

TORE NESSET, HANS-OLAV ENGER OG
LAURA A. JANDA

Er politiet sikker eller sikre? Adjektivkongruens ved kollektiver i norsk

1 Problemet: predikativt adjektiv i entall eller flertall?

I norsk kan predikativt adjektiv stå i entall eller flertall når subjektet er et kollektiv:¹

- (1) Politiet var **sikker** på at mannen [...] var død. (*Sosialnytt* 18.8.2019)
- (2) Politiet var **sikre** på at mannen var død [...]. (*Dagbladet* 1.12.2019)

Takk til Helge Dyvik for hjelp med data fra NorGramBank og til Marianne Lund og Endre Machlik for hjelp til å klassifisere dataene fra NoWaC-korpuset. Vi vil også takke deltagerne på NORKOG-konferansen i 2021 og et digitalt seminar i 2022 for kommentarer og innspill. Takk også til medlemmer av CLEAR-gruppen ved UiT Norges Arktiske Universitet og to anonyme konsulenter for kommentarer på en tidligere versjon av artikkelen.

¹ Eksempelene er herfra: <https://sosialnytt.com/finner-en-mumie-i-skogen-sa-apner-den-oynene/> og <https://www.dagbladet.no/nyheter/bildene-snur-drapsmysteriet-pa-hodet/71872793>.

Nesset, T., professor of Russian linguistics, UiT The Arctic University of Norway, Enger, H.-O., professor of Scandinavian linguistics, University of Oslo, & Janda, L. A., professor of Russian linguistics, UiT The Arctic University of Norway, "Is the police certain-sg or certain-pl? Adjective agreement on collective nouns in Norwegian". *ANF* 137 (2022), pp. 123–144.

Abstract: This is an empirical study of adjectival agreement in Norwegian. We investigate sentences with a collective noun as a subject and a predicative adjective in the singular or plural. On the basis of data from two corpora and a survey, we demonstrate that plural agreement is widespread in Norwegian sentences of the relevant type. The following factors are shown to be significantly associated with the choice between singular and plural: animacy and semantic type of subject, distance between subject and predicative adjective, adjective vs. participle, and the adjective's role as an indicator of agency. The corpus data are analyzed by means of a logistic regression model.

Keywords: Norwegian, agreement, collective noun, adjective, logistic regression, corpus.

DOI: <https://doi.org/10.63420/anf.v137i.27899>

I begge eksemplene er *politiet* subjektet. Dette er et kollektivt substantiv i den forstand at det er et ord i entall som betegner en gruppe individer. I (1) står predikativet i entall (*sikker*), mens (2) har predikativ i flertall (*sikre*).

Ettersom *politi* er et intetkjønnsord, kunne man *a priori* også vente predikativ i intetkjønn entall. Faarlund et al. (1997: 771) stjernemerker likevel eksemplet **Politiet er sikkert på at dei har teki rett mann*. Vi har ikke eksempler med *sikker* i intetkjønn i vårt datamateriale, men intetkjønnsformen *mistenksomt* forekommer i (3), mens vi har flertall (*mistenksomme*) i (4):

- (3) Politiet var ikke **mistenksomt**? spør Paul. (*Historien om Gottlob* av Torolf Elster, fra NorGramBank)
- (4) Hun følte ikke desto mindre at politiet var litt **mistenksomme**. (*Det perfekte mord* av Peter James, kapittel 15, fra Google Books)

I det følgende skal vi ikke diskutere genus, men nøye oss med å undersøke motsetningen mellom entall (som i (1) og (3)) og flertall (som i (2) og (4)).

At kollektive substantiv kan ta predikativ i entall og flertall, er velkjent fra mange språk. Mest kjent er kanskje såkalte komité-substantiv ("committee nouns", Corbett 2006: 211–213) som i enkelte varieteter av engelsk kan veksle mellom entalls- og flertallskongruens. Levin (2001: 9) fant i en og samme artikkel i avisen *The Independent* at det kollektive substantivet *press* kunne forekomme med både entalls- og flertallskongruens på verbalet (*the press writes* og *the press have*). I russisk har vi samme type variasjon i setninger med kvantifiserte nominalfraser som subjekt, for eksempel kvantorer som *mnogo* 'mange, mye' (Corbett 2006: 192). Fenomenet er også kjent fra svensk (Hagåsen 1992) og dansk. Hansen & Heltoft (2011: 875) diskuterer for eksempel forskjellen i dansk mellom *politiet er ikke dygtigt nok* og *politiet er ikke dygtige nok*.

Det er derfor ikke overraskende å finne variasjon mellom entalls- og flertallskongruens i norsk, og fenomenet nevnes allerede av Western (1921: 187–190). Knudsen (1973) drøfter emnet utførlig, og fenomenet er beskrevet i *Norsk referansegrammatikk* (Faarlund et al. 1997: 769–772). I den senere tid er fenomenet blitt behandlet av Åfarli og Vangsnes (2021). Noen omfattende statistisk analyse av korpusdata foreligger imidlertid ikke for norsk. Vårt mål med denne artikkelen er å fylle dette kunnskaps-hullet. Vi undersøker to spørsmål: (a) hvor vanlig er flertallskongruens ved kollektive substantiv? Og (b) hvilke faktorer motiverer valget mellom entall og flertall i predikativet?

Vårt bidrag kan sammenfattes på følgende måte: For det første indikerer våre empiriske undersøkelser at flertallskongruens er vanlig. Men det er stor variasjon idet data fra trebanken har større andel flertallskongruens enn data fra NoWaC-korpuset. For det annet identifiserer vi flere faktorer som påvirker valget mellom entalls- og flertallskongruens. Noen av faktorene er knyttet til subjektet. Dette gjelder om subjektet refererer til mennesker eller dyr, og om subjektet hører til ulike grupper av mennesker (politiske partier, geografiske enheter, organisasjoner osv.). Andre faktorer kan knyttes til predikativet. Om predikativet er et adjektiv eller partisipp, er av betydning, og det samme gjelder om adjektivet indikerer agentivitet eller ikke. En tredje type faktorer refererer til forholdet mellom subjekt og predikativ. Vi finner at avstanden mellom de to setningsleddene er av betydning for valget mellom entalls- og flertallskongruens.

I avsnitt 2 gir vi en oversikt over vårt datamateriale, før vi i avsnitt 3 diskuterer hvor vanlig flertallskongruens er. Våre hypoteser beskrives i avsnitt 4 og testes mot korpusdata i avsnitt 5. Avsnitt 6 og 7 drøfter informantdata. Vi sammenfatter våre funn i avsnitt 8.

