande studie behandlas ålder dock inte som en kognitiv faktor. Ålder sammanfaller nämligen med en annan variabel som kan anses ha större betydelse för andraspråksinläringen hos gymnasieelever, nämligen mängden undervisning (eller exponering) som de har fått i språket i fråga. Exempelvis visar Genelot (1997) att det finns en stark och positiv korrelation mellan antalet undervisningstimmar som franska skolelever haft i engelska som andraspråk och deras faktiska språkfärdigheter i engelska (se även Nikolov 2009). Eftersom alla deltagare i vår studie läser svenska som B1-språk (dvs. de har alla börjat läsa svenska i samma ålder) fungerar ålder som en bra approximation av exponeringstid. Vi valde att operationalisera exponeringstid på detta sätt eftersom en fråga om ålder är lätt för studiedeltagare att besvara, medan en fråga om antalet kurser eller undervisningstimmar i svenska vore mindre entydig, i synnerhet i det årskurslösa gymnasiesystemet i Finland.

### *2.3 Verb i inlärarsvenska – inomspråkliga variabler*

Verb i inlärarsvenska har hittills främst undersökts utifrån processbarhetsteorins grundtes om att inläringen av språkliga strukturer följer en bestämd hierarki. Följaktligen har tidigare forskning fokuserat på att klarlägga ordningsföljden i vilken olika böjningskategorier tillägnas. Den mest omfattande redogörelsen för inläringen av svensk verbböjning presenteras i Paavilainen (2015). Hennes analys av finskspråkiga skolelevers språkbruk visar att dessa

tillägnar sig de morfologiska kategorierna infinitiv och presens innan de lär sig preteritum och supinum. Paavilainen konstaterar också att eleverna bäst behärskar presensböjningen av första konjugationens verb, dvs. svaga verb med stam som slutar på obetonat *-a*, t.ex. *rita*. Vidare noterar hon att första konjugationens böjningsparadigm övergeneraliseras så att verb som borde sluta på *-er* i presens i stället ofta får suffixet *-ar*, t.ex. *\*ätar*. Enligt Paavilainen kan detta förklaras med att första konjugationen täcker det största antalet verb i svenskan och att detta paradigm således står för den mest frekventa inputen som andraspråksinlärare får.<sup>3</sup> Enligt internationell forskning fungerar inputfrekvens som en av de starkaste prediktorerna för inläringen av språkliga strukturer i andraspråk, men även bl.a. strukturens funktion, entydighet och prototypiskhet är av betydelse (se t.ex. Ellis & Ferreira Junior 2009, Wulff m.fl. 2009). Vi har valt att inkludera både frekvens och morfologisk form som förklarande variabler i våra analyser.

Verbbruket i inläraresvenska har tidigare undersökts av bl.a. Muittari (2008), som studerat användningen av finita och infinita verbformer i uppsatser skrivna av finskspråkiga gymnasieelever. Muittari visar att finita former används mer än infinita på alla årskurser, men att andelen infinita verbformer växer från årskurs 1 till årskurs 3. Presens är den överlägset vanligaste tempusformen på alla årskurser, och verben i presens uppvisar också den högsta korrekthetsgraden. Korrekthetsgraden stiger i takt med årskurserna (se även Rahkonen & Håkansson 2008). Muittaris resultat överensstämmer med Bergman (1988) som jämförde det svenska verbbruket hos finskspråkiga och spanskspråkiga inlärare av svenska. Likaså överensstämmer resultaten med Philipssons (2007) slutsatser om att antalet infinita verbformer stiger ju högre nivå andraspråksinläraren uppnår. Enligt Philipsson tyder alla tecken på att andraspråksinlärare måste behärska enkla finita verbformer innan de kan lära sig använda sammansatta verbformer. Högfrekventa ord och uttryck kan dock också läras in som helsekvenser (Sundman 2011), vilket innebär att man exempelvis kan tillägna sig vissa preteritumformer innan själva principen för preteritumböjning har internaliserats.

Vissa forskare påstår att det finns ett universellt inlärningsmönster för tempusböjningen av olika verbtyper. Exempelvis argumenterar Lightbown och Spada (2013) för att preteritumböjningen tillägnas först för kopula- och hjälpverb, sedan för oregelbundna verb och sist för regelbundna verb. Arifin m.fl. (2019) hittar partiellt stöd för denna inlärningshierarki i en studie med indonesiska andraspråksinlärare av engelska; enligt deras resultat behärskas

---

<sup>3</sup> Man kunde dock också tänka sig att ändelsen *-ar* föredras p.g.a. att det formmässigt ligger närmast till verbens infinitivform, som allra oftast slutar på *-a*.

kopulaverb dock sämre än oregelbundna verb.<sup>4</sup> Dyliga inlärningshierarkier för olika verbtyper har inte undersökts för svenskans del. Föreliggande studie inkluderar endast 15 verb och vi kan därför inte heller uttala oss om eventuella inlärningshierarkier, men den kan utgöra en grund för fortsatt forskning om inlärningsprocesser gällande svenskans verbmorfologi.

### 3 Blandad logistisk regression

Begreppet *regressionsanalys* täcker en hel grupp av statistiska analysmetoder vars syfte är att belysa sambandet mellan en beroende och en eller flera oberoende (dvs. förklarande) variabler. Grundtanken bakom regressionsanalys är att skapa en modell (dvs. en förenklad representation av ett fenomen) som passar observerade data. Med 'modell' avses i praktiken en matematisk funktion som uttrycker förhållandet mellan den beroende och de oberoende variablerna. Funktionen kan sedan användas för prediktion, dvs. för att förutsäga värdet på en variabel, vars värde är okänt, utifrån värdet på variabler, vars värden är kända. En utmärkt introduktion till regressionsanalys och dess tillämpningar inom språkvetenskapen ges av Winter (2020). Andra mycket nyttiga källor om regressionsanalys är t.ex. Speelman (2014), Speelman, Heylen och Geeraerts (2018), Sonderegger, Wagner & Torreira (2018) och Gries (2021).

Den enklaste tekniken bland regressionsanalyser är *enkel linjär regression* som används för att undersöka det linjära sambandet mellan två variabler. Linjär regression förutsätter att den beroende variabeln är numerisk. När den beroende variabeln är kategorisk och dikotomisk (dvs. den kan endast ta två olika värden, t.ex. *ja/nej*, *rätt/fel*) använder man en annan teknik: *logistisk regression*. När man inkluderar fler än en oberoende variabel i en logistisk regressionsmodell kallas detta *multipel logistisk regression*, och generellt kallas statistiska analyser som inkluderar flera oberoende variabler *multivariata analyser*.

I normalfallet bygger både linjär och logistisk regression på antagandet att alla observationer i ens data är oberoende av varandra (dvs. observation A har ingen inverkan på observation B osv.). Detta krav uppfylls inte av experiment där samma studiedeltagare står för flera av datapunkterna, p.g.a. individuell variation mellan deltagarna. För att ta i beaktande dyliga beroendeförhållanden mellan observationerna i ens analyser bör man använda sig av *blandade regressionsmodeller* (eng. *mixed-effects regression models*).

---

<sup>4</sup> Detta kan dock bero på att kopulaverb har en något annorlunda funktion i engelskan än i indonesiskan (Steinhauer 2001).



Blandade regressionsmodeller är alltid multivariata, dvs. de inkluderar alltid fler än en oberoende variabel. Mer exakt inkluderar blandade modeller minst en oberoende variabel med s.k. *fast effekt* (eng. *fixed effect*) och minst en oberoende variabel med s.k. *slumpmässig effekt* (eng. *random effect*). Variabler med fast effekt antas ha en konstant/identisk inverkan på den beroende variabeln över alla tänkbara observationer av det fenomen som man studerar, medan variabler med slumpmässig effekt varierar beroende på vilket urval av observationer man granskar. Data för föreliggande studie utgörs av 5 220 observationer av preteritumböjda verb, men alla observationer är inte oberoende av varandra eftersom de kommer från 351 studiedeltagare och gäller 15 olika verb. Om vi hade ett annat urval av deltagare eller om uppgiften innehöll ett annat urval av verb skulle resultaten av studien vara annorlunda. Således bör studiedeltagare och verb behandlas som variabler med slumpmässig effekt i analysen.

