

Katrine Rosendal Ehlers

Number-sensitive reflexive
pronouns in Danish

Optionality, microvariation, and cyclic change

Supervisors: Sten Vikner,
Henrik Jørgensen, Theresa Biberauer

PhD Thesis

Department of English

Aarhus University

December 2023

ISBN: 978-87-7507-564-5
DOI: 10.7146/aui.532

Abstract

English abstract

Danish, in common with the other Scandinavian languages but in contrast with e.g. English and German, has a possessive reflexive pronoun *sin* (with the inflectional variants *sit* and *sine*). *Sin* must have a subject antecedent within a particular binding domain, and this subject antecedent must be singular (or non-plural). This is an unusual demand in a cross-linguistic perspective: In the other Scandinavian languages and in earlier stages of Danish (before 1000 AD, at least), *sin* places no number restrictions on its antecedent. It is also an unusual demand in a linguistic theory-perspective: Various researchers (Burzio 1991 and Reuland 2011, among others) suggest that the defining trait of reflexive pronouns is precisely that they lack morphological content (number, gender, sometimes even person). With the Danish *sin*, we have a reflexive pronoun that can be specified as a 3rd person singular pronoun, which certainly seems like some amount of morphological content. However, this may not be a lasting trait. In modern Danish, it is not unusual to hear or read *sin* used with plural antecedents, and I suggest that *sin* may be in the process of losing its singular feature.

Danish having number-restricted reflexives is not a new feature of the language. It is attested in the earliest non-runic sources of Danish, going back to the 13th century. In the same period of time, Danish object reflexive *sig/sig selv* seems to have changed from patterning with *sin* in mainly allowing singular antecedents to the present situation where it patterns with the other Scandinavian languages in allowing both singular and plural antecedents. This is a relatively recent development, however, and it was finalized as recently as the 20th century in the spoken language (cf. Pedersen 2017). I hypothesize that *sin* may be going the same way as *sig* and changing (back) to allow both singular and plural antecedents. The change in *sin* lags behind because there has been less external pressure on speakers to acquire a system with a number-neutral *sin*.

I investigate the properties of *sin* (and *sig*) in modern and historical Danish in this thesis. I view the topic through a lens of generative grammar with a particular focus on microvariation. My thesis is a contribution to a field of research into reflexive pronouns within generative grammar which has become increasingly eclectic with the inclusion of more languages, more niggling details, and new theory-internal perspectives. My thesis is not aimed at arguing e.g. for or

against an agreement-based or movement-based approach to binding. I offer an explorative, empirically based perspective on diachronic and synchronic microvariation within a single language. I hope to offer some insight into an interesting case of cyclic change and linguistic optionality.

In the first chapter of the thesis, I introduce the theories and technical machinery behind the approach to linguistics that I adopt in this thesis. I situate my thesis within a framework of generative grammar, and within a framework of variation and change with a special focus on microvariation. I sketch the development in the research on reflexives from the relatively unified Government and Binding approach of the 1980s to the current, rather less unified range of approaches.

In the second chapter of the thesis, I narrow the focus and look specifically at Danish and the Danish reflexive system. I present an analysis of the reflexive systems in standard Danish (an analysis which first saw the light of day in Vikner 1985), and I outline some of the ways that the Jutlandic dialects of Danish differ from the standard. The ways that reflexive use in the Jutlandic dialects differ from standard Danish are somewhat stigmatised and I hypothesize that this could lead to speakers hyper-correcting and using *sin* with plural antecedents more frequently. Finally, I present the use of *sin* in standard Danish that I call number restricted.