Før vi går videre, er en terminologisk bemerkning nødvendig. Entallskongruens som i (1) og (3) omtales i faglitteraturen ofte som ”syntaktisk” eller ”formell kongruens” (”syntactic/formal agreement”), mens termene ”semantisk kongruens” eller kongruens ”with greater semantic justification” (Corbett 2006: 156) brukes om eksempler som (2) og (4). Dette begrepsparet er problematisk (Dahl 2000). I vårt tilfelle skaper det et inntrykk av at entallskongruens mangler semantisk motivasjon. Om vi tar *politi* som eksempel, er dette formelt sett et entallssubstantiv som betegner en enhet (politietaten) som består av mange individer (de ulike ansatte i etaten). Ettersom betydningen både inneholder et entalls- og et flertallselement, kan en argumentere for at entallskongruens er like ”semantisk” som flertallskongruens. At entallskongruens også har formell motivasjon, mens flertallskongruens mangler slik motivasjon, er en annen sak. I det følgende skal vi unngå dette terminologiske problemet ved å referere til de to kongruensmønstrene som entalls- og flertallskongruens.

2 Våre data

Vi har hentet data fra tre ulike kilder. Først laget vi en database med eksempler fra NoWaC-korpuset, et stort korpus med eksempler fra inter-

nett.² Dette korpuset har den fordel at det er svært omfattende, men en ulempe er at innholdet ikke er ”kvalitetssikret” (kuratert); eksemplene er hentet direkte fra internett uten noen form for ”siling”. Vi søkte etter følgende substantiv:

- (5) *Arbeiderpartiet, EU, familie, flertall, forsvar, høyreside, kommune, ledelse, menighet, parti, politi, publikum, RBK [Rosenborg], regjering, Rosenborg [fotballklubb], Telenor, ungdom*

Vi valgte disse ordene for å danne oss et inntrykk av hvordan ulike kollektiver oppfører seg. Å lage et representativt utvalg av norske kollektiver er vanskelig, men kollektivene i (5) dekker et bredt spekter. Vi har med egennavn og fellesnavn, forkortelser og ulike grammatiske kjønn. Vi søkte etter substantivene sammen med et kopulaverb (*være* og *bli*) og et adjektiv i entall eller flertall. Søkene omfattet avstander fra 0 til 3 ord mellom kopulaverbet og adjektivet. Eksemplene ble gjennomgått manuelt, og irrelevante eksempler (”støy”) ble luket ut. Etter dette sto vi igjen med en database på 2397 eksempler.

Dataene fra NoWaC ga et godt utgangspunkt for analyse, men vi hadde behov for å undersøke et bredere utvalg av kollektiver. Derfor gransket vi kollektivene i den norske trebanken NorGramBank knyttet til INESS (Infrastructure for the Exploration of Syntax and Semantics, Rosén et al. 2012).³ I NoWaC-korpuset viste det seg vanskelig å analysere predikativer langt unna subjekt og kopulaverb; ved større avstander enn 3 ord ble andelen irrelevante eksempler så stor at den manuelle utlukingen ble for krevende. I en trebank er de syntaktiske relasjonene mellom søkeordene spesifisert, og dermed er det mulig å identifisere relevante eksempler med svært stor avstand mellom subjekt og predikativ. Vi fant eksempler med opp til 11 ord mellom subjekt og predikativ. Resultatet av søkene var en database med 511 eksempler. Her er språket kvalitetssikret i den forstand at tekstene som ligger til grunn for trebanken, er valgt ut og kontrollert av dem som har laget trebanken.

En tredje kilde vi hentet data fra, var et enkelt eksperiment der 92 bachelorstudenter ved Universitetet i Oslo fylte ut et digitalt spørreskjema i september 2021. Studentene studerte nordisk med fokus på norsk grammatikk. De fikk ikke vite hva som var hensikten med undersøkelsen

² NoWaC-korpuset, som inneholder ca. 700 millioner ord, er tilgjengelig her: https://tekstlab.uio.no/glossa2/nowac_1_1.

³ NorGramBank beskrives nærmere her: <https://norgramtall.w.uib.no/om-den-ne-bloggen/>. For INESS, se: <http://clarino.uib.no/iness>.

på forhånd. Respondentene ble bedt om å sette inn en form de ”synes passer av ordene i parentes”. De fikk for eksempel følgende setning hvor de skulle sette inn *klar*, *klart* eller *klare*:

- (6) Da ferien endelig kom, var familien typisk nok alt annet enn... (klar).

Spørreskjemaet inneholdt til sammen 10 setninger med adjektivene *klar* og *stor* i kombinasjon med fem ulike substantiv som subjekt: *gruppe*, *band*, *familie*, *saueflokk* og *hundespenn*. Testsetningene var omgitt av 15 fyllsetninger. Alle deltagerne fikk testsetningene i samme rekkefølge. Eksperimentet gjorde det mulig å teste effekten av animatumhierarkiet og agentivitet – faktorer som ikke lot seg undersøke med utgangspunkt i korpusdataene. Eksperimentet ga en database på 920 datapunkter. Vi analyserte ikke informasjon om talernes kjønn, alder, utdanning eller dialekt, fordi sosiolingvistiske parametre ikke var tema for vår undersøkelse.

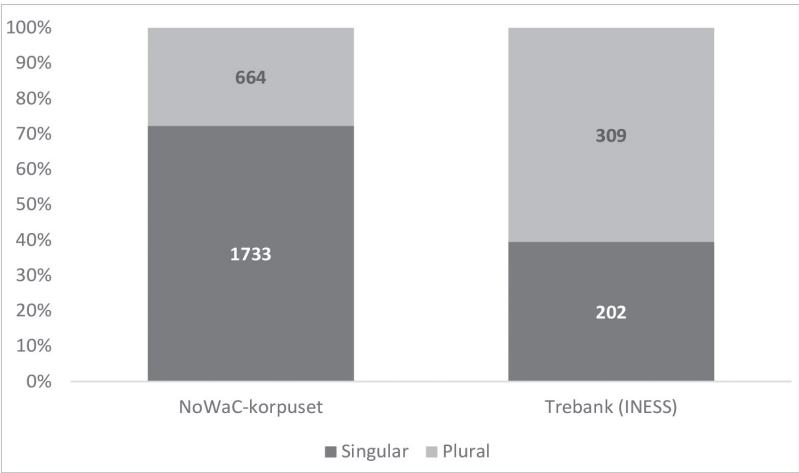
3 Hvor vanlig er flertallskongruens?

Ut fra korpusdataene våre har vi et godt grunnlag for å svare på hvor utbredt flertallskongruens er. Tabell 1 viser absolutte tall for de to korpusene, mens Figur 1 i tillegg får frem den prosentvise fordelingen mellom entalls- og flertallskongruens. Først og fremst viser våre korpusdata at flertallskongruens ikke kan avskrives som et marginalt fenomen i norsk. Selv i dataene fra NoWaC-korpuset, der vi fant minst flertallskongruens, har om lag 28 % av eksemplene våre adjektiv i flertall. I trebankdataene er andelen med flertall hele 60 %.