Det bör noteras att tolkningen av blandade logistiska regressionsmodeller skiljer sig från tolkningen av både linjära regressionsmodeller och logistiska regressionsmodeller utan slumpmässiga effekter (se t.ex. Sonderegger, Wagner & Torreira 2018, Winter 2020). Detta gäller både utvärdering av modellens förklaringskraft (dvs. hur väl den passar observerade data) och utvärdering av effekten av de enskilda oberoende variablerna i modellen. I linjär regression brukar regressionsmodellens förklaringskraft uttryckas med det statistiska måttet  $R^2$  som räknas utifrån s.k. residualer, dvs. differensen mellan observerade och predicerade värden av den beroende variabeln i modellen. Logistiska modeller (såväl blandade som ”vanliga”) har inte residualer och därför utvärderas de i stället genom (i) ett likelihood-kvotest där modellen i fråga jämförs med en s.k. nollmodell (dvs. en modell som inte innehåller några variabler med fast effekt) och (ii) modellens prediktionsgrad (dvs. andelen korrekt predicerade utfall för den beroende variabeln). För att utvärdera effekten av de enskilda oberoende variablerna i blandade regressionsmodeller, inklusive blandade logistiska regressionsmodeller, granskar man först variablernas koefficienter (numeriska värden som uttrycker variablernas effekt i den matematiska funktionen med vilken värdet på den beroende variabeln kan prediceras). I de flesta statistikprogrammen brukar resultaten av blandade logistiska regressionsmodeller inkludera variablernas koefficienter med standardfel samt p-värden som har räknats med det s.k. Wald-testet. Numera brukar man dock avråda från användningen av Wald-testet i signifikansprövningen av regressionsmodeller (se t.ex. Agresti 2002: 172, Williamson m. fl. 2007). I stället prövas signifikansen av var och en av variablerna med ett likelihood-kvotest där modellen som inkluderar variabeln i fråga jämförs med en modell som exkluderar den. Resultaten av likelihood-kvotest rapporteras som  $\chi^2$ -värden (med tillhörande

p-värde) som anger om modellen med variabeln passar data signifikant bättre än jämförelsemodellen.

## 4 Material och metod

### 4.1 De undersökta verben

Svenskans verb brukar traditionellt indelas i svaga och starka verb, och de svaga verben vidare i tre konjugationer beroende på vilka suffixvarianter de kombineras med i preteritum (se SAG: Verb §1, §38, §50–62). Första konjugationens verb har en stam som slutar på obetonat *-a* och de får suffixet *-de* i preteritum (t.ex. *ritade*). Andra konjugationens verb har en stam som slutar på tonande eller tonlös konsonant och de får suffixet *-de* eller *-te* i preteritum (t.ex. *vävde*, *läste*). Tredje konjugationens verb har en stam som slutar på betonad vokal och de får suffixet *-dde* i preteritum (t.ex. *sydde*). Starka verb, som utgör den fjärde konjugationen, bildar oftast preteritum genom vokalväxling i verbstammen, t.ex. *ge* → *gav*. Hos ett fåtal starka verb sker dock inga förändringar i verbstammen, t.ex. *sova* → *sov*. Hos några verb från andra konjugationen bildas preteritum med både suffix och vokalväxling i verbstammen, t.ex. *välja* → *valde*. Således sammanfaller distinktionen mellan svaga och starka verb inte med distinktionen mellan verb som får vokalväxling och verb som inte får vokalväxling i preteritum. Därför behandlar vi dessa två dikotomiska böjningsdrag (svagt/starkt verb och vokalväxling/ingen vokalväxling) som två separata morfologiska variabler i vår analys.

Vår analys bygger på data från en verbböjningsuppgift med 15 vardagliga svenska verb. Vi presenterar verben i tabell 1 där de har grupperats enligt den traditionella indelningen i svaga och starka verb. Den tredje kolumnen i tabellen anger om verbet får vokalväxling i preteritum eller inte. Den sista kolumnen anger preteritumformens relativa frekvens (per 1 miljon löpord, avrundad till närmaste heltal) i Språkbankens PAROLE-korpus ([spraakbanken.gu.se/resurser/parole](http://spraakbanken.gu.se/resurser/parole)). Denna frekvens representerar en approximation för hur vanligt verbet är i svenskan överlag. Frekvensen utgör en av de inomspråkliga förklarande variablerna som vi inkluderar i regressionsanalysen.

Tabell 1. De undersökta verben och deras inomspråkliga egenskaper.

Lexem	Preteritumform	Vokalväxling	Relativ frekvens (per 1 M löpord)
<b>Svaga verb</b>			
Första konjugationen			
<i>studera</i>	<i>studerade</i>	Nej	13
Andra konjugationen			
<i>köpa</i>	<i>köpte</i>	Nej	49
<i>läsa</i>	<i>läste</i>	Nej	50
<i>tänka</i>	<i>tänkte</i>	Nej	151
<i>fortsätta</i>	<i>fortsatte</i>	Ja	102
<i>lägga</i>	<i>lade</i>	Ja	78
<i>säga</i>	<i>sade</i>	Ja	448
<i>välja</i>	<i>valde</i>	Ja	58
Tredje konjugationen			
<i>nå</i>	<i>nådde</i>	Nej	33
<b>Starka verb</b>			
<i>be</i>	<i>bad</i>	Ja	45
<i>ge</i>	<i>gav</i>	Ja	222
<i>gå</i>	<i>gick</i>	Ja	658
<i>ligga</i>	<i>låg</i>	Ja	203
<i>sova</i>	<i>sov</i>	Nej	23
<i>stjåla</i>	<i>stal</i>	Ja	7

De 15 verben i tabell 1 inkluderar nio svaga och sex starka verb. Nio av verben får vokalväxling i preteritum, medan sex av verben har identisk stam i presens och i preteritum. Vi valde ut de 15 verben utifrån ett antal på förhand bestämda kriterier. För det första strävade vi efter att inkludera verb från varje konjugation samt verb som varierar gällande förekomsten av vokalväxling. I synnerhet valde vi att inkludera flera verb ur andra konjugationen som får vokalväxling i preteritum, eftersom denna ”dubbla” böjning kan betraktas som den mest komplexa bland verbens böjningsparadigm. För det andra anpassade vi valet av verben enligt finskspråkiga gymnasieelevers nivå. Mer specifikt inkluderade vi i enkäten endast sådana verb som förekommer i läroboksserien *Grip chansen!* (<https://www.e-oppi.fi/sarja/gripchansen/>) som används i finska gymnasier. För det tredje strävade vi efter att inkludera både högfrekventa och lågfrekventa verb. Slutligen valde vi att exkludera hjälpverb, kopulaverb och verb vars böjning växlar mellan svag och stark (dvs.

verb för vilka det korrekta svaret enligt SAG inte är helt entydigt). Beslutet att begränsa antalet verb till 15 motiverades av att vi ville hålla verbböjningsuppgiften kort (< 5 minuter). Vi ansåg att vi på detta sätt kunde maximera antalet frivilliga deltagare i vår undersökning. De 15 verben utgör alltså inte något representativt urval av svenskans verb, men verbgruppen uppvisar en lämplig mängd variation gällande de för våra ändamål viktigaste variablerna: böjningsparadigm (dvs. starka/svaga verb), vokalväxling och frekvens.

