



En splyv eller *et splyv*? Tilordning av grammatisk genus til pseudosubstantiv i norsk

Olga Urek¹, Terje Lohndal^{2,1} og Marit Westergaard^{1,2}

¹UiT Norges arktiske universitet

²NTNU Norges teknisk-naturvitenskapelige universitet

Tradisjonelt har det norske genussystemet blitt karakterisert som lite transparent. Et viktig spørsmål er om språkbrukere likevel kan være sensitive til visse egenskaper ved substantiver og bruker disse produktivt når de tildeler genus til ukjente ord. I denne artikkelen undersøker vi eksperimentelt språkbrukeres sensitivitet til fonologiske egenskaper som vi har identifisert gjennom korpusundersøkelser. Denne artikkelen formidler resultater fra to studier som begge fokuserer på forholdet mellom hankjønn som standardverdi (*default*) for genus i norsk og sensitivitet til et substantivs endelse når det gjelder genustilordning. Generelt finner vi at hankjønn står i en særstilling og har en klar posisjon som standardverdi, men samtidig ser vi en effekt av fonologiske egenskaper. I studie 1 avdekker vi dessuten at eksperimentell metode kan ha betydning for resultatene, noe vi derfor kontrollerer for i studie 2.

Nøkkelord: fonologi, eksperimentell prosedyre, elisiteringseksperiment, genus, genustilordning, korpusstudie, pseudosubstantiv, semantikk, standardverdi, transparens

1. Introduksjon¹

I norske dialekter som opprettholder et tregenussystem, er det først og fremst semantiske egenskaper som kan brukes til å forutsi hvilket grammatisk genus et substantiv har. For animater er det biologisk kjønn som bestemmer det grammatiske genuset: *jente* og *ku* er hunkjønn, *gutt* og *okse* er hankjønn. Utover dette inneholder nakne substantiv lite eller ingen informasjon som kan brukes til å predikere genus. Det vil si at det ikke er noe ved ordene *spill* og *drill* som kan brukes til å forutsi at *spill* er intetkjønn og *drill* er hankjønn.

Tidligere forskning har likevel kommet til ulike syn på om det er mulig å forutsi genustilordning i norsk, der analysene går fra at genustilordning er helt arbitrært til at det er mulig å formulere en rekke spesifikke regler for tilordning. Denne forskningen er i hovedsak basert på analyser av eksisterende norske ord. Det er et eksempel på det Comrie (1999: 461) kaller «postfactum rationalizations», siden det tar utgangspunkt i det eksisterende og forsøker å lete etter mulige generaliseringer. Et slikt perspektiv sier derfor svært lite om den psykologiske statusen til slike regler (jf. Gagliardi 2012: 110). For å få innsikt i dette er det nødvendig å «uncover psychologically real and productive criteria that speakers exploit in “on-the-spot” gender assignment» (Thornton 2009: 17, jf. Corbett 1991 og Audring 2016). Dette kan gjøres gjennom å se på genustildeling til lånord og det som kalles pseudosubstantiv, altså ord som kunne ha vært norske ord i henhold til norsk fonotaks, men som ikke er det. Da ser vi hvilke produktive regler språkbrukerne eventuelt har etablert. Gjennom eksperimenter basert på pseudosubstantiv ønsker vi i denne artikkelen å undersøke om det finnes produktive regler for genustilordning i norsk.

Artikkelen begynner med en gjennomgang av relevant bakgrunn for studien vår i avsnitt 2. Her ser vi på forskningshistorikken. Denne fører oss fram til forskningsspørsmålene som utgjør grunnlaget for denne artikkelen. Avsnitt 3 beskriver hvordan vi har identifisert prediktive egenskaper for genustilordning. Deretter beskriver avsnitt 4 den første studien vår, mens avsnitt 5 beskriver den andre. Avsnitt 6 gir en overordnet diskusjon av funnene våre og hva dette forteller oss om tilordning av grammatisk kjønn mer generelt i norsk. Konklusjonen kommer til slutt i avsnitt 7.

1. Vi er takknemlige til to anonyme konsulenter og Yvonne van Baal for nyttige kommentarer. Denne forskningen er en del av MultiGender-prosjektet ved Senter for grunnforskning (CAS) 2019-2020 samt NFR-prosjektene MiMS (prosjektnummer 250857) og GenVAC (prosjektnummer 301094).

2. Bakgrunn

Det norske genussystemet karakteriseres tradisjonelt som lite transparent (Næs 1952, Faarlund et al. 1997, Rønhovd 1997, Venås 1997, Conzett 2006). Næs (1952: 181) sier for eksempel følgende: «[...] et ords genus er et vilkårlig og isolert historisk faktum». Den samme konklusjonen kommer *Norsk referansegrammatikk* til: «I de aller fleste tilfellene kan vi ikke se av formen på substantivet hvilket genus det har» (Faarlund, Lie & Vannebo 1997: 150). Likevel inneholder referansegrammatikken flere sider med genustilordningstendenser, slik at det framstår som mulig å stille opp noen regler. Et omfattende arbeid som gjør nettopp dette, er Beito (1986), som gir en rekke regler for tildeling av genus (se også Hjelde 1996 om amerikanorsk og Husby 1990 samt Leirvaag 1999 innenfor fagfeltet norsk som fremmedspråk). På tross av dette har det rådende synet vært at genustildeling i norsk er arbitrært.²

Trosterud (2001) følger i Beitos fotspor, men han er mer preget av internasjonale perspektiver på genusforskning. Et viktig poeng for Trosterud er at hankjønn er standardverdien (*default*), og dette gjør at det ikke er nødvendig å formulere regler for hankjønn (men Trosterud gjør det likevel i sine beskrivelser av unntak til reglene). Videre ser Trosterud de semantiske reglene i sammenheng med hverandre, samtidig som han skiller tydelig mellom semantiske, morfologiske og fonologiske regler. Basert på 31500 substantiver fra *Nynorskordboka*, foreslår han tre overgripende regler, som vist i (1) (Trosterud 2001: 35).