Som vist i Tabell 1 og Figur 1 er det stort sprik mellom de to korpusene. Hva kan være årsaken til at vi fant mer flertall i trebankdataene? Det er ikke mulig å gi noe sikkert svar på dette, men i utgangspunktet er det ikke unaturlig at vi får forskjellige resultater, ettersom de to korpusene er forskjellige, og ettersom dataene ble samlet inn på ulike måter. En mulig forklaring kan være at dataene fra NoWaC-korpuset bare omfatter de 17 substantivene i (5), mens trebankdataene har et rikere tilfang av substantiv. Levin (2001: 147–148, 2006), som analyserte engelske data, kom frem til at det var stor variasjon mellom leksemer i hans data. Våre data fra NoWaC-korpuset tyder på at dette også er tilfellet i norsk. Som vi ser i Tabell 2 og Figur 2, er det store forskjeller mellom substantivene vi har undersøkt. Vi lister først ordene som har størst andel flertall, og så

	Singular	Plural	Totalt
NoWaC-korpuset	1733	664	2397
Trebank (INESS)	202	309	511

Tabell 1. Fordeling av entalls- og flertallskongruens i data fra to korpus.



Figur 1. Fordeling av entalls- og flertallskongruens i data fra to korpus.

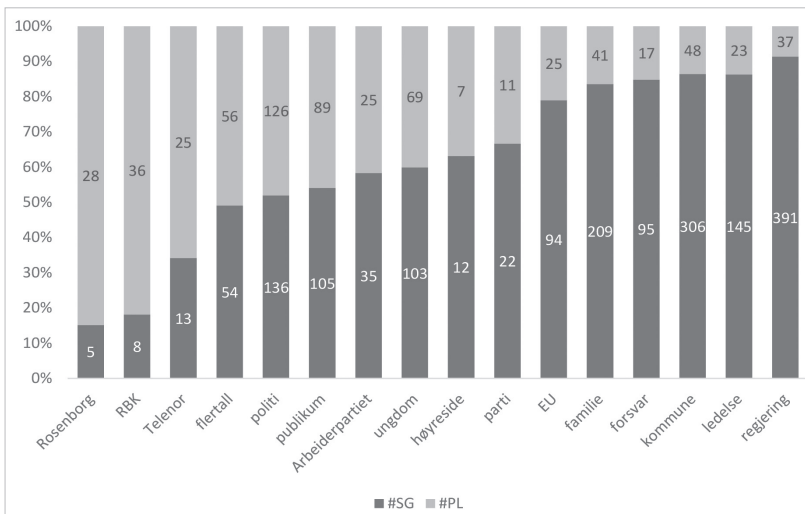
i synkende rekkefølge. Mest flertall har fotballklubben *Rosenborg* med 28 av 33 eksempler i flertall (85 %). I den andre enden av skalaen finner vi *regjering*, som i vårt materiale bare hadde flertall i 9 % av eksemplene (37 av 428).

En annen forskjell mellom dataene fra de to korpusene har med avstanden mellom subjekt og predikativ å gjøre. Som nevnt i avsnitt 2 inneholder trebankdataene eksempler med avstand opp til 11 ord, mens vi i NoWaC-dataene opererer med avstander fra 0 til 3 ord. Som vi skal se i avsnitt 5, øker sjansen for flertallskongruens med økende avstand mellom subjekt og predikativ. Dette kan være en av grunnene til at vi fant mer flertallskongruens i trebankdataene.

En tredje faktor som kan være relevant, er forholdet mellom skriftlig og muntlig språkbruk. Det ligger nær å anta at flertallskongruens er hyppigere i muntlig enn skriftlig norsk, ettersom muntlig språk generelt

	#SG	#PL
Rosenborg	5	28
RBK	8	36
Telenor	13	25
flertall	54	56
politi	136	126
publikum	105	89
Arbeiderpartiet	35	25
ungdom	103	69
høyreside	12	7
parti	22	11
EU	94	25
familie	209	41
forsvar	95	17
kommune	306	48
ledelse	145	23
regjering	391	37
menighet	0	1
Totalt	1733	664

Tabell 2. Fordelingen av entalls- og flertallskongruens for ulike substantiv i NoWaC-korpuset.



Figur 2. Fordelingen av entalls- og flertallskongruens for ulike substantiv i NoWaC-korpuset. *Menighet* fra Tabell 1 er utelatt i figuren, da dette substantivet bare er representert med ett eksempel i vårt materiale.

er mindre preget av normativ grammatikk. Trebankdataene inneholder eksempler fra forhandlinger i Stortinget, som er nærmere spontan muntlig tale enn data fra skriftlige kilder, selv om stortingsforhandlingene er skrevet ned og derfor kan være påvirket av normativ grammatikk. Vi har dessverre ikke stort nok materiale til å teste effekten av muntlig språk. Uansett er det klart at flertallskongruens ikke er et marginalt fenomen.

4 Hvilke faktorer kan være relevante?

I faglitteraturen omtales en rekke faktorer som kan ha betydning for kongruens. Vi har allerede nevnt avstanden mellom subjekt og predikativ som en mulig faktor. Corbett (2006: 235–236) behandler denne faktoren fra en språktypologisk synsvinkel, og fenomenet er kjent fra bl.a. engelsk (Levin 2001: 92–99). *Norsk Referansegrammatikk* (1997: 766) sier (riktignok ikke med direkte adresse til kollektiver): ”Jo større avstand det er i setninga mellom basen og predikativet, jo lettere har språkkjensla for å godta manglande kongruens” (se Åfarli og Vangsnes 2020: 535 for kritisk diskusjon, men se også Denstad 2021: 53–54 for evidens for at avstand i alle fall er relevant for et annet kongruensfenomen i norsk). Referansegrammatikkens analyse innebærer at vi skulle forvente mer flertallskongruens (som er et eksempel på det som referansegrammatikken kaller ”manglande kongruens”) jo større avstanden er mellom subjekt og predikativ. Vi vil teste denne hypotesen mot korpusdata.