#### 4.2 Enkät

Vi samlade in data för studien genom en webbenkät som vi skapade på plattformen *SoSci Survey* (Leiner 2019). All text i enkäten var på finska<sup>5</sup>, förutom de 15 verb som utgjorde uppgiften i svenskans preteritumböjning (se figur 1). Enkäten bestod av fem delar: (1) ett formulär för informerat samtycke, (2) frågor om deltagarens bakgrund, (3) uppgiften i verbböjning, (4) kontrollfrågor (se nedan) samt (5) en avslutande sida med de rätta svaren till uppgiften. Det var inte möjligt att gå bakåt till en tidigare sida i enkäten. Studiedeltagarna använde i genomsnitt 8 minuter för att fylla i enkäten i sin helhet.

Figur 1. Skärmdump av webbenkäten (en del av sidan med verbböjningsuppgiften).

**Tehtävä 1**

Tehtäväsi on taivuttaa ruotsin kielen verbejä.  
Kirjoita kunkin verbin viereen sen menneen ajan (preteritum) muoto ruotsiksi.  
Arvaa vaikka et olisikaan varma oikeasta vastauksesta! Älä käytä apuvälineitä, aläkä kysy apua.

Esimerkiksi:  
prata (puhua) - pratade (puhui)  
ta (ottaa) - tog (otti)

1. studera (opiskella) -  (opiskeli)

2. be (pyytää) -  (pyysi)

<sup>5</sup> Vi ville vara säkra på att alla elever förstår anvisningarna. Dessutom hade svensk text eventuellt kunnat påverka elevernas svar i uppgiften.

Uppgiften i verbböjning bestod av 15 svenska verb. Dessa var uppdelade på två sidor i enkäten med 5 respektive 10 verb. De 15 verben presenterades i identisk ordningsföljd för alla informanter. Uppgiften hade ingen tidsbegränsning. Vi gav studiedeltagarna följande instruktioner (här översatta till svenska): ”Din uppgift är att böja följande svenska verb. Skriv vid varje verb in dess preteritumform på svenska. Gissa även om du är osäker på vad det rätta svaret är. Använd inga hjälpmedel och be inte någon om hjälp.” Vi gav två exempel på verb som hade böjts enligt anvisningarna och vi inkluderade de finska översättningarna av alla verb i uppgiften i både grundform och preteritum för att minimera risken för att studiedeltagarna skulle missförstå vilken form vi frågade efter.

Efter uppgiften bad vi deltagarna att bekräfta att de hade gjort uppgiften självständigt och att de inte hade använt några hjälpmedel. Vi bad dem också bekräfta att de hade gjort sitt bästa i uppgiften och att de inte hade svarat på enkäten förut (vi försäkrade dem om att svaren på dessa frågor inte skulle ha några konsekvenser för dem). Utifrån svaren på dessa kontrollfrågor exkluderade vi ur våra analyser de deltagare som inte hade följt anvisningarna eller som inte hade tagit uppgiften på allvar.

### 4.3 Studiedeltagare

Vi spred enkäten genom e-postlistan för *Svensklärarna i Finland rf* – en förening för alla svensklärare i Finland med ca 1 500 medlemmar. Vi riktade vårt meddelande i första hand till gymnasielärare som undervisar i B1-svenska och bad dem att i mån av möjlighet inkludera enkäten i sin undervisning. Deltagande i enkäten var helt frivilligt. Enkäten var öppen under två veckors tid i april 2019.

Vi samlade in uppgifter om studiedeltagarnas kön, ålder, hemort, nuvarande yrke/syssla, hemspråk och motivation. Motivation operationaliserades som en ordinal variabel på en skala från -2 till 2 (där -2 = ”Jag är inte alls motiverad att förbättra mina kunskaper i svenska” och 2 = ”Jag är väldigt motiverad att förbättra mina kunskaper i svenska”). Vi fick in sammanlagt 462 svar, varav 437 kom från gymnasieelever som läser svenska som B1-språk. I våra analyser inkluderade vi de deltagare som angav ’gymnasieelev’ som yrke/syssla, ’finska’ som hemspråk och ålder mellan ’16’ och ’20’. Dessutom angav de att de hade försökt sitt bästa, att de inte hade deltagit i enkäten förut och att de inte hade använt sig av hjälpmedel.

Utifrån de ovanstående kriterierna kunde vi inkludera svaren från 351 deltagare i våra analyser. Av dessa var 223 kvinnor, 125 män och 3 identifierade sig som annat. De två överlägset största åldersgrupperna var 17-

åringar (124 elever) och 18-åringar (177 elever). En stor del av eleverna kom från Egentliga Finland (116 elever) eller Norra Österbotten (95 elever) men i övrigt var fördelningen mellan hemorter mycket heterogen med relativt få elever från respektive region. På grund av detta valde vi att inte inkludera hemort som variabel i regressionsanalysen.

#### 4.4 Analysmetoder

I första steget av analysen gick vi igenom svaren i verbböjningsuppgiften och klassade dem som antingen *rätt* eller *fel*. Vi följde SAOL 14 (2015) för att definiera vad det rätta svaret var för respektive verb. Alla andra svar klassade vi som fel svar. Sedan använde vi oss av blandad logistisk regression (se avsnitt 3) för att komma underfund med vilka utomspråkliga (dvs. inlärrrelaterade) och inomspråkliga (dvs. verbrelaterade) variabler som har en inverkan på att någon behärskar preteritumformerna. Svarens Korrekthet<sup>6</sup> (*rätt/fel svar*) utgjorde den dikotomiska beroende variabeln i regressionsanalysen, medan Kön, Ålder, Motivation, Verbfrekvens, Vokalväxling och Böjningsparadigm (variablerna med fast effekt) samt Verb och Studiedeltagare (variablerna med slumpmässig effekt) utgjorde de oberoende variablerna vars effekt vi prövade i analysen. Vi operationaliserade Verbfrekvens för respektive preteritumform som logaritmen ( $\log_{10}$ ) av dess relativa frekvens<sup>7</sup> i PAROLE-korpusen i Språkbanken ([spraakbanken.gu.se/resurser/parole](http://spraakbanken.gu.se/resurser/parole); se tabell 1). Vokalväxling och Böjningsparadigm operationaliserades som dikotomiska variabler (*ja/nej* respektive *stark/svag*) i enlighet med uppgifterna i tabell 1. Eftersom det var så få deltagare som angett 'Kön: Annat' exkluderade vi dessa deltagare ur regressionsanalysen. Vi utförde alla analyser med programmeringsspråket R (R Core Team 2021). Vi använde funktionen *glmer* i R-paketet *lme4* (Bates m.fl. 2015) för att utföra regressionsanalysen och paketet *ggplot2* (Wickham 2016) för att skapa figurerna i denna artikel.<sup>8</sup>

Utöver regressionsanalysen genomförde vi även en mer detaljerad felanalys av elevernas svar gällande verbens preteritumformer. I felanalysen kategoriserade vi de skriftliga svaren utifrån deras grammatiska egenskaper. Vi fokuserade framförallt på den morfologiska strukturen hos felsvaren, dvs. hur dessa former hade bildats (t.ex. med suffixet *-de* eller med vokalväxling).

---

<sup>6</sup> För åskådllighetens skull markerar vi variablerna i regressionsanalysen med stor begynnelsebokstav i texten.

<sup>7</sup> Detta är vanlig praxis i kvantitativa analyser som berör ordfrekvens (se t.ex. Baayen 2008).

<sup>8</sup> R-koden för analyserna kan fås av första författaren på begäran.

Eftersom vårt syfte var att upptäcka allmänna mönster gällande preteritumböjningen hos finskspråkiga elever, inkluderade vi i felanalysen endast de svar som har minst 5 belägg i våra data (dvs. svar som gavs av minst 5 deltagare).