- (1) a. **Regel 1 Default: Alle norske ord er m**
bil, båt, ironi, SAS, web
- b. **Regel 2 Tostava ord på trykklett -e (svake substantiv) er f**
fille, hytte, kake, snekke
- c. **Regel 3 Einstava ord på vokal er f**
bru, klo, krå, tå

Trosterud gir også fire generelle regler som sammen med reglene i (1) utgjør grunnlaget for de mer detaljerte reglene. Totalt gir han 43 ulike regler som predikerer 94% av substantivene i korpuset han bruker. Mange av disse har en rekke unntak, noe som viser vanskelighetene med å formulere binære regler

2. Vi skiller her mellom genustildeling og genuskongruens. For eksempel er standardverdien til genustildeling hankjønn (mer om dette nedenfor), mens standardverdien til genuskongruens er nøytrum. Se Enger (2001: 178) og Lohndal & Westergaard (2021) for mer om dette.

som enten gjelder eller ikke gjelder. Noen av disse reglene er også svært spesifikke. For eksempel sier regel 13 at substantiver som refererer til bygninger som ikke er permanente husvære for mennesker, er intetkjønn. Det er ikke klart hvordan et barn tilegner seg regler som er så spesifikke på grunnlag av input alene. Samtidig er det viktig å undersøke om reglene også har en psykologisk realitet. En naturlig måte å gjøre dette på er gjennom eksperimenter.

Halse (2004) tester Trosteruds regler ved hjelp av datamaskinelle læringsmetoder, mer spesifikt maskinlæring. En slik test sier ikke nødvendigvis noe om psykologisk realitet, men den gir en kvantitativ etterprøving samt mulig optimalisering av reglene. Halse ser på tre ulike maskinlæringsalgoritmer der to av disse genererer regler basert på datagrunnlaget, mens den tredje er en minnebasert modell som ikke genererer regler. Datamaterialet skulle ha vært alle Trosteruds 31500 substantiver, men Halse fikk bare tilgang til 13384. Disse hadde fordelingen 22,2% hunkjønn, 58% hankjønn og 19,8% intetkjønn. Hun gjorde også noen videre justeringer i utvalget som vi ikke går inn på her (se Halse 2004: 23-24). For å finne ut hvor gode de ulike modellene er, gjennomførte hun en rekke eksperimenter der hun så på feilraten ved de ulike eksperimentene for alle tre modellene. Eksperimentene så bort fra enkelte egenskaper, f.eks. semantikk eller bøyingsmorfologi, for å finne ut hvilken betydning de ignorerte egenskapene hadde for de forskjellige modellenes evne til å predikere korrekt genus.

Resultatene til Halse (2004) viser at den minnebaserte modellen greier seg best, noe som ikke er overraskende siden denne modellen ikke abstraherer fra datasettet. Vi går ikke inn på forskjeller og likheter mellom de to regelbaserte modellene her, men fokuserer på de overordnede resultatene. De viser at morfologi er den egenskapen som i størst grad bidrar til riktig klassifikasjon. Interessant nok viser modelleringen at avledningsmorfologi, suffiks og generelle fonologiske endelser er viktige faktorer ved genustildeling, siden disse sammen kan danne et regelsett (Halse 2004: 56). Derimot er semantikk mye mindre viktig enn Trosteruds regler skulle tilsi. Begge de regelbaserte modellene genererer «mange færre regler enn Trosterud inkluderer, noko som gjer regelsettet meir oversiktleg og lettare å lære» (Halse 2004: 56).³ Samtidig

3. Halse (2004: 56) peker også på et viktig forbehold med læringsmodeller av den typen hun diskuterer:

Dei regelbaserte læringsmetodane genererer ein del reglar som er kompliserte og lite intuitive, og difor lite praktisk nyttige. Det viser seg også at ulike algoritmar genererer ulike reglar ut i frå det same datasettet. Slike metodar kan difor ikkje produsere ferdige regelsett for genustilordning, men gi ein peikepinn på moglege reglar.

fant ikke modellen andre mulige regler enn de som er formulert av Trosterud. Dette støtter også opp om Trosteruds generelle hypotese, nemlig at det er mulig å formulere regler for genustildeling i norsk.

Gagliardi (2012) bruker pseudosubstantiv for å teste egenskapene som har vært hevdet å best forutsi genus: For hankjønn er dette referanse til biologisk hankjønn, for hunkjønn er dette endelsen *-e* og referanse til biologisk hunkjønn.⁴ Hun finner at norske barn i alderen 4;2-7;2 ikke gjør særlig bruk av denne informasjonen når de skal tildele genus til pseudosubstantiv. I stedet har de en sterk preferanse for å klassifisere alle ordene som hankjønn (Gagliardi 2012: 121). Imidlertid ser hun også at «there is a very slight preference to classify nouns as Feminine with either semantic or phonological cues for Feminine» (Gagliardi 2012: 121). Denne tendensen gjelder uavhengig av hva substantivenes referanse er, det vil si at morfofonologisk informasjon overstyrer semantisk referanse.

Et annet arbeid som også bekrefter at språkbrukere til en viss grad er sensitive til enkelte egenskaper ved substantivene og bruker disse når de tildeler genus til ukjente ord, er Bobrova (2013). Hun viser at voksne språkbrukere ganske systematisk skiller mellom intetkjønn og fellekjønn basert på animathet. En rekke semantiske kategorier ble også undersøkt, der flere av pseudosubstantivene fikk tilskrevet en generell betydning (eks.: *spønn* 'et slags dyr'). Samlet sett viser hun at semantikken har en klar påvirkning når det gjelder genustilordning, men hun undersøker ikke om andre egenskaper også kan ha betydning.

Til slutt er det verd å ta med at det finnes forskning på tildeling av genus til engelske lånord i norsk. Graedler (1998) og Johansson & Graedler (2002) bruker korpusdata for å undersøke hvordan genus tildeles til engelske substantiv i ytringer som ellers er norske. Resultatene deres viser at 80-90% av substantivene er hankjønn, 10-20% er intetkjønn, og svært få er hunkjønn (Johansson & Graedler 2002: 183). Veldig mange lånord veksler mellom hankjønn og intetkjønn. Samtidig er det tendenser også i dette materialet: engelske substantiver som ender på *-ing* eller *-er* blir typisk tildelt hankjønn, mens enstavsessubstantiver gjerne blir tildelt intetkjønn.