The third chapter of the thesis is a diachronic study of the use of *sin* and *sig* in Danish over the last millennium. There is textual evidence from runic stones from Denmark that *sin* was used with both plural and singular antecedents before 1000 AD. There is no direct evidence that this was also the case for *sig*. This use is, however, a direct continuation from the Common Germanic stage where both *sig* and *sin* were used with antecedents of all numbers. By the 13th century, where the earliest Danish manuscripts are from, the use of *sin* and *sig* had changed so that both forms are primarily used with singular antecedents and their non-reflexive counterparts (*deres* and *dem*) are used with plural antecedents. Texts in the following 6-700 years shows different developmental trajectories for *sig* and *sin*. *Sig* becomes the predominant form in the written language with both singular and plural antecedents after the Reformation (with a great deal of variation). This is probably due to influence from German, and locally bound *dem* remains frequent in the spoken language until the early 20th century. Locally bound *dem* must be considered a very marginal form in the modern language outside of the contexts where there is structurally conditioned optionality between *sig* and *dem*. *Sin* stays limited to mainly singular antecedents, although

examples of plural antecedent *sin* can be found sporadically all through the period.

The fourth chapter of the thesis is a corpus study. I investigate the occurrence and distribution of plural antecedent *sin* in KorpusDK, which is a written corpus with texts from 1983 to 2002. I found 1218 examples of plural antecedent *sin* in the corpus out of 188,585 instances of *sin* with any kinds of antecedents. Overall, plural antecedent *sin* occurs relatively more often with inanimate antecedents, with full noun antecedents (rather than pronoun antecedents), with coordinated antecedents with singular coordinands, and in complex clauses (particularly a type of clause where a partitive is modified by a relative clause). The data does not support suggestions from the literature that plural antecedent *sin* is better or more likely to occur within distributive contexts.

The fifth chapter is an acceptability judgment experiment with a sample of 550 young Danish speakers from various Danish gymnasiums (upper secondary schools). I find in the study that plural antecedent *sin* is more acceptable if it is contained with a plural nominal, if the antecedent is inanimate, and if the antecedent is a full noun. I find a little statistically significant evidence of sociolinguistically conditioned variation but the most striking result here is that the effects of e.g. region or gender are not very striking. Participants from all over the country rated plural antecedent *sin* as quite acceptable, just like participants from all over the country rated plural antecedent *sin* as quite unacceptable. I believe that this could show a linguistic state of ongoing change where some speakers have plural antecedent *sin* as part of their grammars, and some speakers do not. Whether or not a speaker accepts plural antecedent *sin* seems more dependent on other factors in that speaker's own grammar than on the sociolinguistic factors that I coded the data for. Specifically, those speakers who rate other non-standard uses of reflexives (such as using locally bound singular possessive pronouns) higher also tend to rate plural antecedent *sin* higher.

In the final chapter of the thesis I sum up the results from the previous chapters, suggest areas of improvement and topics for follow-up studies, and speculate on the factors that could have an impact on this possible ongoing change. I suggest that some of these factors are the change in number-restriction on *sig*, hyper-correction driven by other stigmatised features of the Danish reflexive system, semantic agreement, and – very speculatively – the great deal of optionality or variation in number agreement on Danish adjectives, which could in principle also work as evidence for acquiring a reflexive system with less number-sensitivity on *sin*.

Dansk resumé

Dansk har ligesom de andre skandinaviske lande, og i modsætning til f.eks. engelsk og tysk, et refleksivt possessivpronomen *sin*. *Sin* skal have en subjektsantecedent der er tilpas tæt på, og denne antecedent skal desuden være i tredje person ental (eller i hvert fald ikke teknisk set i flertal). Dette er usædvanligt på tværs af sprog: I de andre skandinaviske sprog og i tidligere stadier af dansk (før år 1000) er der ingen tilsvarende krav om numerus af antecedenten. Det er også et usædvanligt krav i et teoriinternt perspektiv: Forskellige forskere (Burzio 1991 og Reuland 2011, for at nævne to) har foreslået at refleksive pronominer kan defineres ved netop at mangle morfologisk indhold (numerus, genus, måske også person). Med det danske *sin* har vi et refleksivt pronomen der kan beskrives som tredje person ental, hvilket netop må tolkes som en form for morfologisk indhold. Dét ser dog ud til at være i forandring i moderne dansk. Det er ikke usædvanligt at høre eller læse eksempler på *sin* med flertalsantecedent, og jeg foreslår at *sin* kan være ved at miste sit numerus-krav.