## 5 Resultat

### 5.1 Korrekthetsprocent

Vi inleder resultatavsnittet med en översikt över andelen korrekta svar i verbböjningsuppgiften. Tabell 2 åskådliggör att korrekthetsprocenten varierar starkt mellan de enskilda verben. Vid en närmare granskning av tabell 2 går det att urskilja några relativt tydliga grupperingar bland verben. De fem första verben (*gå*, *studera*, *tänka*, *köpa* och *läsa*) behärskas av de flesta gymnasieeleverna (65–72 %), medan de fyra följande (*välja*, *ge*, *säga* och *sova*) behärskas av knappt hälften (42–49 %). De övriga verben i uppgiften behärskas endast av en relativt liten andel av eleverna (< 30 %). Verbet *stjåla* har den överlägset lägsta korrekthetsprocenten av alla verb i uppgiften. Tabell 2 åskådliggör också att det inte finns något entydigt samband mellan konjugation och korrekthetsprocent. Både *gå* och *stjåla* (med 72 respektive 3 procent korrekta svar) tillhör den fjärde konjugationen. Vi går nu över till att granska resultaten av den blandade logistiska regressionsanalysen för att titta närmare på vilka inom- och utomspråkliga variabler som har en signifikant inverkan på elevernas behärskning av preteritumformerna.

Tabell 2. Verbens korrekthetsprocent i verbböjningsuppgiften.

Verb	Konjugation	Korrektthetsprocent
<i>gå</i>	4	72 %
<i>studera</i>	1	70 %
<i>tänka</i>	2	69 %
<i>köpa</i>	2	68 %
<i>läsa</i>	2	66 %
<i>välja</i>	2	49 %
<i>ge</i>	4	47 %
<i>säga</i>	2	46 %
<i>sova</i>	4	42 %
<i>nå</i>	3	29 %
<i>lägga</i>	2	27 %
<i>ligga</i>	4	23 %
<i>fortsätta</i>	2	17 %
<i>be</i>	4	16 %
<i>stjäla</i>	4	3 %

## 5.2 Regressionsanalys

I tabell 3 återges resultaten av den blandade logistiska regressionsanalysen i det format som de presenteras i programmet R (se avsnitt 4.4).<sup>9</sup> Med tanke på studiens syfte hittas den viktigaste informationen i tabell 3 under ”Fixed effects” (dvs. variabler med fast effekt), där kolumnen ”Estimates” anger variablernas koefficienter i modellen. Den sista kolumnen anger om variablerna har en signifikant inverkan på den beroende variabeln i modellen utifrån Wald-testet. Resultaten tyder på att alla oberoende variabler i modellen utom Böjningsparadigm (dvs. distinktionen mellan starka och svaga verb) bidrar till modellens förklaringskraft. Eftersom Wald-testet dock anses problematiskt av många forskare (se avsnitt 3) prövade vi också signifikansen av Böjningsparadigm med ett likelihood-kvottest där vi jämförde modellen som inkluderar alla variabler mot en modell som inkluderar alla variabler utom Böjningsparadigm. Likelihood-kvottestet bekräftar att modellen inte förbättras av att variabeln Böjningsparadigm inkluderas i den ( $\chi^2 = .45$ ,  $df = 1$ ,  $p = .50$ ). Vi diskuterar implikationerna av detta resultat i avsnitt 6.

<sup>9</sup> Vi använder detta format för att underlätta tolkningen av resultaten både för dem som redan har erfarenhet med regressionsanalys i R och för dem som vill lära sig att göra sådana analyser, i.o.m. att R är det verktyg som de flesta kvantitativt orienterade språkvetare föredrar att använda för sina analyser.



Tabell 3. Resultaten av den blandade logistiska regressionsanalysen.

---

```

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial (logit)
Formula: Korrekt_svar ~ Kön + Ålder + Motivation + Log_Verbfrekvens +
Verbparadigm + Vokalväxling + (1 | Verb) + (1 | Deltagare)
Data: Verbböjningsuppgift_clean_data
Control: glmerControl(optimizer = "bobyqa")

          AIC      BIC   logLik deviance df.resid
4434.1    4493.1  -2208.0  4416.1    5211

Scaled residuals:
   Min       1Q   Median       3Q      Max
-10.2127  -0.3984  -0.0842   0.3915  11.7581

Random effects:
 Groups      Name      Variance Std.Dev.
Deltagare (Intercept) 3.665    1.914
Verb (Intercept)      1.274    1.129
Number of obs: 5220, groups: Deltagare, 348; Verb, 15

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -23.4699    2.9955  -7.835 4.69e-15 ***
KönMan       -0.5635    0.2476  -2.276 0.0229 *
Ålder        0.9408    0.1525   6.169 6.86e-10 ***
Motivation   0.7955    0.1118   7.116 1.11e-12 ***
Log_Verbfrekvens 2.4736    0.6218   3.978 6.95e-05 ***
VerbparadigmSvag 0.4461    0.6576   0.678 0.4976
VokalväxlingNej 2.9833    0.7231   4.125 3.70e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr) KönMan Ålder  Motvtn Lg_Vrb VrbprS
KönMan       -0.025
Ålder        -0.895 -0.009
Motivation   0.046  0.314 -0.092
Lg_Vrbfrkvn -0.423 -0.003  0.012  0.013
Vrbprdgmsvg -0.049  0.000  0.001  0.002 -0.120
VoklvxlngNj -0.223 -0.004  0.012  0.014  0.432 -0.396

```

---

Regressionsmodellen som bäst passar våra data inkluderar således de följande oberoende variablerna: Kön, Ålder, Motivation, Verbfrekvens och Vokalväxling (variabler med fast effekt), samt Verb och Studiedeltagare (variabler med slumpmässig effekt). Vi kallar hädanefter denna modell Modell A. Modell A predicerar Korrekthet rätt för 86 procent av de 5 220 observationerna. Det kan nämnas att modellen som inkluderar Böjningsparadigm har samma prediktionsgrad: 86 procent. Detta förstärker slutledningen att variabeln Böjningsparadigm är ”onödig” i modellen.

Prediktionsgraden hos Modell A är klart högre än hos den s.k. nollmodellen (dvs. modellen som inte inkluderar några variabler med fast effekt) som predicerar rätt utfall för endast 57 procent av observationerna. Likelihoodkvottest bekräftar att Modell A är signifikant bättre än nollmodellen ( $\chi^2 = 126$ ,  $df = 5$ ,  $p < .0001$ ) och att de enskilda variablerna är signifikanta prediktorer för Korrekthet: Kön ( $\chi^2 = 5.12$ ,  $df = 1$ ,  $p = .02$ ), Ålder ( $\chi^2 = 36.6$ ,  $df = 1$ ,  $p < .0001$ ), Motivation ( $\chi^2 = 48.3$ ,  $df = 1$ ,  $p < .0001$ ), Verbfrekvens ( $\chi^2 = 11.2$ ,  $df = 1$ ,  $p = .0008$ ) och Vokalväxling ( $\chi^2 = 13.8$ ,  $df = 1$ ,  $p = .0002$ ). Också de två variablerna med slumpmässig effekt bidrar till modellens förklaringskraft: Verb ( $\chi^2 = 709.4$ ,  $df = 1$ ,  $p < .0001$ ) och Studiedeltagare ( $\chi^2 = 1043.7$ ,  $df = 1$ ,  $p < .0001$ ).

Koefficienterna för variablerna med fast effekt i tabell 3 belyser närmare hurdan inverkan variablerna har på den beroende variabeln. Strängt taget anger koefficienterna den naturliga logaritmen av förändringen i oddset att den beroende variabeln har värdet 'rätt svar'. Positiva koefficienter innebär ett positivt samband, medan negativa koefficienter innebär ett negativt samband. Exempelvis visar tabell 3 att det finns ett positivt samband mellan svarens Korrekthet och Ålder ( $\approx$  exponeringstid; se avsnitt 2.2), vilket betyder att sannolikheten att ett visst svar i våra data är rätt är högre ju äldre studiedeltagaren ifråga är. För kategoriska variabler specificerar tabellen vilket värde av variabeln som jämförs mot variabelns referensvärde, t.ex. för Kön anger tabellen den logaritmerade förändringen i oddset om variabelns värde ändras från referensvärdet 'Kön: Kvinna' till 'Kön: Man'. Att den ifrågavarande koefficienten är negativ betyder att sannolikheten att ett visst svar i våra data är rätt är lägre om svaret getts av en manlig deltagare än om det getts av en kvinnlig deltagare.