Samlet sett viser tidligere forskning at det er mulig å formulere stokastiske regler for tildeling av genus i norsk, det vil si regler som har en viss sannsyn-

4. For å finne fram til disse, oversatte Gagliardi 833 engelske substantiver fra MacArthur-Bates CDI (Fenson et al. 1993). Deretter brukte hun et beslutningstre for å finne ut hvilke egenskaper som var mest prediktive for kjønn. Se avsnitt 3 for mer om beslutningstre som metode.

lighet for å gjelde. Imidlertid er det vanskelig å se at det finnes noen kategoriske regler, siden alle regler som har blitt formulert, har unntak. Mye av forskningen som har vært gjort så langt, har også vært basert på nynorsk, og det er ikke gitt at bildet er det samme for bokmål. Vi vet heller ikke i hvilken grad språkbrukere har internalisert kunnskap om korrelasjoner mellom formen på substantivet og det genuset som substantivet tilordnes. Dersom de har slik kunnskap, blir det neste spørsmålet hvordan denne er representert og lagret i vår interne grammatikk. I denne artikkelen vil vi belyse disse spørsmålene gjennom to eksperimentelle undersøkelser. På forhånd bruker vi maskinlæring for å identifisere de mest sannsynlige prediktive egenskapene for genustildeling i bokmål, og deretter bruker vi disse til å lage pseudosubstantiver som vi så tester ved hjelp av ulike metoder.

Denne artikkelen tar utgangspunkt i følgende forskningsspørsmål:

1. Finnes det pålitelige fonologiske egenskaper for genustildeling i norsk?
Kan grammatisk kjønn predikeres basert på den fonologiske formen til substantivet?
2. Gjør talere av norsk bruk av probabilistiske fonologiske egenskaper når de tildeler grammatisk kjønn til pseudosubstantiver?
3. Spiller den eksperimentelle prosedyren en rolle?
4. Spiller tilstedeværelsen eller fravær av visuelle stimuli en rolle?

3. Prediktive egenskaper for genustildeling: En korpusstudie

For å identifisere hvilke egenskaper som predikerer genustildeling i norsk, ekstraherte vi ca. 13000 substantivlemma med genus samt frekvensinformasjon fra korpuset Universal Dependencies Treebank Bokmål (Øvrelid & Hohle 2016).⁵ Vi markerte hvert lemma for ordfinalt grafem og antall stavelser. Listen ble analysert med et såkalt beslutningstre for å identifisere de prediktive egenskapene. Et beslutningstre er en organisering av tester. Hver node i et beslutningstre spesifiserer en test for en egenskap (attributt), og hver gren som kommer fra denne noden, korresponderer med en mulig verdi for egenskapen. Halse (2004) viser hvordan det er mulig å lage et beslutningstre basert på reg-

5. En fagfelle stiller spørsmål ved hvorfor vi ser på substantivlemma når disse er mindre frekvente i input enn bestemte former samt flertallsformer. Vi valgte lemmaformer fordi vi antar at det er nakne former som er lagret i en språkbrukers mentale leksikon. Da er det de nakne formene som danner grunnlaget for å generere sannsynligheten for at gitte endelser kobles sammen med et bestemt genus.

lene til Trosterud (2001). Det er viktig å presisere at det ikke er et en-til-en-forhold mellom attributtverdier og genustilordningsregler, siden attributtverdiene er formulert slik at de viser til en eller flere tilordningsregler (jf. Halse 2004: 26). I vårt arbeid brukte vi et J48 beslutningstre (Weka software, Witten & Frank 2005).

Resultatet av dette beslutningstreet var at final *-v* og *-e* er de mest prediktive egenskapene for henholdsvis intetkjønn og hunkjønn. Substantiver som ender på *-g* er ikke med her, selv om 43% av disse er hunkjønn. Årsaken til denne eksklusjonen er at i flertallet av disse substantivene som slutter på *-g*, er *-g* en del av det produktive derivasjonelle morfemet *-ing*. Denne derivasjonelle endelsen er en tydelig *morfologisk* egenskap for hunkjønn. Prediktive egenskaper er ikke nødvendigvis i samsvar med frekvens. Om vi ser på det samme korpuset som beslutningstreet ble brukt på, er 23% av substantivene som ender på *-e* hunkjønn (av totalt 2386 lemma med denne endelsen), 63% er hankjønn og 14% intetkjønn. Her overlapper altså ikke frekvens og prediktivitet. Ser vi på final *-v*, er 60% av totalt 208 *-v*-finale lemma intetkjønn, 38% er hankjønn og 2% er hunkjønn. Frekvens og prediktivitet overlapper altså i dette tilfellet.

Korpusstudien viser altså at det er mulig å identifisere prediktive egenskaper for genustildeling. I de neste to avsnittene undersøker vi om disse egenskapene har psykologisk realitet hos språkbrukerne.

4. Studie 1

4.1. Språklige og visuelle stimuli

I studie 1 benyttet vi 30 pseudosubstantiver som var konstruert slik at de fulgte norske fonotaktiske regler. Disse ikke-eksisterende ordene fordelte seg på følgende tre kategorier: i) ti substantiver med to stavelser som ender på *-e* (f.eks. *kvumme*, *tryspe*, *dilfe*); ii) ti substantiver med én stavelse som ender på *-v*, enten alene eller i en konsonantkombinasjon (f.eks. *klirv*, *trulv*, *glyv*); og iii) ti substantiver med én stavelse som ender på en annen konsonant enn *-v*, igjen enten alene eller i en kombinasjon av flere konsonanter (f.eks. *prylk*, *sprygg*, *smulp*). Vi forsikret oss om at disse ordene ikke hadde høyfrekvente fonologiske naboer (ord som skiller seg fra dem kun ved én lyd) ved å foreta et søk ved hjelp av bokmål-versjonen av *Norwegian Computational Lexicon* (Nordgård 1998) og et Python-skript utviklet av Tekstlaben ved Universitetet i Oslo. I tillegg valgte vi ut 30 bilder av såkalte non-objekter fra databasen NOUN (Horst & Hout 2016). Alle bildene viste inanimate tellelige objekter, slik som i (2).

(2) Eksempel på et non-objekt



4.2. *Deltakere og prosedyre*

Studie 1 hadde 26 voksne deltakere med norsk som morsmål. Alle deltakerne snakket en nordnorsk dialekt. De var studenter ved UiT Norges arktiske universitet, og eksperimentet foregikk på campus. Deltakerne ble tilfeldig valgt ut til å delta i en av to forskjellige prosedyrer for eksperimentet (N=13 deltakere i hver).