Det er ikke nyt at de danske refleksiver stiller krav til numerus af deres antecedent. Helt tilbage til de tidligste håndskrifter fra det 13. århundrede findes der eksempler på at både *sin* og det ikke-possessive reflektiv *sig* overvejende optræder med entalsantecedenter. I de seneste 800 år har *sin* og *sig* udviklet sig i forskellige retninger. *Sig* har genvundet muligheden for at optræde med både entals- og flertalsantecedenter, mens *sin* har bibeholdt sit krav om primært at optræde med entalsantecedenter. Ændringen i *sig* blev endeligt gennemført i starten af det 20. århundrede i talesproget (jf. Pedersen 2017). Min hypotese er at *sin* følger *sig* i denne udvikling, men at forandringen i *sin* sakker bagud fordi der har været mindre anledning for danske talere til at tilegne sig et sprog uden numerus-restriktioner på *sin*.

Jeg undersøger egenskaberne ved *sin* (og *sig*) i moderne og historisk dansk i denne afhandling. Jeg tilgår emnet med en baggrund i generativ grammatik med et specifikt fokus på mikrovariation. Min afhandling skriver sig ind i forskningstradition inden for generativ grammatik hvor udforskningen af refleksive pronominer er gået fra at være et relativt samlet felt til en nuværende forskningsvirkelighed der peger i mange forskellige retninger i takt med at man har inddraget flere sprog, flere fine detaljer, og nye teoriinterne perspektiver. Min afhandling er ikke skrevet for f.eks. at argumentere for eller imod en *agreement*-baseret eller en *movement*-baseret tilgang til refleksiverne. Det jeg kommer med, derimod, er et eksplorativt, empirisk funderet perspektiv på diakron og synkron mikrovaria-

tion inden for et enkelt sprog. Jeg håber på derved at kunne bidrage med indsigt i hvad der er muligt inden for cyklisk sprogforandring og valgfrihed inden for et lingvistisk paradigme.

I afhandlingens første kapitel introducerer jeg de teorier og tekniske værktøjer som ligger bag den tilgang til lingvistik som jeg bruger i afhandlingen. Jeg placerer min afhandling inden for en ramme af generativ grammatik, og inden for en ramme af sproglig variation og sprogforandring med et specifikt fokus på mikrovariation. Herunder beskriver jeg udviklingen i den generative forskning i reflexiver fra 1980'ernes umiddelbart nogenlunde samlede *Government and Binding*-tilgang til det nuværende, noget mindre samlede forskningsfelt.

I afhandlingens andet kapitel indsnævrer jeg mit fokus til dansk og det danske reflexivsystem. Jeg præsenterer en analyse af reflexivsystemet i standardsproget (en analyse som først så dagens lys i Vikner 1985), og beskriver hvordan brugen af *sin* specielt i visse jyske dialekter afviger fra standardbrugen. Disse afvigelser er hyppige men samtidig noget stigmatiserede i dansk og jeg fremsætter en hypotese om at denne stigma kan få nogen talere til at hyper-korrigere deres brug af *sin* sådan at de også i højere grad bruger *sin* med flertalsantecedenter. Som afslutning på kapitlet præsenterer jeg den brug af *sin* i dansk som jeg beskriver som havende et numerus-krav.