För att underlätta tolkningen av koefficienterna illustrerar vi i figur 2 hur respektive oberoende variabel förhåller sig till andelen korrekta svar i verbböjningsuppgiften. Figurerna visar effekten av variablerna var för sig (dvs. de illustrerar univariata analyser), vilket utgör en intressant jämförelsepunkt för resultaten av vår multivariata analys. Figur 2a visar att de kvinnliga eleverna behärskar preteritumformerna bättre än de manliga eleverna. Detta överensstämmer med den negativa koefficienten för 'Kön: Man' i tabell 3. Medianen för antalet korrekta svar (av 15 möjliga) är åtta för kvinnliga studiedeltagare och sex för manliga studiedeltagare. Figur 2b illustrerar den redan nämnda effekten av Ålder. Äldre elever har läst svenska längre och detta syns i deras svar: medianen för antalet korrekta svar stiger från fyra hos 16-åringar till nio hos 19-åringar<sup>10</sup>. Figur 2c påvisar en påfallande effekt av Motivation som motsvaras av en positiv koefficient för Motivation i tabell 3.

---

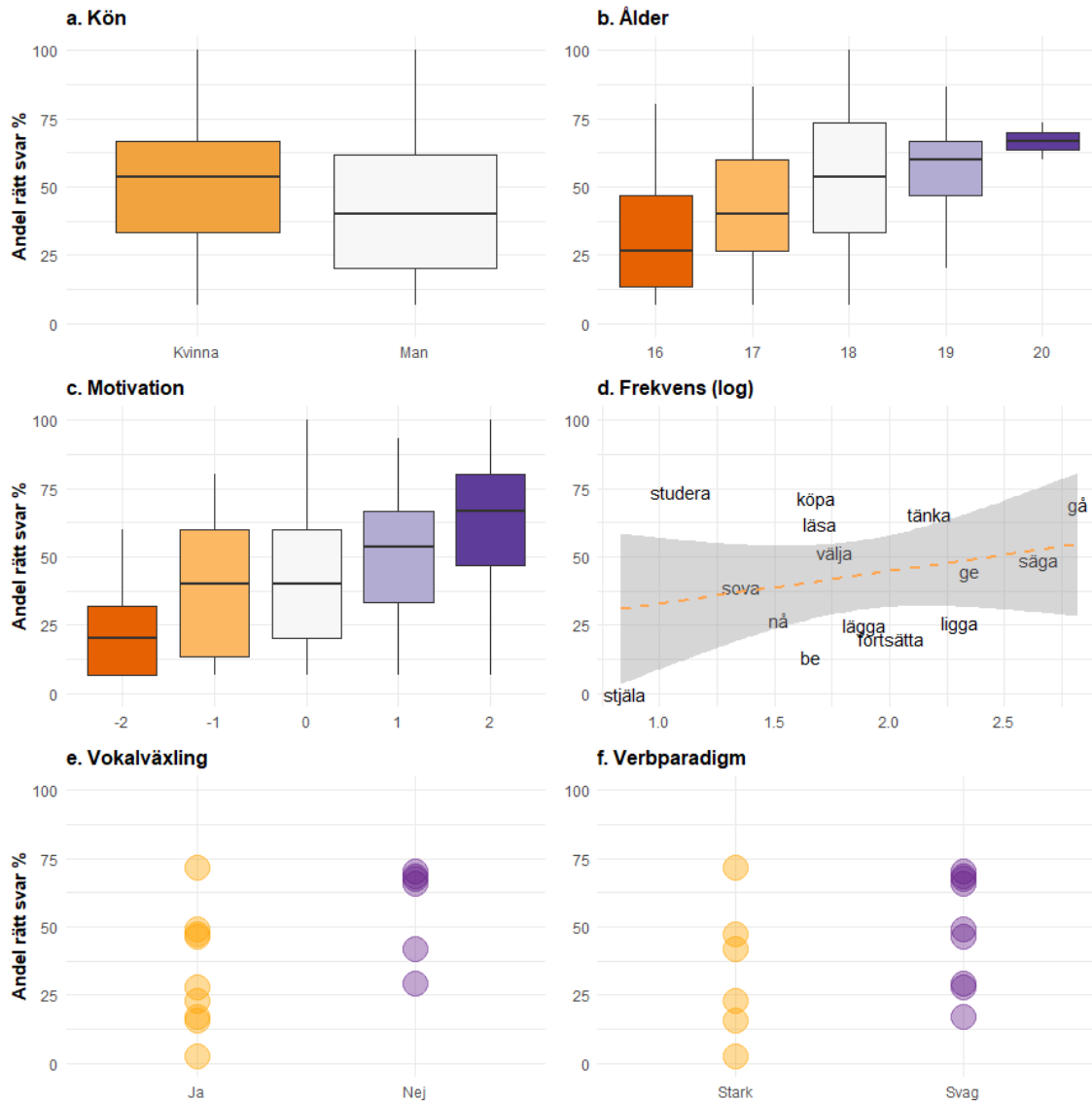
<sup>10</sup> Antalet 20-åringar i våra data är lågt så vi kommenterar inte deras resultat.

För de elever som angav att de inte alls är motiverade att förbättra sina kunskaper i svenska är medianen tre rätt svar av 15, medan medianen är tio rätt för de som angav att de är väldigt motiverade. Figur 2d åskådliggör ett positivt samband mellan andelen korrekta svar och Verbfrekvens: ju mer frekvent en verbform är i svenskan, desto mer sannolikt behärskas den av eleverna.<sup>11</sup> Figur 2e illustrerar sambandet mellan Korrekthet och Vokalväxling. Varje halvtransparent cirkel i figuren motsvarar ett av de 15 verben. Figuren visar att verben som har samma vokal i stammen i både presens och preteritum behärskas bättre än verben där preteritum bildas med vokalväxling. Slutligen illustrerar figur 2f sambandet mellan Korrekthet och Verbparadigm. Samtidigt åskådliggör figuren fördelen med multivariata analyser. Om vi endast hade analyserat vilken inverkan Verbparadigm har på Korrekthet (utan att ta de andra variablerna i beaktande) hade det varit lockande att på basis av figur 2f dra slutsatsen att stark böjning innebär svårigheter för inläringen av svenskans preteritumformer. När vi däremot gör en multivariat analys som tar i beaktande effekterna av Vokalväxling och Verbfrekvens samt effekterna av studiedeltagarnas Ålder, Kön och Motivation är effekten av Verbparadigm inte signifikant.

---

<sup>11</sup> Det är värt att notera att verbet *stjåla* utgör ett relativt extremt fall gällande både Korrekthet och Verbfrekvens (såsom framgår av Figur 2d). För att granska om detta extrema värde påverkar våra resultat genomförde vi den blandade logistiska regressionsanalysen även utan att inkludera verbet *stjåla* i våra data. Den signifikanta effekten av Verbfrekvens består dock även i detta fall ( $\chi^2 = 4.54$ ,  $df = 1$ ,  $p = .03$ ).

Figur 2. Förhållandet mellan de oberoende variablerna och Korrekthet var för sig.