Som vi så i avsnitt 3, er det stor leksikalsk variasjon i vårt materiale fra NoWaC-korpuset. Noen leksemer har oftere flertallskongruens enn andre. Vi ønsker å kartlegge den leksikalske variasjonen mot data fra trebanken. Men i stedet for å se på enkeltsubstantiver, har vi valgt å dele dem inn i grupper etter betydningen: politiske partier (f.eks. *Høyre*), geografiske enheter (f.eks. land som *Norge* og verdensdeler som *Europa*) og organisasjoner, etater og institusjoner (f.eks. *NAV* (også kjent som ”Ny arbeids- og velferdsforvaltning”). I vårt materiale er *politiet* svært hyppig, og vi skilte derfor dette ordet ut som en egen kategori. I tillegg opererer vi med en restkategori av ”andre kollektiver”.⁴

⁴ I kategorien ”andre kollektiver” finner vi følgende substantiv: *ambulansepersonale*, *befolkning*, *betjening*, *borgerskap*, *bygdefolk*, *Cassidy-familien*, *dansebandpublikum*, *familie*, *fjordfolk*, *flertall*, *flestepart*, *FN-personale*, *folk*, *gammelfolk*, *generasjon*, *gjeng*, *gruppe*, *gutttegning*, *Henie-familien*, *HIV-vaksinemiljø*, *Holinfolket*, *hovedpart*, *hulderfolk*, *jegerfolk*, *klimaforskningsmiljø*, *komite*, *kommisjon*, *krets*, *mannskap*, *mengde*,

En faktor som av og til nevnes, er om predikativet er et adjektiv eller partisipp (Knudsen 1973: 39–40). I bokmål er nemlig adjektiv bøyelige i større grad enn partisipper (se f.eks. Dyvik 2019 for inngående diskusjon). Vi forventer dermed hyppigere flertallskongruens ved adjektiv enn ved partisipp. Her må det riktignok nevnes at det er vanskelig å trekke en klar grense mellom de to kategoriene, slik Dyvik (2019) peker på. For å gjøre det mulig å gjennomføre en statistisk analyse har vi likevel klassifisert våre eksempler som enten adjektiv eller partisipp. Her har vi tatt utgangspunkt i et tradisjonelt kriterium: Kan man sette *svært* foran ordet, kan det klassifiseres som et adjektiv (Lie 2003: 77). Dette kan illustreres med konstruerte eksempler. I *Han var nedslått* er det mulig å sette inn *svært*, og dermed kan *nedslått* klassifiseres som adjektiv, selv om det historisk har sammenheng med verbet *slå*. I *Han var slått* virker det lite rimelig å sette inn *svært*, og derfor klassifiseres *slått* som partisipp.

En faktor som så vidt har vært nevnt i litteraturen om kongruens i norsk, er ”blokking” (engelsk: ”chunking”), dvs. det fenomen at flere ord som brukes ofte sammen, ”smelter sammen” til en enhet (Bybee 2010: 33–34). Rad (2021) hevder i tråd med Knudsen (1973) at ”blokker” som *glad i* og *redd for* ofte forekommer i entall selv om subjektet er et substantiv i flertall. Med utgangspunkt i korpusdata vil vi undersøke om denne hypotesen lar seg overføre til kollektiver.

Barlow (1999: 92) sier at i mange språk er det ikke flertallskongruens ved inanimate subjekter. Dermed er det relevant å trekke inn ”animatumhierarkiet”. Hierarkiet, som også kalles ”individueringshierarkiet” (Sasse 1993), har vist seg relevant for mange ulike fenomener i forskjellige språk. Hierarkiet har mange varianter, men en enkel versjon går fra mennesker (øverst) via dyr og konkrete gjenstander til masser og abstrakte begreper (nederst). Med utgangspunkt i Barlow skulle en forvente mindre flertallskongruens jo lenger ned i animatumhierarkiet subjektet kommer. Vi ønsket å teste denne hypotesen mot korpusdata, men vi fant ikke eksempler med inanimate subjekter i vårt datamateriale. I stedet tester vi hypotesen mot data fra eksperimentet (se avsnitt 2 og 6).

En faktor som er beslektet med animatumhierarkiet, er ”agentivitet”, dvs. evnen til å opptre som en ”agens” som utfører en kontrollert handling. Sammenlign setningene nedenfor:

- (7) Debutanten ga Drillo beskjed om at han var **klar**. (Aftenposten 2013)

mestedel, miljø, målfolk, overklasse, panel, personale, politikarpanel, publikum, storpart, vertsfolk og åttitallegenerasjon.

- (8) Sandnes Ulf-nykommeren hylles av treneren: – **Han var stor.** (Aftenposten 2019)

I (7) innebærer adjektivet *klar* at subjektet er rede til å utføre en handling, altså opptre som agens. I (8), derimot, foreligger det ikke noen slik implikasjon. Adjektivet *stor* beskriver bare en (metaforisk) egenskap ved fotballspilleren det er snakk om. I eksperimentet sammenligner vi bruken av disse adjektivene. Vår hypotese er at adjektiv som impliserer agentivitet, har større sjanse for å ha flertallskongruens.

5 Statistisk undersøkelse av trebankdata

Med utgangspunkt i faktorene diskutert i avsnitt 4 gjennomførte vi en logistisk regresjonsanalyse av data fra NorGramBank.⁵ Vi fulgte prosedyrene for denne type analyse som er beskrevet i detalj i Gries (2021). Hypotesen var:

- (9) Numerus ~ Avstand.bcn.sc + Subst_kategori + Adj_partisipp + Blokking⁶

Dette kan leses slik: ”numerus i adjektivet kan predikeres som en funksjon av avstand mellom subjekt og predikativ, substantivets betydning (”substantiv_kategori”), adjektiv vs. partisipp og blokking. Blokking viste seg å ikke være signifikant, og en drop1()-funksjon viste at denne faktoren burde fjernes fra modellen. Vi vil ikke diskutere denne faktoren videre, siden analysen vår tyder på at den ikke har noen betydning for valget mellom entall og flertall i adjektivet. Den optimale modellen blir dermed:

- (10) Numerus ~ Avstand.bcn.sc + Subst_kategori + Adj_partisipp

Tabell 3 oppsummerer resultatene av regresjonsanalysen med den optimale modellen. ”Intercept” representerer setninger der *politi* er subjekt og kongruerer med et adjektiv (ikke et partisipp). Avstanden mellom subjekt

⁵ Data og skript er tilgjengelig i TROLLing: <https://doi.org/10.18710/TGW1EH>.

⁶ I formelen står bcn for “Box-Cox transformation to normality”. Denne transformasjonen er gjennomført fordi distribusjonen for ulike avstander er svært skjev i vårt materiale. Se: https://www.researchgate.net/publication/322920095_COMBINING_THE_BOX-COX_POWER_AND_GENERALISED_LOG_TRANSFORMATIONS_TO_ACCOMMODATE_NONPOSITIVE_RESPONSES_IN_LINEAR_AND_MIXED-EFFECTS_LINEAR_MODELS. Forkortelsen sc står for “scaled centered”.