Afhandlingens tredje kapitel er en diakron undersøgelse af brugen af *sin* og *sig* i dansk i de sidste tusind år. Der er eksempler på *sin* med både entals- og flertalsantecedenter på runesten fra før år 1000 fra det daværende danske område. Der er ingen tilsvarende eksempler på *sig* med flertalsantecedenter, men givet at ord tilsvarende *sig* optrådte i fællesgermansk med både entals- og flertalsantecedenter, ville brugen af *sig* i dansk på samme måde være en logisk følge derfra. I danske håndskrifter fra det 13. århundrede har brugen af *sin* og *sig* ændret sig, sådan at begge former primært bliver brugt med entalsantecedenter og deres ikke-refleksive modstykker (*deres* og *dem*) bliver brugt med flertalsantecedenter. I tekster fra de efterfølgende 6-700 år udvikler brugen af *sin* og *sig* sig forskelligt. I tiden efter Reformationen bliver *sig* den mest brugte form med både entals- og flertalsantecedenter (med en hel del variation i brugen). Denne forandring blev formentlig hjulpet på vej af sproglig påvirkning fra tysk. *Dem* frem for *sig* med flertalsantecedenter er blevet i talesproget frem til det tidlige 20. århundrede, og der kan stadig findes enkelte eksempler på det i det moderne talesprogskorpus LANCHART (ud over i de situationer hvor der er strukturelt baseret valgfrihed mellem *dem* og *sig*). *Sin* bliver anvendt langt overvejende med entalsantecedenter fra 1200 til nu, omend der kan findes mere eller mindre sporadiske eksempler

på *sin* med flertalsantecedenter gennem hele perioden.

Afhandlingens fjerde kapitel er et korpusstudie. Jeg undersøger forekomsten og fordelingen af *sin* med flertalsantecedenter i KorpusDK, som er et dansk skriftsprogskorpus med tekster fra 1983 til 2002. Jeg har fundet 1218 eksempler på *sin* med flertalsantecedenter ud af i alt 188.585 eksempler på *sin* i det samlede korpus. *Sin* forekommer overordnet set oftere med flertalsantecedenter når antecedenten ikke er animat, når antecedenten er et fuldt nominal (frem for et pronomen), når antecedenten består af et eller flere sideordnede nominaler, og når *sin* er indeholdt i en underordnet sætning (relativsætning eller infinitivsætning) i forhold til sin antecedent. Tallene fra KorpusDK understøtter ikke at *sin* skulle være mere hyppigt forekommende med flertalsantecedenter hvis den indgår i en distributiv kontekst.

Det femte kapitel i afhandlingen er et acceptabilitetsstudie med 550 forsøgspersoner som på daværende tidspunkt var elever på forskellige danske gymnasier. Resultaterne fra undersøgelsen viser at *sin* med flertalsantecedent bliver bedømt som mere acceptabel hvis *sin* er indeholdt i en nominal der selv er flertal (dvs. hvis formen er *sine*), hvis antecedenten ikke er animat, og hvis antecedenten er et fuldt nominal (frem for et pronomen). Jeg finder enkelte statistisk signifikante eksempler på at (noget af) variationen i acceptabiliteten af *sin* med flertalsantecedent kan forklares sociolingvistisk. Overordnet set er det dog mest slående at f.eks. deltagernes region eller køn ikke lader til at have den store indflydelse på hvordan de vurderer *sin* med flertalsantecedent. Der er forsøgsdeltagere fra alle egne af landet der vurderer *sin* med flertalsantecedent som helt fint, og tilsvarende forsøgsdeltagere fra alle egne af landet der vurderer *sin* med flertalsantecedent som helt umuligt. Jeg tolker dét resultat som et tegn på netop variation og måske en igangværende sproglig forandring på tværs af talere hvor nogen talere har en grammatik som tillader *sin* med alle slags antecedenter og nogen talere har en grammatik som kun tillader *sin* med entalsantecedenter. Hvorvidt en taler har den ene eller den anden grammatik lader til i højere grad at være betinget af andre faktorer i talerens sprog frem for af de sociolingvistiske faktorer som jeg har kodet mit data efter. Specielt finder jeg at de talere som også giver højere vurderinger til reflexivbrug der falder uden for standardformerne (f.eks. brugen af lokalt bundne entalspossessiver), også giver højere vurderinger af *sin* med flertalsantecedent.