För att komma underfund med hur stor effekt respektive oberoende variabel har på den beroende variabeln är det nyttigt att granska vilka prediktioner Modell A (dvs. modellen med den bästa förklaringskraften) ger upphov till. Vi kan utifrån modellen predicera sannolikheten att ett svar i verbböjningsuppgiften är rätt när värdet på en av de oberoende variablerna ändras men värdet på de övriga variablerna hålls konstant. Gällande variabeln Kön är den predicerade sannolikheten av ett rätt svar 14 procent när svaret ges av en kvinnlig deltagare (konfidensintervall 95 % CI [.06, .26]) och 8 procent (95 % CI [.04, .18]) när det ges av en manlig deltagare.<sup>12</sup> Gällande variabeln Ålder

<sup>12</sup> Dessa predicerade sannolikheter gäller när Ålder = 17.48, Motivation = 0.51, Verbfrekvens = 1.85 och Vokalväxling = Ja.

stiger den predicerade sannolikheten av ett rätt svar från 4 procent (95 % CI [.01, .09]) för 16-åriga deltagare till 63 procent (95 % CI [.36, .84]) för 20-åriga deltagare.<sup>13</sup> Gällande variabeln Motivation sträcker sig den predicerade sannolikheten av ett rätt svar från 2 procent (95 % CI [.01, .06]) för deltagare som inte alls är motiverade att förbättra sina kunskaper i svenska till 34 procent (95 % CI [.18, .55]) för deltagare som angav att de är väldigt motiverade.<sup>14</sup> Dessa resultat tyder således på att Ålder har den största effekten på elevernas behärskning av svenskans preteritumböjning i.o.m. att förändringen i den predicerade sannolikheten av ett rätt svar är störst när Ålder varierar. Effekten av Kön är däremot relativt liten i jämförelse med både Ålder och Motivation. Likadana prediktioner kan göras också gällande de inomspråkliga variablerna, men eftersom våra data endast täcker 15 verb bör dessa prediktioner tolkas med viss reservation. Gällande Verbfrekvens stiger den predicerade sannolikheten av ett rätt svar från 0 procent (95 % CI [.00, .02]) för verb som förekommer en gång per en miljon löpord till 74 procent (95 % CI [.41, .92]) för verb som förekommer 1 000 gånger per en miljon löpord.<sup>15</sup> Gällande Vokalväxling är den predicerade sannolikheten av ett rätt svar 14 procent (95 % CI [.06, .27]) för verb som får vokalväxling i preteritum och 79 procent (95 % CI [.58, .91]) för verb som har samma stamvokal i presens och preteritum.<sup>16</sup>

### 5.3 Felanalys

I det följande gör vi en kvalitativ granskning av svaren i verbböjningsuppgiften där vi sätter fokus på vilka typer av fel eleverna gör. Tabell 4 ger en överblick över svaren per respektive verb. De första kolumnerna i tabellen anger antalet rätt och fel svar samt antalet olika felsvar som har getts av minst 5 studiedeltagare. De två sista kolumnerna anger det mest frekventa felsvaret, respektive vilken böjningsform detta svar representerar. I de fall, där flera felsvar har ett markant större antal belägg i förhållande till de andra felsvaren för verbet, har de två vanligaste felformerna inkluderats i tabellen. Asterisker anger att svaren representerar en ogrammatisk form.

---

<sup>13</sup> Dessa predicerade sannolikheter gäller när Kön = Kvinna, Motivation = 0.51, Verbfrekvens = 1.85 och Vokalväxling = Ja.

<sup>14</sup> Dessa predicerade sannolikheter gäller när Kön = Kvinna, Ålder = 17.48, Verbfrekvens = 1.85 och Vokalväxling = Ja.

<sup>15</sup> Dessa predicerade sannolikheter gäller när Kön = Kvinna, Ålder = 17.48, Motivation = 0.51 och Vokalväxling = Ja.

<sup>16</sup> Dessa predicerade sannolikheter gäller när Kön = Kvinna, Ålder = 17.48, Motivation = 0.51 och Verbfrekvens = 1.85.

Tabell 4. Verbspecifik överblick av felsvaren i verbböjningsuppgiften.

Verb	Rätt	Fel	Olika felsvar	Vanligaste felsvar	Böjningsform
<i>gå</i>	267	110	3	<i>går</i> (35)	Presens
<i>studera</i>	261	117	5	<i>studerat</i> (37)	Supinum
<i>köpa</i>	256	122	4	<i>köpt</i> (58)	Supinum
<i>tänka</i>	254	132	7	<i>*tänkade</i> (33)	*Preteritum
<i>läsa</i>	243	134	6	<i>läser</i> (27)	Presens
<i>välja</i>	184	190	8	<i>*väljade</i> (38)	*Preteritum
<i>säga</i>	175	202	11	<i>såg</i> (31), <i>säger</i> (25)	Fel verb, Presens
<i>ge</i>	174	193	8	<i>ger</i> (40)	Presens
<i>sova</i>	155	221	7	<i>*sovde</i> (57), <i>*sovade</i> (51)	*Preteritum
<i>nå</i>	105	252	11	<i>*nåde</i> (51)	*Preteritum
<i>lägga</i>	99	265	11	<i>*läggade</i> (56)	*Preteritum
<i>ligga</i>	82	293	10	<i>*liggade</i> (76)	*Preteritum
<i>fortsätta</i>	61	311	14	<i>*fortsättade</i> (73)	*Preteritum
<i>be</i>	58	303	12	<i>bet</i> (31), <i>ber</i> (27)	Fel verb, Presens
<i>stjäl</i>	10	352	12	<i>*stjälade</i> (71), <i>*stjälde</i> (55)	*Preteritum

Tabell 4 åskådliggör att antalet unika felsvar stiger med antalet felaktiga svar totalt (fjärde respektive tredje kolumnen i tabellen). Detta är inte oväntat; man brukar ofta hitta mer variation ju större underlag (av felsvar i detta fall) man undersöker. Anmärkningsvärt är däremot att antalet belägg på det mest frekventa felsvaret inte sjunker med ökat antal unika felsvar. Till exempel står det vanligaste felsvaret till verbet *köpa* för 58 av de totalt 122 felaktiga svaren som är fördelade mellan fyra unika svar. Det vanligaste felsvaret för *fortsätta* står i sin tur för 73 belägg av totalt 311 felsvar fördelade mellan 14 unika svar. Med andra ord är felsvaren inte jämnt fördelade mellan de olika alternativen – varken för verb som behärskas relativt bra eller för verb som behärskas relativt dåligt. Detta tyder på att många av studiedeltagarna delar samma ”språkkänsla”, m.a.o. att deltagarna följer samma mönster för hur de bildar (både de korrekta och de felaktiga) preteritumformerna.

Tabell 4 åskådliggör också att det vanligaste felsvaret för många av verben utgörs av en form som slutar på *-de*, t.ex. *\*tänkade*, *\*väljade*, *\*sovde*, *\*nåde*. Med andra ord övergeneraliserar gymnasieeleverna mönstret att preteritum bildas med suffixet *-de*; sådana former utgör 33 procent av felsvaren totalt sett. Den största delgruppen bland denna typ av felsvar utgörs i sin tur av former där suffixet *-de* har fogats till verbets infinitivform, dvs. former som följer första konjugationens böjningsparadigm. Sådana former står för 17 procent av alla felsvar. Andra, tredje och fjärde konjugationens böjningsparadigm övergeneraliseras däremot inte. Många av felsvaren beror således på att eleverna tolkar verbets infinitivform som deras böjningsstam.

Förklaringen till detta kan vara antingen att eleverna inte är medvetna om att det finns en distinktion mellan infinitivform och böjningsstam eller att de inte vet vilka verb denna distinktion gäller. Eftersom verbböjningsuppgiften inkluderade endast 15 verb kan vi inte dra slutsatser om i hur hög grad respektive förklaring gäller för våra studiedeltagare. Att överanvända första konjugationens böjningsparadigm är dock en logisk strategi med tanke på att de flesta verben i svenskan tillhör första konjugationen. Söker man med Språkbankens Korp-sökverktyg (Borin, Forsberg & Roxendal 2012) på ord som har annoterats som 'Verb: Preteritum: Aktiv' i Språkbankens PAROLE-korpus ([spraakbanken.gu.se/resurser/parole](http://spraakbanken.gu.se/resurser/parole)), utgör formerna som slutar på *-ade* 61 procent av alla unika ordformer (eng. *types*) bland sökresultaten. Detta tal fungerar som en bra approximation av andelen verb som följer första konjugationens böjningsparadigm (trots att söksträngen utöver första konjugationens verb även fångar några få starka verb, t.ex. *lade*, *sade*). Följer man första konjugationens böjningsparadigm gör man alltså rätt i en majoritet av fallen.