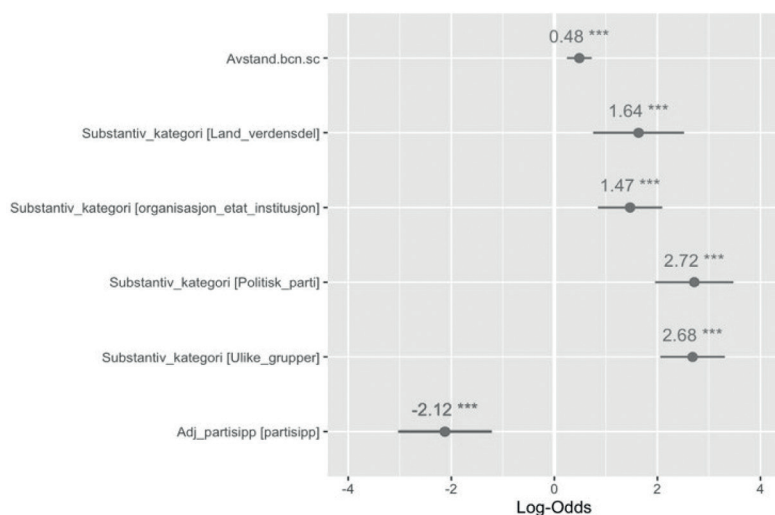
Coefficients:	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.6353	0.1728	-3.677	0.000236	***
Avstand.bcn.sc	0.4847	0.1225	3.957	7.60e-05	***
Subst_kategoriLand_verdensdel	1.6350	0.4502	3.632	0.000281	***
Subst_kategoriorg_etat_inst	1.4702	0.3180	4.623	3.79e-06	***
Subst_kategoriPolitisk_parti	2.7156	0.3873	7.011	2.36e-12	***
Subst_kategoriUlike_grupper	2.683	0.3192	8.406	< 2e-16	***
Adj_partisipppartisipp	-2.1229	0.4626	-4.590	4.44e-06	***

Tabell 3. Resultat av regresjonsanalysen – koeffisienter.

og predikativ er tilnærmet lik 2.⁷ Modellen sammenligner intercept med andre kombinasjoner av verdier for de ulike faktorene. Estimater for intercept er et negativt tall. Dette indikerer at *politi* foretrekker entall fremfor flertall i adjektivet. Avstand i neste rad i tabellen har et positivt estimat. Dette indikerer at sjansen for flertallskongruens øker når avstanden mellom subjekt og predikativ blir større. De positive estimatene for de ulike semantiske gruppene av substantiv viser at disse faktorene har større sjanse for flertallskongruens enn adjektiv ved intercept. For faktoren adjektiv vs. partisipp i nederste rad i tabellen har vi et negativt estimat. Dette innebærer at partisipp øker sjansen for entallskongruens sammenlignet med adjektiv ved intercept. Kolonnene for standardavvik ("Std. Error") og z-verdi ("z value") vil ikke bli omtalt nærmere. Kolonnen Pr(>|z|) angir statistisk signifikans (p-verdi). Alle verdiene er lavere enn 0.001 og er derfor merket med tre stjerner (***). Dette angir at det er svært lav sannsynlighet for at resultatene skyldes tilfeldigheter.

Resultatet fra Tabell 3 visualiseres i Figur 3. Faktorene som er markert med blått på høyre side i figuren, øker sjansen for flertallskongruens. Som vi ser, gjelder dette avstand mellom subjekt og visse semantiske grupper av substantiv. Adjektiv vs. partisipp nederst i figuren kommer ut i rødt. Det betyr at partisipp har mindre sannsynlighet for flertallskongruens enn adjektiver. Hver faktor er representert med et punkt på en linje. Linjen representerer konfidensintervallet på 95 %. Jo kortere linje, jo sikrere kan vi være på at modellens prediksjoner er korrekte.

⁷ Modellen tar utgangspunktet i den gjennomsnittlige avstanden mellom subjekt og predikativ som er 2,121.

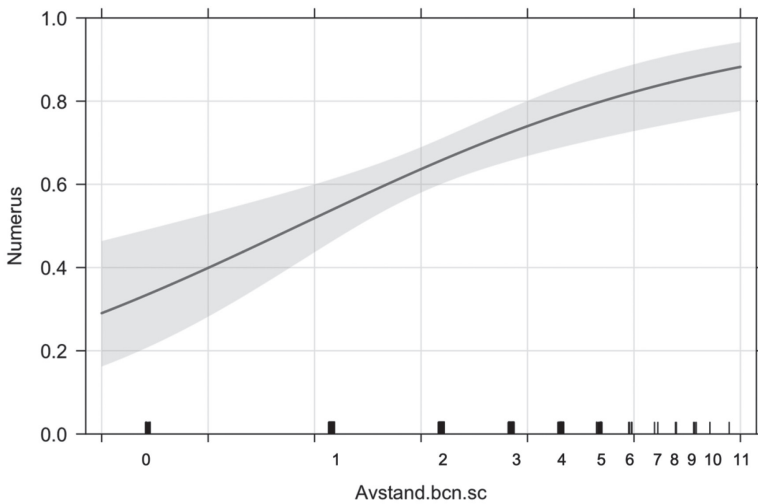


Figur 3. Visualisering av logistisk regresjonsanalyse.

R-kvadratet for modellen er 0,455. Dette viser at modellen vår dekker 45,5 % av variasjonen i datamaterialet. C-score er kalkulert til 0,846. Her regnes alt over 0,8 som utmerket. Modellen gir korrekte prediksjoner i 77,5 % av tilfellene. Vi testet kollinearitet og fikk vif-verdier ("variance inflation factor") lavere enn 1,2. Restleddene ("residuals") har normal distribusjon. En valideringstest viste at det ikke foreligger overtilpasning ("overfitting"). I sum viser disse diagnostiske kriteriene at modellen gir et svært godt bilde av situasjonen.

La oss se nærmere på hver faktor for seg. Figur 4 visualiserer effekten av avstanden mellom subjekt og predikativ. X-aksen angir avstanden i antall ord fra 0 til 11, mens y-aksen angir andelen av eksempler med flertallskongruens.⁸ Når verdiene på y-aksen overstiger 0,5, predikerer vi flertall i adjektivet, mens verdier lavere enn 0,5 innebærer entallskongruens. Som vi ser av kurven, øker sjansen for flertallskongruens med økende avstand. Ved minste avstand (0 ord mellom subjekt og predikativ) har vi flertallskongruens i omtrent 30 % av eksemplene, mens for største avstand (11 ord mellom subjekt og predikativ) er andelen med flertallskongruens økt til ca. 90 %.

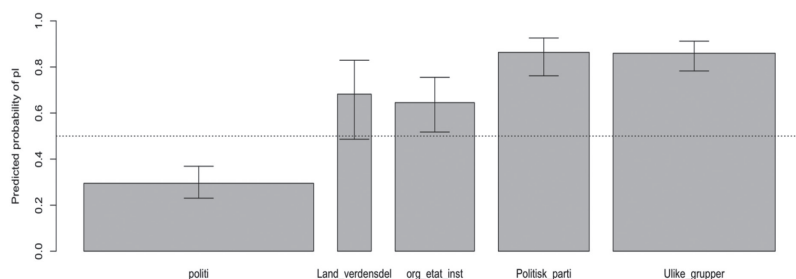
⁸ Mellomrommene mellom tallene langs x-aksen er ulike. Dette er fordi vi gjennomførte en "Box-Cox transformation to normality". Se fotnote 6 ovenfor.