Prosedyre 1: Elisitering med bilder

Prosedyre 1 var en justert versjon av en eksperimentell teknikk som er ofte brukt i genusstudier, f.eks. Rodina & Westergaard (2015, 2017), Westergaard & Rodina (2016), Mitrofanova, Rodina, Urek & Westergaard (2018) og Rodina, Kupisch, Meir, Mitrofanova, Urek & Westergaard (2020). Deltakerne fikk se to bilder av non-objekter ved siden av hverandre på en dataskjerm. Objektene var identisk i form, men hadde forskjellig farge. Det tilhørende pseudosubstantivet ble introdusert i en setning som var formulert slik at substantivets genus ikke var gitt (se dialogen nedenfor). Deltakerne skulle så svare på et spørsmål om hva de så på skjermen og ble øvd opp i et mønster der de skulle produsere substantivet sammen med en ubestemt artikkel og et adjektiv, som begge kongruerer med substantivet i genus. Etter dette forsvant et av bildene på skjermen og deltakerne skulle svare på hvilket av objektene som forsvant. Svaret på dette spørsmålet var en substantivfrase med dobbel bestemthet – altså en prenominal determinativ og bestemthetsuffikset. I (3) ser vi en forventet dialog mellom forsker og deltaker for denne prosedyren.

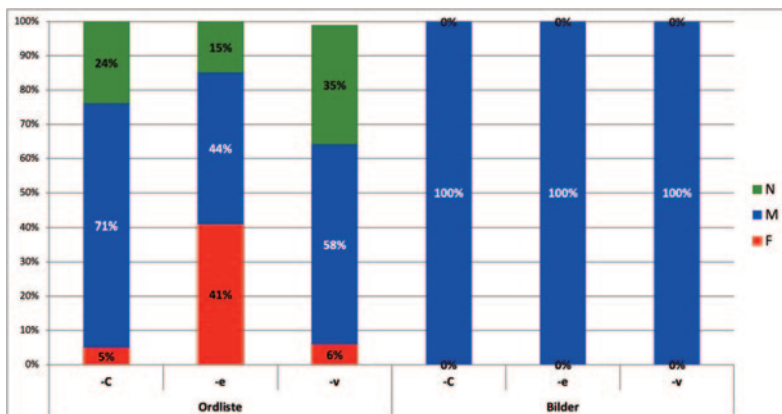
- (3) Forventet dialog for Prosedyre 1
- FORSKER: *Dette kalle vi for **splyv**. Korsen farge e dem?*
- DELTAKER: ***Et blå-tt splyv** og **et gul-t splyv**.*
- FORSKER: (trykker på en knapp som får et av objektene til å forsvinne)
 Ka som forsvant?
- DELTAKER: ***Det blå(e) splyv-et***

Prosedyre 2: Ordliste uten bilder

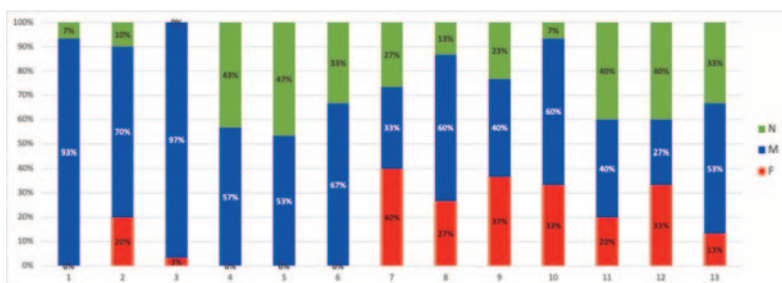
I prosedyre 2 fikk deltakerne en liste med de 30 pseudosubstantivene i tilfeldig rekkefølge (alle deltakerne fikk samme rekkefølge), med beskjed om å lese dem høyt sammen med en av de tre ubestemte artiklene *en*, *ei* eller *et*. Deltakerne ble fortalt at det ikke fantes noe rett eller galt svar, men at de skulle bruke den artikkelen som “hørtes mest naturlig ut for dem”.

4.3. Resultater og analyse

Resultatene i studie 1 viste en drastisk forskjell mellom de to prosedyrene. Alle deltakerne i Prosedyre 1 produserte uten unntak hankjønn for alle pseudosubstantivene, uten hensyn til hvilken ending substantivet hadde, altså *en tryspe*, *en splyv* og *en smulp*. I motsetning til dette så vi at de deltakerne som gjennomførte prosedyre 2, produserte alle tre genus på en måte som syntes å være i samsvar med endelsen på pseudosubstantivet – i alle fall til en viss grad: 71% av alle pseudosubstantiv som endte på en konsonant (bortsett fra *-v*) ble tilordnet hankjønn, 41% av alle pseudosubstantiv som endte på *-e* ble tilordnet hunkjønn, og 35% av alle pseudosubstantiv som endte på *-v* ble tilordnet intetkjønn. Dette er illustrert i Figur 1. Det er verd å merke seg at alle tre genus brukes av flertallet av deltakerne; det er fem deltakere som ikke produserer alle de tre ubestemte artiklene i entall. Av disse er det to som bruker hankjønn hele tiden (henholdsvis 28 og 29 tilfeller av 30 mulige). De gjenværende tre bruker ikke hunkjønn, men de produserer både hankjønn og intetkjønn. Den individuelle distribusjonen er illustrert i Figur 2. Forholdet mellom endelsen og sannsynligheten for å tilordne det tilsvarende genus ble analysert ved hjelp av en generalisert (binomial) blandet modell (lme4 R package, Bates, Mächler, Bolker & Walker 2014). Denne modellen inkluderte sannsynligheten for en respons som er i samsvar med en gitt egenskap (en binær variabel som indikerer om responsen svarer til genusegenskapen, for eksempel *ei* for *-e*, *et* for *-v*, og *en* for *-C*) som en avhengig variabel og substantivendelsen som prediktor og “deltaker” og “pseudosubstantiv” som tilfeldige effekter. Post hoc parvise sammenligninger



Figur 1: Genustilordning for pseudosubstantiv med endelsene -C, -e og -v for henholdsvis hankjønn, hunkjønn og intetkjønn ifølge to eksperimentelle prosedyrer.