I afhandlingens sidste kapitel opsummerer jeg resultaterne fra de foregående kapitler. Jeg foreslår områder hvor mine undersøgelser og forsøgsdesigns kunne forbedres og emner som kunne være oplagte som opfølgende studier. Jeg giver

også mine bud på nogle faktorer der kunne spille ind i den mulige sprogforandring i brugen af *sin* med flertalsantecedenter. Herunder foreslår jeg at udviklingen i brugen af *sig* kunne medføre en forandring i brugen af *sin*, at hyppigere brug af *sin* med flertalsantecedent kunne opstå på grund af hyperkorrektion, og – meget spekulativt – at der findes en hel del variation i adjektivernes numeruskongruens på dansk som i princippet også kunne påvirke førstesprogtaleres tilegnelse af et system med mindre numerus-restriktion i refleksiverne.

Acknowledgements

This thesis is the labour of almost 9 years that have brought with them a very wide assortment of nice and less nice experiences. The work on this thesis has been a relatively stable constant through an increasingly weird sea of boardgames, barbells, late-night cigarette runs and karaoke, a blasphemous wedding, autumnal poets and Jonah Ryan, an early-November trip to New York to celebrate the 2016 election of Hillary Clinton (...), moving vans, a pandemic, three kids, and everything else between the lines.

I am very thankful to my main PhD advisor Sten Vikner, who has been precisely the advisor that I needed throughout this ever so slightly prolonged project. Thank you for answering e-mails very thoroughly in the middle of the night, for thorough and constructive criticism and advice, for a hands-off approach and faith in my ability to get there (here!) in the end despite various *svinkeærinder*, and for being an all-around supportive across-the-hall co-noisy Christmas ornament have-er.

I am very thankful to my two PhD co-advisors, Henrik Jørgensen and Theresa Biberauer, for help with refining my data collection, for excitement about the project, for really specialized knowledge about the state of the Jutlandic reflexives, and for an excellent class on emergent parameters and microvariation at the LOT Summer School.

I very thankful to Anne Mette Nyvad for not only letting me join your project (and for listening to all of my school-related woes), but also for letting me finish my thesis *on the clock*. Without you, I certainly would not have finished the thesis on this side of ~~ever~~ Christmas.

I very gratefully acknowledge the help from Prof. Søren Feodor Nielsen (CBS) with creating an initial search script for the offline version of KorpusDK. Without that script, that entire project would not have got off the ground.

A number of people helped with translations of reflexive sentences into a wide array of languages (and/or with establishing contact to those people who could translate): Rob Truswell, Ivo Youmerski, Hanna H. Poulsen, Steingrímur Óttar Steingríms-

son, Rui Xu, Sidsel Rasmussen, Michaela Hejná, Pramay Rai, Keisuke Yoshimoto, Caroline Heycock, Junji Hamamatsu, Chung-hye Han, Joanna Panowicz, Amra Sabotic, and Zehra Yücel. I have not been able to fit in that cross-linguistic investigation in the thesis but I hope to do it justice at some later point. Thank you very much!

Rob Truswell made me feel welcome far from home in Edinburgh and put a nice, pragmatic spin on my far-reaching theorizing. Thank you very much!

Tanya Karoli Christensen and Torben Juel Jensen made me welcome in Copenhagen and helped me greatly with access to the LANCHART database. Those data also did not end up playing a large role in the thesis but will hopefully see the light of day in another context. Thank you very much!

Thank you to Jørg Asmussen for giving me access to the offline version of KorpusDK and for explaining how it works.