Figur 4. Effekten av avstand mellom subjekt og predikativ.

Det grå båndet rundt kurven angir konfidensintervallet (95 %). Når det grå båndet er smalt, betyr det at prediksjonen er sikrere enn når båndet er bredere. Figuren viser at båndet er smalest på midten (avstander fra 1–3 ord). Dette er de avstandene som vi har flest eksempler med. Når materialet er større, blir prediksjonene sikrere. Figuren viser at hypotesen bekreftes av den statistiske undersøkelsen. Med større avstand mellom subjekt og predikativ øker sjansen for flertallskongruens.

Figur 5 visualiserer effekten av subjektets semantikk på valget mellom entalls- og flertallskongruens. Hver søyle representerer en av de semantiske gruppene beskrevet i avsnitt 4. Søylen har ulik tykkelse. Tykke søyler representerer kategorier med mange eksempler, mens tynne søyler betyr at kategorien har færre eksempler i vårt datasett. Høyden på søylene angir hvor sannsynlig flertallskongruens er. Jo høyere søyler, jo mer sannsynlig er flertallskongruens. Søyler over den vannrette prikkede linjen ved 0,5 viser at sjansen er størst for flertall i adjektivet, mens søyler under linjen angir at entallskongruens er mest sannsynlig. "Værhårene" (engelsk: "whiskers"), linjene øverst på hver søyle, representerer konfidensintervallet på 95 %. Jo kortere linjene er, jo sikrere er modellens prediksjoner.



Figur 5. Effekten av subjektets semantikk.

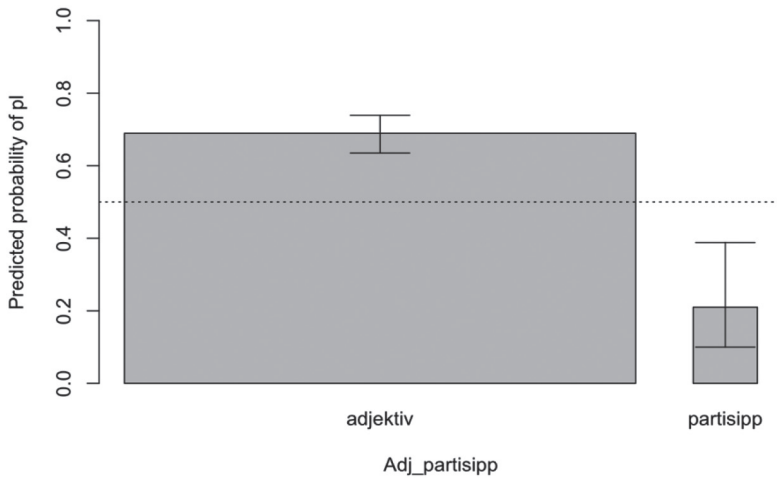
Vi ser at *politi* har den klart laveste andelen med flertallskongruens. Dette er den eneste kategorien der sannsynligheten for flertallskongruens er under 50 %. For de andre gruppene er flertall det mest sannsynlige valget. Vi kan uttrykke dette som et hierarki der sjansen for flertallskongruens minker fra venstre mot høyre:

- (11) Politisk parti/ulike grupper > land/verdensdel > organisasjon, etat, institusjon > politi

Legg merke til at restkategorien ”ulike grupper”, som er beskrevet nærmere i fotnote 4 i avsnitt 4, har stor sannsynlighet for flertallskongruens. Her ville det være ønskelig med en grundigere undersøkelse av mer finkalibrerte inndelinger av dataene, men dessverre er eksempelmaterialet vårt, når det blir brutt ned i slik detalj, for lite til at det er mulig å gjennomføre meningsfulle statistiske undersøkelser.

Det er også verdt å merke seg at *politi* står ved siden av kategorien ”organisasjon, etat, institusjon” i (11). Dette er betryggende, ettersom politiet er en offentlig etat som bare ble skilt ut som egen kategori fordi ordet forekommer hyppig i vårt materiale. Samtidig har *politi* betydelig lavere sannsynlighet for flertallskongruens enn andre etater i vårt data-materiale.⁹ Det er altså grunn til å tro at det kan være stor variasjon blant kollektiver som betegner ulike organisasjoner, etater og institusjoner.

⁹ En anonym konsulent foreslår at normativ grammatikk kan være en årsak til at *politi* har lavest sannsynlighet for flertallskongruens i Figur 5. Som konsulenten påpeker, har mange organisasjoner egne skriveveiledninger som kan tenkes å påvirke kongruensforholdene ved spesielle substantiv. Dette er utvilsomt en mulighet, men vi har ikke funnet noe konkret om *politi* på Språkrådets nettsider eller andre kilder (f.eks. Vinje 2002) som underbygger konsulentens idé.



Figur 6. Effekten av adjektiv vs. partisipp.

Igien ville en mer detaljert analyse vært ønskelig, men til dette er vårt datamateriale ikke stort nok. Uansett viser vår statistiske analyse klart at subjektets semantikk har betydning for valget mellom entalls- og flertallskongruens.

Figur 6 viser effekten av adjektiv versus partisipp for valget mellom entalls- og flertallskongruens. Som vi ser, er søylen for partisipp mye tynnere enn søylen for adjektiv. Dette er fordi partisipp utgjør en relativt liten del av vårt datamateriale. Når det er sagt, viser figuren klart at denne faktoren er relevant. Flertallskongruens er langt mindre sannsynlig for partisipp enn for adjektiv.

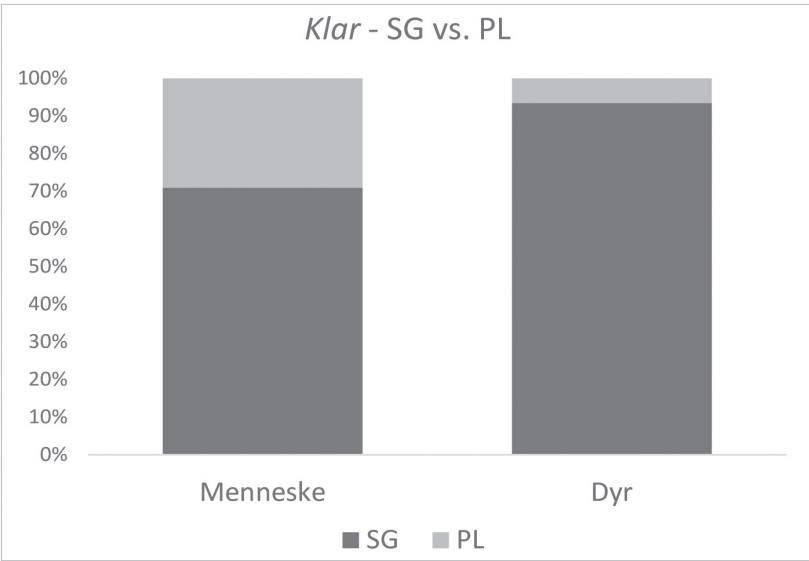
6 Animatumhierarkiet – eksperimentdata

Som nevnt i avsnitt 4 ønsker vi å teste hypotesen om at flertallskongruens er mest sannsynlig når subjektet betegner mennesker. Alle kollektivene i korpusmaterialet betegner mennesker, så for å teste hypotesen var det nødvendig å gjennomføre et eksperiment. Som nevnt i avsnitt 2 omfattet eksperimentet følgende substantiver: *gruppe*, *band*, *familie*, *saueflokk* og *hundespann*. Tabell 4 og Figur 7 viser resultatene for adjektivet *klar* sammen med disse substantivene. Som vi ser, er sjansen for flertallskon-