Figur 2: Distribusjonen til hunkjønn (F), hankjønn (M) og intetkjønn (N) hos individuelle deltakere i leseprosedyren.

(Ismeans, Lenth 2016) viste at det var signifikant større sannsynlighet for at deltakerne produserte hankjønn for pseudosubstantiver som endte på konsonant (-C) enn det var for henholdsvis intetkjønn for -v og hunkjønn for -e (SE = 0,288, z ratio = 5,446, $p < 0,0001$ og SE = 0,283, z ratio = 4,691, $p < 0,0001$). Det var imidlertid ingen statistisk forskjell mellom sannsynligheten for å tilordne intetkjønn til pseudosubstantiv med endelsen -v og hunkjønn til endelsen -e ($p = 0,648$).

Separate sett av blandede binomiale modeller som inkluderte sannsynligheten for et gitt genus som en avhengig variabel og «vilkår» som prediktor,

avdekket at det var signifikant større sannsynlighet for at et spesifikt genus ble tilordnet pseudosubstantiv som hadde den tilsvarende endelsen enn at dette genuset skulle tilordnes et pseudosubstantiv med en annen endelse. Dette betyr at andelen hankjønn var høyere ved endelsen *-e* enn ved både *-C* (SE = 0,577, *z* ratio = 5,839, $p < 0,0001$) og *-v* (SE = 0,563, *z* ratio = 5,711, $p < 0,0001$). Dette gjaldt også andelen intetkjønn ved endelsen *-v* i forhold til *-e* (SE = 0,352, *z* ratio = -3,58, $p = 0,001$), mens forskjellen mellom tilordning av intetkjønn til pseudosubstantiv med endelsen *-v* og andre konsonanter (*-C*) ikke var statistisk signifikant ($p = 0,1$). Når det gjaldt andelen hankjønn, var den eneste signifikante forskjellen mellom endelsene *-C* og *-e* (SE = 0,33, *z* ratio = 4,198, $p = 0,0001$).

4.4. Diskusjon og nye forskningsspørsmål

Resultatene av studie 1 støtter klart en analyse der hankjønn er standardverdien for grammatisk genus i norsk (f.eks. Faarlund, Lie & Vannebo 1997, Trosterud 2001): Ikke bare brukes hankjønn helt konsekvent ved prosedyre 1, vi ser også at hankjønn faktisk er foretrukket ved alle tre substantivendelsene i prosedyre 2 (71% ved *-C*, 44% ved *-e* og 58% ved *-v*), til tross for at vi ved de to sistnevnte endelsene også ser en effekt av den fonologiske formen på substantivet. I tillegg viser resultatene at fonologisk form faktisk spiller en viss rolle for genustilordning i norsk, i motsetning til tidligere forskning som primært argumenterer for betydningen av semantikk og delvis morfologi (jf. avsnitt 2). Dette er tilfelle også der semantiske trekk ikke er en del av bildet (deltakerne i Prosedyre 2 fikk ikke opplyst noen betydning for pseudosubstantivene) og man har utelukket påvirkning fra mulige fonologiske naboer. Dette betyr at selv om de fleste substantiver på *-e* i norsk er hankjønn (f.eks. *en pinne*, *en bolle*) og også hovedsakelig tilordnes hankjønn av våre norsktalende deltakere (44%), så vil disse substantivene med større sannsynlighet tilordnes hankjønn enn substantiv med en annen fonologisk form (41% vs. 5% og 6% for endelsene *-C* og *-v*). Noen av resultatene fra studie 1 er ikke statistisk signifikante, noe som kan skyldes at vi har et begrenset antall deltakere (N=13).

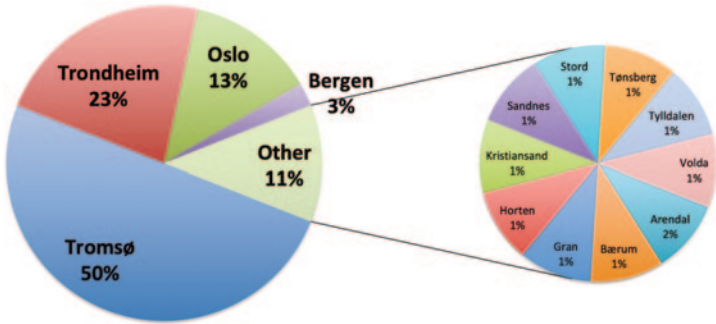
Det mest overraskende resultatet av denne studien er den store diskrepansen mellom prosedyre 1 og 2. Dette leder naturligvis til noen nye spørsmål som vi skal utforske i studie 2: Hvorfor brukte deltakerne i prosedyre 1 kun hankjønn, mens de som så ordlisten, tok hensyn til formen på substantivet? Er det sånn at semantikken spiller inn her, slik at inanimate tellelige objekter vanligvis er hankjønn i norsk? Trosterud (2001) antyder at dette kan være tilfelle, men det finnes ingen tidligere forskning som har undersøkt om hankjønn faktisk er et mer sannsynlig genus for de non-objektene som ble avbildet i prosedyre 1. Det er også

mulig at forklaringen ligger i måten den lesebaserte prosedyren ble utført på: Ved at deltakerne her fikk et valg mellom alle tre artiklene (*en, ei, et*), ble de eksplisitt gjort oppmerksom på at studien dreide seg om genus, og de ble sannsynligvis dermed mer fokusert på substantivenes form. Lesing kan også innebære at skriftspråket og kunnskap om dette blir aktivert, selv om dette i seg selv ikke skulle tilsi de mønstrene vi ser. I tillegg var det slik at deltakerne fikk se hele ordlisten (30 ord) på én gang. Dette kan ha ført til at de følte en viss forventning om at de skulle bruke alle artiklene, og kanskje til og med at de skulle bruke hver artikkel omtrent like ofte. Likevel var det ikke slik at de fordelte artiklene tilfeldig, i og med at det er helt klart at den fonologiske formen har hatt en effekt på deltakernes valg. Vi må dermed anta at norsktalende – på et eller annet nivå – har en viss internalisert kunnskap om at det er korrelasjoner mellom formen på et substantiv og det genus substantivet tilordnes. Men hvordan er denne kunnskapen representert og lagret i vår interne grammatikk? Og hvorfor kommer den til uttrykk bare i noen tilfeller (i prosedyre 2, men ikke i prosedyre 1)? I studie 2 utforsker vi disse spørsmålene videre.