En annan viktig iakttagelse är att många av de vanligaste felsvaren utgörs av former som är grammatiska ordformer men som representerar fel böjningskategori (dvs. inte preteritum). Tabell 4 åskådliggör t.ex. att det vanligaste felsvaret till verbet *gå* är presensformen *går* och att det vanligaste felsvaret till *studera* är supinumformen *studerat*. Överlag är andelen svar som representerar kategorierna presens respektive supinum ungefär lika hög (9 % respektive 10 % av alla svar i verbböjningsuppgiften). Denna typ av fel är vanlig i synnerhet för de verb som har relativt få felaktiga svar totalt (som är högre upp i tabell 3). Detta tyder på att studiedeltagarna är relativt medvetna om hur de korrekta ordformerna av de mer bekanta verben ser ut, men de är inte helt säkra på vilken av ordformerna som representerar kategorin preteritum. För vissa av verben är det vanligaste felsvaret i sin tur en korrekt preteritumform men av fel verb, t.ex. formen *såg* för verbet *säga* och formen *bet* för verbet *be*, vilket betyder att deltagarna i vissa fall är osäkra på vilken preteritumform som tillhör vilket verb. Detta gäller uteslutande verb som får vokalväxling i preteritum (6 % av svaren för dessa verb representerar denna feltyp).

## 6 Sammanfattning och diskussion

Denna studie utgör en pilotundersökning av verbböjning i inläraarsvenska genom multivariata analyser, med utgångspunkt i hur finskspråkiga gymnasieelever behärskar preteritumböjningen av 15 vardagliga svenska verb. Tidigare studier av verb i inläraarsvenska har byggt på förhållandevis små mängder data

(framförallt vad antalet informanter/studiedeltagare beträffar) och de har granskat effekten av enskilda variabler (såsom morfologisk form) utan att ta andra samvarierande variabler (såsom frekvens) i beaktande. Föreliggande studie utgör ett viktigt metodologiskt bidrag till forskningsfältet genom att den utnyttjar multivariata analyser för att skilja mellan effekten av de enskilda variablerna och för att belysa samspelet dem emellan.

Resultatet av vår analys visar att man relativt tillförlitligt kan predicera hur väl finska gymnasieelever behärskar de ifrågavarande preteritumformerna med hjälp av en regressionsmodell som inkluderar de utomspråkliga variablerna *ålder/exponeringstid*, *motivation* och *kön* och de inomspråkliga variablerna *frekvens* och *vokalväxling* (och som dessutom tar i beaktande den slumpmässiga effekten av urvalet av studiedeltagare och verb i studien). Modellen predicerar rätt utfall för 86 procent av observationerna i våra data. De effekter som de enskilda variablerna kan förutsägas ha enligt vår modell överensstämmer med de effekter som har beskrivits genom univariata analyser i tidigare forskning (se avsnitt 2): äldre elever (som läst svenska längre) presterar bättre än yngre, motiverade elever bättre än omotiverade och kvinnliga elever bättre än manliga. Vidare behärskas frekventa verbformer bättre än mindre frekventa, och verb vars stam inte växlar i preteritum behärskas bättre än verb som får vokalväxling. Våra resultat åskådliggör dessutom vilka fördelar multivariata analyser har gentemot univariata analyser. För det första hjälper multivariata analyser oss att identifiera de variabler som är relevanta för beskrivningen av ett fenomen: våra resultat tyder på att den traditionella distinktionen mellan starka och svaga verb inte är lika relevant för behärsksningen av svenskans preteritumböjning som distinktionen mellan verb som får vokalväxling i preteritum och verb som behåller samma stamvokal i presens och preteritum. För det andra hjälper multivariata analyser oss att jämföra hur stor effekten av de olika variablerna är. Våra resultat tyder på att *ålder/exponeringstid* har en klart större effekt än *motivation* och *kön* på hur väl preteritumböjningen behärskas.

Resultatet av vår felanalys tyder på att det finns allmänna mönster som finska gymnasieelever följer då de böjer svenska verb i preteritum. Många elever använder i synnerhet mönstret {INFINITIV + *-de*}. Med andra ord övergeneraliserar de suffixet *-de* till de flesta verben i svenskan med undantag av vissa högfrekventa verb, såsom *gå*. Samtidigt tolkar de infinitivformen som verbets böjningsstam, vilket ger upphov till former som t.ex. *\*tänkade*, *\*liggade* och *\*nåde*. Dessa resultat stämmer väl överens med Paavilainens (2015) analys av presensböjningen i finska skolelevers tal. Där är presensformer som slutar på *-ar*, dvs. former som kan tolkas som {INFINITIV + *-r*}, klart vanligast. Orsaken till att eleverna föredrar att bilda preteritumformer som slutar på *-ade* är alltså sannolikt formernas strukturella



genomskinlighet (i inlärares ögon) och likhet med infinitivformen. Dessutom är eleverna antagligen medvetna om att denna strategi fungerar för majoriteten (ca 60 %) av svenskans verb.

På en generell nivå problematiserar våra resultat de s.k. universella inlärningshierarkierna gällande verb som läggs fram i bl.a. Lightbown och Spada (2013) och Arifin m.fl. (2019). Dessa hierarkier utgår från en grov indelning av verben i kopulaverb, hjälpverb, regelbundna verb och oregelbundna verb. Vi argumenterar däremot för att inläringen av verbmorfologi är en betydligt mer mångfacetterad process som innebär ett samspel mellan åtskilliga inomspråkliga och utomspråkliga variabler. Inlärares tillägnar sig inte alla regelbundna former samtidigt och alla oregelbundna former är inte oregelbundna på samma sätt. Våra resultat tyder på att vokalväxling är en signifikant försvårande faktor vid inläringen av preteritumformerna, medan förekomsten eller avsaknaden av preteritumsuffix (dvs. distinktionen mellan svaga och starka verb) inte nödvändigtvis är det. För att kunna dra meningsfulla slutsatser om eventuella inlärningshierarkier bör ordens egenskaper följaktligen behandlas på en mycket mer detaljerad nivå än vad som gjorts inom andraspråksforskningen hittills. Observationerna i föreliggande studie har också några intressanta implikationer för undervisningen i svenska som andraspråk. Utifrån våra resultat behöver undervisningen om svenskans preteritumböjning inte nödvändigtvis fokusera på distinktionen mellan svaga och starka verb utan den kunde betona andra grammatiska faktorer istället, i synnerhet vokalväxlingen. Undervisningen kunde dessutom i större grad framhäva distinktionen mellan infinitiv och verbstam så att eleverna lär sig vilken form böjningssuffixet ska fogas till. Ett sätt att framhäva denna distinktion vore att öva mer på imperativformer i.o.m. att dessa är lika med verbens böjningsstam i svenskan.

De inomspråkliga och utomspråkliga variabler som vi har granskat i denna artikel utgör endast ett begränsat urval av faktorer som kan tänkas påverka inläringen av svenskans preteritumböjning. Studien bör utökas med en undersökning av ett mycket större antal verb från varje konjugation så att effekten av fler inomspråkliga variabler (t.ex. stavelselängd och *j*-bortfall) kunde granskas. En analys av ett större antal verb kunde dessutom ytterligare precisera effekten av de variabler som enligt våra resultat är signifikanta, t.ex. eventuella skillnader mellan olika mönster för vokalväxling (är mönstret *ä* → *a* exempelvis lättare att tillägna sig än *i* → *å*?). Samtidigt borde även fler inlärarrelaterade variabler tas i beaktande, såsom personlighet, kognitiv stil och övriga språkkunskaper. En bredare åldersfördelning kunde vara givande för att konkretisera vilka stadier inläringen av böjningsparadigm typiskt genomgår och hur länge dessa stadier varar. Vi hoppas att denna studie kan

fungera som en tematisk och metodologisk inspirationskälla för fortsatt forskning om verb i inläraarsvenska.