gruens langt høyere for subjekter som refererer til mennesker enn til dyr. Forskjellen er statistisk signifikant med en moderat effektstørrelse.¹⁰ Hypotesen om at flertallskongruens er mer sannsynlig for subjekter som betegner mennesker, bekreftes dermed av våre data.

Substantiv	SG	PL	Totalt
Gruppe	58	31	89
Band	62	29	91
Familie	73	19	92
Saueflokk	85	7	92
Hundespann	86	5	91
Totalt	364	91	455

Tabell 4. Resultater fra eksperiment for adjektivet *klar*.



Figur 7. Resultater fra eksperiment for adjektivet *klar* fordelt på kategoriene menneske (*gruppe, band, familie*) og dyr (*saueflokk, hundespann*).

¹⁰ Pearson's Chi-kvadrat-test med Yates' kontinuitetskorreksjon (X-kvadrat = 33.182, df = 1) ga en p-verdi = 8.391e-09. Dette innebærer at de observerte forskjellene er svært signifikante, dvs. at det er svært lite sannsynlig at de beror på tilfeldigheter. Cramers V-verdi = 0,3, noe som er en moderat effektstørrelse.

7 Agentivitet – mer eksperimentdata

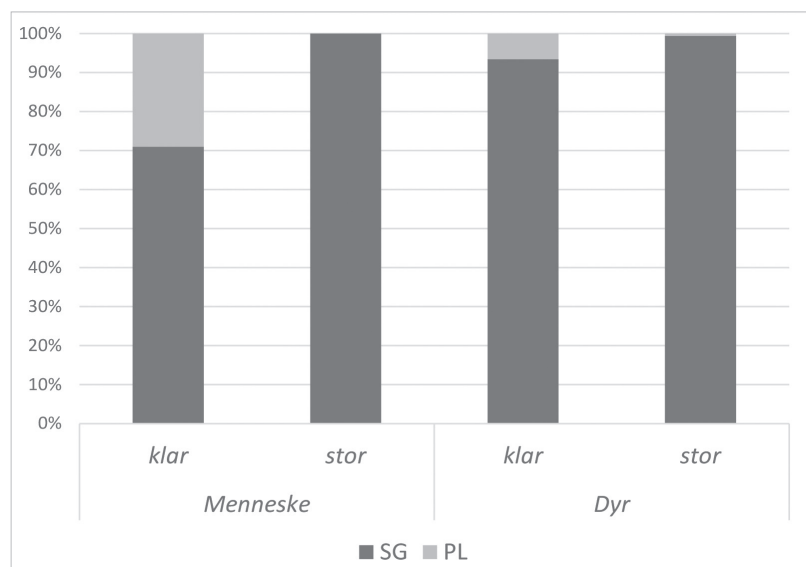
I avsnitt 4 nevnte vi at vi ønsket å finne ut om agentivitet påvirker sannsynligheten for flertallskongruens. Vår hypotese er at adjektiv som impliserer agentivitet, dvs. innebærer at subjektet skal utføre en handling, har større sjanse for å ha flertallskongruens enn adjektiv som ikke har agentiv betydning.

I eksperimentet sammenlignet vi setninger med adjektivet *klar*, der adjektivet innebærer agentivitet, med setninger med *stor*, der agentivitet ikke er implisert (jf. eksempel (7) og (8) i avsnitt 4). Resultatene oppsummeres i Tabell 5 og Figur 8. Som vi ser, er forskjellene tydelige. Mens *klar* ofte forekommer med flertallskongruens, særlig når subjektet refererer til mennesker, har vi kun ett eneste eksempel på at en deltager i eksperimentet godtok flertallskongruens for *stor*. Dette tyder på at hypotesen vår er riktig: agentivitet kan gjøre flertallskongruens sannsynlig.¹¹

	<i>klar</i>		<i>stor</i>	
	SG	PL	SG	PL
Gruppe	58	31	90	0
Band	62	29	90	0
Familie	73	19	90	0
Saueflokk	85	7	90	0
Hundespann	86	5	88	1
Totalt	364	91	448	1

Tabell 5. Resultater for adjektivene *klar* og *stor* i eksperimentet.

¹¹ En svakhet med eksperimentet må nevnes her. I eksperimentet er avstanden mellom subjekt og predikativ større i setningene med *klar* enn i setningene med *stor*. Som vi har sett i avsnitt 5, er det slik at større avstand øker sjansen for flertallskongruens. Imidlertid skulle vi ikke forvente oss en kategorisk forskjell mellom *klar* og *stor* hvis det bare var avstanden som spilte en rolle. Den kategoriske forskjellen vi observerer i Tabell 5, kan dermed ikke bare skyldes avstanden mellom subjekt og predikativ, og vi konkluderer derfor med at agentivitet er en relevant faktor som påvirker valget mellom entalls- og flertallskongruens. En anonym konsulent påpeker at eksemplene med *hundespann* og *stor* kan være problematiske, siden egenskapen 'stor' predikeres over kollektivet, men ikke over hver enkelt hund. Hundespannet har mange medlemmer, men hver enkelt hund behøver ikke være stor. Dette gjør trolig flertallskongruens mindre sannsynlig. Vi vil påpeke at denne betimelige innvendingen likevel ikke gjelder eksemplene med *gruppe* og *band*. Her er flertall mulig, som i følgende eksempel (fra internett) om den Los Angeles-baserte gruppen Flogging Molly: *Grappa er store i USA og Asia, og har hele verden som tumleplass*. Her predikeres egenskapen over hvert enkelt gruppemedlem (som oppfattes som "store musikere"), og samtidig gjelder også egenskapen for gruppen som helhet. Like fullt velger ingen av informantene flertall i eksperimentet.



Figur 8. Resultater for adjektivene *klar* og *stor* i eksperimentet for kategoriene menneske og dyr.