5. Studie 2

5.1. Deltakere

I studie 2 brukte vi de samme pseudosubstantivene som i studie 1, men vi benyttet en ny eksperimentell metode og nye deltakere. Vi rekrutterte 66 morsmålsbrukere av norsk (av dem 45 kvinner) til denne studien, hovedsakelig gjennom sosiale medier. Deltakerne fylte ut et spørreskjema på forhånd, som ga oss informasjon om alder og geografisk tilhørighet (nåværende bosted). Aldersmessig var det stor spredning på deltakerne, fra 15 til 75 år (gjennomsnitt = 36, SD = 17,8). På grunn av at vi brukte en digital eksperimentell metode (se nedenfor), var det også betraktelig større geografisk spredning enn i studie 1. Bare halvparten av deltakerne bodde i Tromsø, mens Trondheim, Oslo og Bergen var representert med henholdsvis 15%, 13% og 3%, og de resterende 11% fordelte seg på 10 forskjellige steder. Denne distribusjonen er illustrert i Figur 3.




Figur 3: Deltakernes geografiske tilhørighet

5.2. Eksperimentell metode

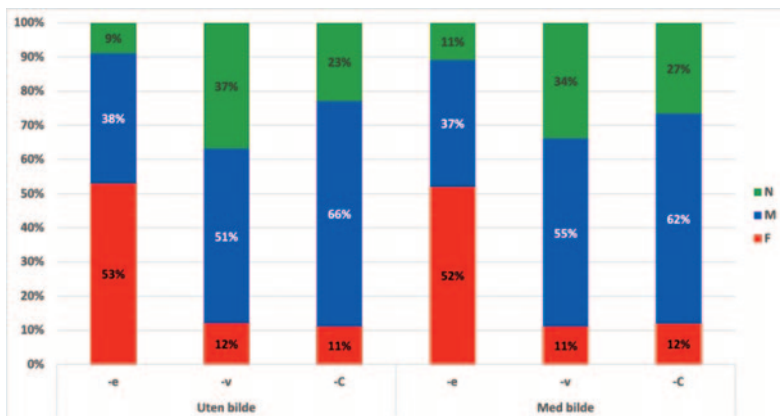
For studie 2 valgte vi å bruke IbexFarm, en digital plattform der deltakerne mottar en lenke til eksperimentet (Drummond 2012). Etter at de hadde gitt samtykke til deltakelsen og fylt ut bakgrunnsskjemaet, ble deltakerne helt tilfeldig valgt ut til en av to forskjellige eksperimentelle prosedyrer, en med bilder og en uten. I begge prosedyrene så deltakerne de aktuelle pseudosubstantivene – ett av gangen – plassert midt på dataskjermen, sammen med de tre ubestemte artiklene *en*, *ei* og *et*. Rekkefølgen ordene ble presentert i, var randomisert for hver deltaker. Deltakerne fikk beskjed om å velge den ubestemte artikkelen de syntes passet best til hvert ord, og det ble samtidig presisert at det ikke fantes noe rett eller galt svar. Forskjellen mellom de to prosedyrene var at i prosedyre 1 ble pseudosubstantivet presentert sammen med et bilde, mens deltakerne i prosedyre 2 kun så selve ordet på dataskjermen, illustrert i (4a, b). Bildene var de samme som ble brukt i studie 1, tatt fra NOUN-databasen (Horst & Hout 2016).

(4) Eksempel på et skjermbilde fra prosedyre 1 og prosedyre 2

<p>a.</p>  <p>tryspe</p> <p>en et ei</p>	<p>b.</p> <p>tryspe</p> <p>en et ei</p>
---	---

5.3. Resultater

I motsetning til studie 1 så vi ingen nevneverdig forskjell mellom de to prosedyrene her: Deltakernes valg av genus ble ikke påvirket av om de så et bilde av et objekt ordet skulle referere til eller ikke. Dette er illustrert i Figur 3, som viser nærmest identiske valg for hankjønn, hunkjønn og intetkjønn i prosedyre 1 og 2. På grunn av dette ble resultatene for begge prosedyrene slått sammen i den statistiske analysen.



Figur 4: Resultater av genustilordning for pseudosubstantiv ved prosedyre 1 (med bilde) og prosedyre 2 (uten bilde).

Sannsynligheten for at deltakerne valgte genus ut fra endelsen på substantivet ble analysert ved hjelp av en generalisert (binomial) blandet modell (lme4 R package, Bates *et al.* 2014) som inkluderte sannsynligheten for en respons i tråd med en gitt egenskap som avhengig variabel, «vilkår» som prediktor, og «deltaker» og «pseudosubstantiv» som tilfeldige effekter. Post hoc parvise sammenligninger (*lsmeans*, Lenth 2016) viste at sannsynligheten for å velge intetkjønn ved endelsen -v var lavere enn ved å velge hunkjønn for endelsen -e (SE = 0,219, z ratio = -3,427, p = 0,002) og også lavere enn å velge hankjønn ved substantiver som ender på andre konsonanter (-C) (SE = 0,221, z ratio = -5,642, p < 0,0001). Forskjellen mellom hankjønn ved -C og hunkjønn ved -e var bare marginalt signifikant (p = 0,0623). Det viktigste funnet er likevel at et separat sett med andre statistiske modeller viste at andelen av *en*, *ei* eller *et* er signifikant høyere ved substantiv med de tilsvarende endelsene (henholdsvis -C, -e og -v) enn ved de andre kategoriene: Valg av hunkjønn er signifikant

høyere ved substantiver med endelsen *-e* enn ved både *-C* (SE = 0,182, z ratio = 12,993, $p < 0,0001$) og *-v* (SE = 0,181, z ratio = 12,951, $p < 0,0001$). Det samme gjelder for valg av intetkjønn, som er høyere ved endelsen *-v* enn ved endelsen *-e* (SE = 0,248, z ratio = 6,690, $p < 0,0001$) og *-C* (SE = 0,251, z ratio = 4,347, $p < 0,0001$). Valg av hankjønn er signifikant høyere ved substantiver med endelsen *-C* enn med *-e* (SE = 0,227, z ratio = 5,269, $p < 0,0001$), mens den kun er marginalt høyere enn ved substantiver med endelsen *-v* ($p = 0,077$).