### Litteratur

- Agresti, Alan (2002). *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons.
- Arifin, Muhammad Ahkam, Jihad, Suryani, Mulyani, Sri, Ardin, Hardiani & Nurwahida, Nurwahida (2019). The acquisition order of past tenses: An interlanguage analysis. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, vol. 7, nr 2, s. 147–153.
- Baayen, R. Harald (2008). *Analyzing linguistic data*. Cambridge, UK: Cambridge University Press.
- Bates, Douglas, Maechler, Martin, Bolker, Ben & Walker, Steven (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, vol. 67, nr 1, s. 1–48. Doi: 10.18637/jss.v067.i01.
- Bergman, Pirkko (1988). Tid och tempus i nästan infödd svenska. I: Hyltenstam, Kenneth och Lindberg, Inger (red.). *Första symposiet om svenska som andraspråk. Volym 1. Föredrag om språk och interaktion*. Stockholm: Stockholms universitet. S. 157–164.
- Borin, Lars, Forsberg, Markus & Roxendal, Johan (2012). Korp – the corpus infrastructure of Språkbanken. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, s. 474–478.
- Boyle, Joseph P. (1987). Sex differences in listening vocabulary. *Language Learning*, vol. 37, nr 2, s. 273–284.
- Burstall, Clare (1975). Factors affecting foreign-language learning: A consideration of some recent research findings. *Language Teaching*, vol. 8, nr 1, s. 5–25.
- Csizér, Kata & Dörnyei, Zoltán (2005). Language learners' motivational profiles and their motivated learning behavior. *Language Learning*, vol. 55, nr 4, s. 613–659.
- Davies, Beatrice (2004). The gender gap in modern languages: A comparison of attitude and performance in year 7 and year 10. *Language Learning Journal*, vol. 29, nr 1, s. 53–58.
- DeKeyser, Robert M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, vol. 63, s. 52–67.
- Doughty, Catherine J. och Long, Michael (red.) (2003). *The Handbook of Second Language Acquisition*. Oxford: Blackwell.
- Ehrman, Madeline (1990). The role of personality type in adult language learning: An ongoing investigation. I: Parry, Thomas S. och Stansfield, Charles W. (red.). *Language aptitude reconsidered. Language in education: Theory and practice*. Washington, D.C.: ERIC Clearinghouse on Languages and Linguistics. S. 126–178.
- Ellis, Nick C., & Ferreira-Junior, Fernando (2009). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, vol. 93, nr 3, s. 370–385.
- Ellis, Rod (1989). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Genelot, Sophie (1997). L'enseignement des langues vivantes à l'école élémentaire: Éléments d'évaluation des effets au collège. *Revue Française De Pédagogie*, vol. 118, s. 27–42.

- Gries, Stefan Th. (2021). *Statistics for Linguistics with R. 3rd revised and extended edition*. Boston & Berlin: De Gruyter.
- Hyltenstam, Kenneth, & Abrahamsson, Niclas (2003). Maturational constraints in SLA. I: Doughty, Catherine J. och Long, Michael H. (red.). *The handbook of second language acquisition*. Malden, MA: Blackwell Publishing. S. 538–588.
- Leiner, Dominik J. (2019) SoSci Survey (Version 3.1.06) [computer software]. <https://www.socisurvey.de>.
- Lightbown, Patsy M. och Spada, Nina (2013). *How languages are learned*. 4th edition. Oxford: Oxford University Press.
- Marton, Enikő & Kovács, Magdolna (2019). What motivates Finnish-speaking high-school students in learning Swedish? *Nordand*, vol. 3, nr 2, s. 101–116.
- Moody, Raymond (1988). Personality preferences and foreign language learning. *The Modern Language Journal*, vol. 72, nr 4, s. 389–401.
- Muittari, Veikko (2008). Finita och infinita verbformer i finskspråkiga gymnasisters inlärnarsvenska. I: Lindström, Jan, Kukkonen, Pirjo, Lindholm, Camilla och Mickwitz, Åsa (red.). *Svenskan i Finland 10*. Helsingfors: Helsingfors universitet. S. 212–224.
- Nikolov, Marianne (2009). Early modern foreign language programmes and outcomes: Factors contributing to Hungarian learners' proficiency. I: Nikolov, Marianne (red.). *Early Learning of Modern Foreign Languages: Processes and Outcomes*. Bristol: Multilingual matters. S. 90–107.
- Paavilainen, Marika (2015). *Inlärnning och behärskning av svenskans verb- och adjektivböjning samt negationens placering hos finska grundskoleelever*. Doktorsavhandling. Jyväskylä Studies in Humanities 260. Jyväskylä: University of Jyväskylä.
- Philipsson, Anders. (2007). *Interrogative clauses and verb morphology in L2 Swedish: Theoretical interpretations of grammatical development and effects of different elicitation techniques*. Ph.D. Dissertation. Stockholm: University of Stockholm.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Rahkonen, Matti & Håkansson, Gisela (2008). Production of written L2-Swedish-processability or input frequencies. I: Kessler, Jörg-U. (red.). *Processability approaches to second language development and second language learning*. Cambridge: Cambridge Scholars Publishing. S. 135–161).
- SAG = Teleman, Ulf, Hellberg, Staffan och Andersson, Erik (1999). *Svenska Akademiens grammatik*. Stockholm: Svenska Akademien/Norstedts.
- SAOL 14 = *Svenska Akademiens ordlista över svenska språket* (2015). Stockholm: Norstedts.
- Saville-Troike, Muriel & Barto, Karen (2017). *Introducing second language acquisition*. Cambridge: Cambridge University Press.
- Sonderegger, Morgan, Wagner, Michael & Torreira, Francisco (2018). *Quantitative methods for linguistic data. Version 1.0*. <http://people.linguistics.mcgill.ca/~morgan/qmld-book/>.
- Speelman, Dirk (2014). Logistic regression: A confirmatory technique for comparisons in corpus linguistics. I: Glynn, Dylan och Robinson, Justyna A. (red.). *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins. S. 487–533.
- Speelman, Dirk, Heylen, Kris & Geeraerts, Dirk (red.) (2018). *Mixed-effects regression models in linguistics*. Cham: Springer.

- Sundman, Marketta (2011). Helsekvenser i finska grundskoleelevers skrivande. I: Edlund, Ann-Catrine och Mellenius, Ingmarie (red.). *Svenskans Beskrivning 31*. Umeå: Umeå universitet. S. 327–336.
- Steinhauer, Hein (2001). *Leerboek Indonesisch*. Leiden: KITLV Uitgeverij.
- Van Der Slik, Frans W. P., Van Hout, Roeland W. & Schepens, Job J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PloS ONE*, vol. 10, nr 11. Doi: 10.1371/journal.pone.0142056.
- Wickham, Haley (2016). *Ggplot2: Elegant graphics for data analysis* [computer software]. New York: Springer-Verlag.
- Williamson, John M., Lin, H., Lyles, Robert H. & Hightower, Allen W. (2007). Power calculations for ZIP and ZINB models. *Journal of Data Science*, vol. 5, nr 4, s. 519–534.
- Winter, Bodo (2020). *Statistics for linguists: An introduction using R*. New York & London: Routledge.
- Wulff, Stefanie, Ellis, Nick C., Römer, Ute, Bardovi-Harlig, Kathleen & Leblanc, Chelsea J. (2009). The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal*, vol. 93, nr 3, s. 354–369.