8 Sammenfatning

I denne artikkelen har vi undersøkt valget mellom entalls- og flertalls-kongruens i norske setninger med predikativt adjektiv og kollektiv som subjekt. Vi har analysert data fra to korpus statistisk og gjennomført en undersøkelse med informanter. Vårt bidrag kan oppsummeres som følger. For det første har våre empiriske undersøkelser vist at flertallskongruens er ganske vanlig. Samtidig er det stor variasjon; våre trebankdata har en større andel flertallskongruens enn vi finner i data fra NOWAC-korpuset (avsnitt 3). For det annet har vi vist at det er flere faktorer som påvirker valget mellom entall og flertall. Noen faktorer refererer til subjektet (12a), andre har med predikativet å gjøre (12b), mens en tredje type faktorer har med relasjonen mellom subjekt og predikativ å gjøre (12c):

- (12) a. Subjektets egenskaper:
- menneske >> dyr
 - Politisk parti >> Land, verdensdel >> Organisasjon, etat, institusjon >> Politi

b. Predikativets egenskaper:

- Adjektiv vs. partisipp: Adjektiv >> Partisipp
- Agentivitet: Agentivt >> ikke-agentivt
- c. Relasjonen mellom subjekt og predikativ:
- Avstand subjekt – predikativ: Stor >> Liten

Her betegner >> at det som står til venstre, har større sannsynlighet for å utløse flertallskongruens enn det som står til høyre. Vi undersøkte også om blocking hadde noen effekt, men fant ikke noe evidens for det i våre data.

Vårt fokus har vært empirisk og deskriptivt. Avslutningsvis kan vi likevel antyde noen teoretiske implikasjoner. Vi ser at animatumkategorien spiller en rolle. Kanskje kan våre funn knyttes til det mer generelle fenomenet individueringsgrad. En hypotese som kan forfølges i videre forskning, er at høy individueringsgrad gir stor sjanse for flertallskongruens. Kontrasten mellom *parti* (som har stor sannsynlighet for flertallskongruens) og *politi* som i mindre grad har flertallskongruens, kan tyde på at individueringsgrad er relevant, siden politiet opptrer i uniform og det er lite fokus på enkeltindivider, mens enkeltindivider får stort spillerom i politiske partier. Et annet funn som indikerer at individueringsgrad har betydning, er at agentivitet påvirker kongruensen. Agentivitet henger sammen med animathet, siden bare animata kan utføre kontrollerte handlinger, og dermed med individueringsgrad.

Et annet teoretisk poeng er at reell avstand, for eksempel målt i antall mellomliggende ord, har betydning for kongruens, og dermed for syntaks. Våre empiriske funn støtter her tidligere antagelser om norsk (Faarlund et al. 1997) og iakttagelser fra andre språk (Corbett 2006). At det er slik, kan forklares med utgangspunkt i prosessering. Ved større avstand mellom subjekt og predikativ kan taleren ha glemt hvilket ord som er subjekt, men likevel huske at det dreier seg om en flerhet av personer, og dermed velge flertallskongruens. Om denne forklaringen har noe for seg, har prosessering direkte konsekvenser for syntaktisk form.

Fra et typologisk synspunkt er mange av faktorene vi har pekt på i denne studien, ikke kontroversielle. Men vi har foretatt den første større empiriske og statistiske undersøkelsen av norsk adjektivkongruens ved kollektiver og dermed kastet nytt lys over et klassisk problem i norsk og allmenn språkvitenskap.

Litteraturliste

- Barlow, M. (1999): "Agreement as a Discourse Phenomenon". *Folia Linguistica* XXXIII, s. 187–211.
- Bybee, J.L. (2010): *Language, Usage and Cognition*, Cambridge University Press, Cambridge.
- Corbett, G.G. (2006): *Agreement*, Cambridge University Press, Cambridge.
- Dahl, Ö. (2000): "Animacy and the notion of semantic gender". B. Unterbeck et al. (red.): *Gender in Grammar and Cognition*, Mouton de Gruyter, Berlin/New York, s. 99–117.
- Denstad, I. (2021): *En av barna: Generalisering av felleskjønns determinativ*, MA-avhandling, Universitetet i Oslo. <https://www.duo.uio.no/handle/10852/87006>
- Dyvik, H.J.J. (2019): "Er vi fokusert eller fokuserte? Partisippenes indre kamp i bokmål", *NORGRAMTALL: Data om norsk fra NorGramBank*. <https://nogramtall.w.uib.no/2019/02/19/er-vi-fokusert-eller-fokuserte-partisippenes-indre-kamp-i-bokmal/>
- Faarlund, J.T., S. Lie & K.I. Vannebo (1997): *Norsk referansegrammatikk*, Universitetsforlaget, Oslo.
- Gries, S.Th. (2021): *Statistics for linguistics with R*, 3. reviderte og utvidede utgave, De Gruyter, Berlin.
- Hagåsen, L. (1992): "Sverige var intresserad och facket ska vara obunden – en typ av neutruminkongruens". G. Widmark et al. (red.): *Svenska i harmoni*, Hallgren & Fallgren, Uppsala, s. 60–75.
- Hansen, E. & L. Heltoft (2011): *Grammatik over det Danske Sprog II: Syntaktiske og semantiske helheder*, Det Danske Sprog- og Litteraturselskab, København.
- Knudsen, T. (1973): "Bøyningen av adjektiver som predikatsord". *Maal og Minne*, s. 29–41.
- Levin, M. (2001): *Agreement with collective nouns in English*, Lund University Press, Lund.
- Levin, M. (2006): "Collective nouns and language change". *English Language and Linguistics* 10, s. 321–343.
- Lie, S. (2003): *Innføring i norsk syntaks*, 5. utgave, Universitetsforlaget, Oslo.
- Rad, K. (2021): *Bortfall av flertallssuffiks ved norsk adjektivkongruens*, MA-avhandling, Universitetet i Oslo. <https://www.duo.uio.no/handle/10852/87139>
- Rosén, V., K. De Smedt, P. Meurer & H. Dyvik (2012): "An open infrastructure for advanced treebanking". J. Hajič et al. (red.): *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, European Language Resources Association (ELRA), Istanbul, s. 22–29.
- Sasse, H.-J. (1993): "Syntactic categories and subcategories". Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (red.): *Syntax: An International Handbook of Contemporary Research*, De Gruyter, Berlin, s. 646–686.
- Vinje, F.E. (2002): *Moderne norsk*, 5. utgave, Fagbokforlaget, Bergen.

- Western, A. (1921): *Norsk Riksmåls-grammatikk*, Aschehoug, Kristiania (Oslo).
- Åfarli, T.A. & Ø.A. Vangsnes (2021): "Formell og semantisk adjektivkongruens i norsk". K. Hagen, A. Hjelde, K. Stjernholm & Ø.A. Vangsnes (red.): *Bauta: Janne Bondi Johannessen in memoriam*, Oslo Studies in Language 11(2), s. 527–540.