5.4. *Diskusjon*

Et viktig funn i studie 2 er at tilstedeværelse av bilde ikke har noe å si: Det var ingen statistiske forskjeller mellom de to prosedyrene, verken for andelen responser knyttet til de ulike egenskapene eller andelen responser fordelt på de tre genus. Samtidig ser vi altså at metoden har mye å si: I de to prosedyrene som ble gjennomført i studie 2, var det ingen nevneverdige forskjeller, mens dette, som nevnt, var tilfelle i studie 1. Dermed er det avgjørende hvilken metode som brukes for å undersøke genustildeling ved bruk av pseudosubstantiver.

I begge prosedyrene i studie 2 er det slik at deltakerne ble gjort eksplisitt oppmerksom på at studien dreide seg om genus. Dette øker sjansen for at de fokuserer mer på substantivenes form. Samtidig behandler de pseudosubstantivene forskjellig, slik at det er grunn til å tro at mønstrene vi ser, er reelle og ikke tilfeldige.

Det kanskje mest interessante funnet i studie 2 er at vi ser at korpusundersøkelsen får støtte fra de eksperimentelle dataene. Andelen hunkjønnsreponser er signifikant høyere for stimuli som ender på *-e*, og andelen intetkjønnsreponser er signifikant høyere for stimuli som ender på *-v*. Dette viser at reglene som beslutningstreet identifiserte, faktisk har psykologisk realitet. Språkbrukerne gjør bruk av disse reglene når de tildeler grammatisk genus til pseudosubstantiver. Vi ser også at hankjønn er det hyppigst brukte genus for både hunkjønns- og intetkjønnsbetingelsene, noe som bekrefter rollen hankjønn har som standardverdi for genustildeling i norsk (jf. Trosterud 2001).

6. Generell diskusjon

Vi stilte følgende forskningsspørsmål for denne artikkelen:

1. Finnes det pålitelige fonologiske egenskaper for genustildeling i norsk? Kan grammatisk kjønn predikeres basert på den fonologiske formen til substantivet?

2. Gjør talere av norsk bruk av probabilistiske fonologiske egenskaper når de tildeler grammatisk kjønn til pseudosubstantiver?
3. Spiller den eksperimentelle prosedyren en rolle?
4. Spiller tilstedeværelsen eller fravær av visuelle stimuli en rolle.

Vi så i avsnitt 2 at det er et uavklart spørsmål om det finnes regler for tildeling av grammatisk genus i norsk. I denne artikkelen har vi spesifikt sett på fonologiske egenskaper. Hovedfunnet vårt er at genustildeling i norsk for inanimater ikke er helt arbitrært: Språkbrukerne gjør bruk av egenskaper ved substantivene, og særlig gjelder dette endelsene *-e* og *-v*. Dette viser at de er sensitive til substantivenes form og bruker disse til å abstrahere regler som kan brukes på en produktiv måte. Det ser ikke ut til å være noen geografiske forskjeller mellom deltakerne, men her er også antall deltakere for flertallet av stedene for lavt til å utforske dette nærmere.

Bruken av disse fonologiske egenskapene er ikke kategorisk. Det er snakk om probabilistiske regler. Det er altså mer sannsynlig at en språkbruker vil tildele intetkjønn til et substantiv som ender på *-v* enn til substantiv med andre endelser. Det samme gjelder *-e* for hunkjønn. Men det betyr også at vi vil finne en rekke tilfeller der regelen ikke gjelder. Siden regelen er probabilistisk, gir det derfor ikke mening å betrakte disse som unntak. Vi har ingen konkrete forslag til hvordan en slik probabilistisk regel skal formaliseres, og eksperimentene våre har ikke blitt laget for å kunne skille ulike formaliseringer fra hverandre. Derfor går vi ikke videre inn på dette her.

Den eksperimentelle prosedyren har stor betydning for hvilke resultater vi får. Totalt har vi brukt fire ulike metoder, og tre av disse gir relativt like resultater, prosedyre 2 i studie 1 samt begge prosedyrene i studie 2. De tydelige forskjellene som kommer frem avhengig av metode, er uventet. Det er kanskje særlig overraskende at det første eksperimentet bare produserte hankjønn, all den tid dette paradigmet har vært brukt med suksess på en rekke språk og ulike norske dialekter. Dette tyder på at hankjønn som standardverdi for genus i norsk har en svært sterk stilling i genussystemet, slik at når deltakere i et eksperiment ikke blir eksplisitt gjort oppmerksom på genus, så har de ingen problemer med å bruke hankjønn overalt, helt uavhengig av hva formen på substantivet er. Det er kun når de fokuserer på genus at substantivets form spiller en (viss) rolle for genustilordning. Ellers viser likheten mellom resultatene fra de to prosedyrene i studie 2 at tilstedeværelse eller fravær av visuelle stimuli ikke har noe å si for genustilordningen. Dette tyder på at semantikken til et inanimat tellelig objekt ikke spiller noen rolle for genustilordning ved disse substantivene. Ellers vil vi

fremheve at en stor fordel med metoden vi brukte i studie 2 er at den er enkel å bruke og administrere, siden den er heldigital.

7. Konklusjon

Ved å bruke pseudosubstantiver for å studere forholdet mellom fonologiske egenskaper og genustildeling viser denne artikkelen at norske språkbrukere er sensitive til slike egenskaper ved substantiv. Dette gjelder særlig endelsene *-e* og *-v*, som predikerer henholdsvis hunkjønn og intetkjønn. Det er viktig å understreke at det ikke er snakk om kategoriske regler. De er probabilistiske, og dermed vil man også forvente unntak og tilfeller der de ikke gjelder. Resultatene våre viser at semantiske faktorer ikke spiller noen rolle for genustildeling ved inanimate tellelige substantiver og også at hankjønn helt klart har status som standardverdi for genustildeling i norsk.

Referanser

- Audring, Jenny. 2016. Gender. *Oxford Research Encyclopedia in Linguistics*. doi:10.1093/acrefore/9780199384655.013.43
- Bates, Douglas, Martin Mächler, Ben Bolker og Steve Walker. 2014. Fitting Linear Mixed-Effects Models Using lme4. *arXiv preprint arXiv:1406.5823*.
- Comrie, Bernard. 1999. Grammatical gender systems: a linguist's assessment. *Journal of Psycholinguistic Research* 28: 457–466.
- Conzett, Philipp. 2006. Gender assignment and the structure of the lexicon. *Sprachtypologie and Universalienforschung* 59: 223–240.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Drummond, Alex. 2012. IbxFarm (versjon 0.3.7). Online: <http://spellout.net/ibxfarm>
- Enger, Hans-Olav. 2001. Genus i norsk bør granskes grundigere. *Norsk Lingvistisk Tidsskrift* 19: 163–183.
- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Fenson, Larry, Philip S. Dale, J. Steven Reznik, Donna Thal, Elizabeth Bates, J. Paul Hartung, Steve Pethick og Judy S. Reilly. 1993. *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore: Paul H. Brookes Publishing Co.
- Gagliardi, Ann C. 2012. Input and intake in language acquisition. Doktorgradsavhandling, University of Maryland.

- Graedler, Anne-Line. 1998. Morphological, semantic and functional aspects of English lexical borrowings in Norwegian. Doktoravhandling, Universitetet i Oslo.
- Halse, Gro Egset. 2004. Genustilordning i nynorsk: Ei datamaskinell etterprøving. Hovedoppgave, Universitetet i Bergen.
- Hjelde, Arnstein. 1996. The gender of English nouns in American Norwegian. P. Sture Ureland og Iain Clarkson (red.): *Language Contact across the North Atlantic*, Tübingen: Max Niemeyer Verlag, 297–312.
- Horst, Jessica S. og Michael C. Hout. 2015. The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior research methods* 48: 1393–1409.
- Husby, Olaf. 1990. *Norske ord: Ordlagingslære med arbeidsoppgaver*. Oslo: Friundervisningens forlag.
- Johansson, Stig og Anne-Line Graedler. 2002. *Rocka, hipt og snacksy. Om engelsk i norsk språk og samfunn*. Kristiansand: Høyskoleforlaget.
- Leirvaag, Ole Gunnar. 1999. *Norsk på grunnlag av samisk*. 2. utg. Kautokeino: Samisk utdanningsråd.
- Lenth, Russell V. 2016. Least-Squares Means: The R Package ‘lsmeans’. *Journal of Statistical Software* 69. doi:10.18637/jss.v069.i01
- Lohndal, Terje og Marit Westergaard. 2021. Grammatical Gender: Acquisition, Attrition, and Change. *Journal of Germanic Linguistics* 33: 95–121.
- Mitrofanova, Natalia, Yulia Rodina, Olga Urek og Marit Westergaard. 2018. Bilinguals’ sensitivity to grammatical gender cues in Russian: the role of cumulative input, proficiency, and dominance. *Frontiers in Psychology* 9:1894. doi:10.3389/fpsyg.2018.01894
- Nordgård, Torbjørn. 1998. Norwegian computational lexicon (NorKompLeks). *Proceedings of the 11th Nordic Conference of Computational Linguistics (NODALIDA 1998)*.
- Næs, Olav. 1952. *Norsk grammatikk 1, ordlære: bokmål og nynorsk på bakgrunn av språkhistorie og dialekter*. Oslo: Fabritius.
- Rodina, Yulia, Tanja Kupisch, Natalia Meir, Natalia Mitrofanova, Olga Urek og Marit Westergaard. 2020. Internal and External Factors in Heritage Language Acquisition: Evidence from Heritage Russian in Israel, Germany, Norway, Latvia and the UK. *Frontiers in Education* 5:20. doi:10.3389/educ.2020.00020
- Rodina, Yulia og Marit Westergaard. 2017. Grammatical gender in bilingual Norwegian-Russian acquisition: The role of input and transparency. *Bil-*

- lingualism: Language and Cognition* 20: 197–214. doi:10.1017/S1366728915000668
- Rodina, Yulia og Marit Westergaard. 2015. Grammatical gender in Norwegian: Language acquisition and language change. *Journal of Germanic Linguistics* 27: 145–187.
- Rønhovd, Jarle. 1997. *Norsk morfologi*. Oslo: Gyldendal.
- Thornton, Anna M. 2009. Constraining gender assignment rules. *Language Sciences* 31: 14–32.
- Trosterud, Trond. 2001. Genustilordning i norsk er regelstyrt. *Norsk Lingvistisk Tidsskrift* 19: 29–58.
- Venås, Kjell. 1997. Annenopponent professor Kjell Venås, Universitetet i Oslo. *Norsk Lingvistisk Tidsskrift* 15: 90–100.
- Westergaard, Marit og Yulia Rodina. 2016. Hvor mange genus er det i Tromsø-dialekten? *Maal og Minne* 2 2016: 159–189.
- Witten, Ian H. & Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Øvrelid, Lilja og Petter Hohle. 2016. Universal Dependencies for Norwegian. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, 1579–1585.

Summary

The Norwegian grammatical gender system has traditionally been characterized as opaque. However, it is a matter of debate whether language users might nevertheless be sensitive to certain properties of nouns and also use these properties productively when assigning gender to unfamiliar nouns. In this article, we investigate whether and to what extent speakers are sensitive to phonological gender cues extracted from a text corpus. We report on two experimental studies that both focus on the tradeoff between masculine gender default and speakers' reliance on nominal endings in gender assignment. We find clear evidence supporting the status of masculine as the default gender in Norwegian, while we also demonstrate that phonological cues play a significant role in gender assignment. In addition, we show that the elicitation technique being used may significantly influence the results.

Olga Urek
Postdoktor i lingvistikk
UiT Norges arktiske universitet
9037 Tromsø
olga.urek@uit.no

Marit Westergaard
Professor i engelsk lingvistikk
UiT Norges arktiske universitet
9037 Breivika
Professor II i lingvistikk
NTNU Norges teknisk-natur-
vitenskapelige universitet
7491 Trondheim
marit.westergaard@uit.no

Terje Lohndal
Professor i engelsk lingvistikk
NTNU Norges teknisk-natur-
vitenskapelige universitet
7491 Trondheim
Professor II i lingvistikk
UiT Norges arktiske universitet
9037 Tromsø
terje.lohndal@ntnu.no