A great number of gymnasium teachers contributed their time and in many cases their students(!) to my project. Thank you to Just, Nanna Smith Vinther, Mette Solsø Nielsen, Erik Hundrup Rasmussen, Gitte Lautrup, Monica Christensen Birkler, Carina K. Ørskov, Marie Trier, William Bjarnø, Mads Dam, Alice Sahinkuye, Ann-Sophie Spaabæk, Nanna Helene Jørgensen, Helle Elgaard, Charlotte Frølich, Pernille Jensen, and Niels Christian Krogh Hansen. (And my apologies if I have forgotten any names here!) I am very grateful to the hundreds of students who participated in the study more or less voluntarily. Your input is greatly appreciated, and I hope that you did not lose too many brain cells from reading through that many examples of outlandish sentences. I am also very grateful to the hundreds of people who participated in the pilot study. I hope to do your participation justice with a proper write-up of those data in another publication. A last thank you to my mother-in-law who helped with collecting data from your colleagues and church-goers – that is another bit of data that I hope to do justice at some point in the future.

I would like to thank my linguist colleagues on the third floor and beyond – past and present – for lunch discussions, interesting talks and classes, book recommendations, cakes, (lenience towards) Christmas ornaments, and good company over the last nine years: Anne Mette Nyvad, Birgitte Poulsen, Camilla Søballe Horslund Marqvorsen, Clara Ulrich Eggers, Hanne Wachter Kjærgaard, Helle Kaalund Tornbo, Johanna Wood, Johannes Kizach, Jonas Villumsen, Julie Maria Rohde, Ken Ramshøj Christensen, Krestina Vendelbo Christensen, Mette Hjortshøj Sørensen, Michaela Hejná, Ocke-Schwen Bohn, Sidsel Rasmussen (my great office-mate), Sten Vikner, and William

Miki Thorsen. (Apologies to Sten for alphabetizing by first name. And I really hope I did not forget anybody here!)

A final but very large and heartfelt thank you goes out to my friends (if you feel slightly neglected at the expense of this project at this point, I would understand!) and family. Mutti/mormor og Claus/manden der siger ord og Marie Louise/Wiewie. Lasse og Slippe. Anne og Jens Ole, og Peter som jeg aldrig nåede at møde. Thomas Ehlers. Mormor (som så starten men ikke slutningen – jeg savner dig!) og morfar i Hygum. Sigrid, Ask og Magne. Martin – jeg elsker dig.

Any errors, omissions, superficial analyses, and far too frequent use of (really important) comments in parentheses are entirely my own fault.

Contents

English abstract	i
Dansk resumé	iv
Acknowledgements	viii
Contents	xi
1 Introduction: Generative grammar, variation, and binding theory	1
1.1 Variation and change in a generative framework	6
1.1.1 Language change is driven by language acquisition	7
1.2 Generative tree structures	8
1.2.1 CP – IP – VP	8
1.2.2 Movement	10
1.2.3 Arguments, adjuncts, and theta-role assignment	13
1.2.4 Case-assignment	16
1.2.5 Subjects – overt and non-overt	17
1.2.6 DP: Nominal structure	19
1.3 Theoretical framework: Principles and Parameters	25
1.3.1 A general generative view of language	25
1.3.2 Government and Binding – rich UG and levels of representation	28
1.3.3 Minimalist framework and terminology	30
1.4 Binding theories	35
1.4.1 Traditional binding theory a la Chomsky (<i>Lectures on govern-</i> <i>ment and binding</i> (LGB) 1981)	36
1.4.2 Newer approaches to binding	43
1.4.3 Summary	52
2 Danish language and Danish reflexives	54

2.1	The <i>sig</i> and <i>sin</i> reflexives in the Germanic languages	54
2.1.1	Reflexive possessive <i>sin</i> in the Germanic languages	55
2.1.2	Reflexive <i>sig</i> in Danish and other Germanic languages	59
2.2	A general introduction to Danish grammar	60
2.2.1	Clause structure and verbal system	60
2.2.2	Nominal system	62
2.3	The Danish reflexive system	69
2.3.1	+/- <i>selv</i> (1st and 2nd person examples)	69
2.3.2	± <i>sig</i> (<i>sin</i> examples)	71
2.3.3	Full system (<i>sig/sig selv</i> examples)	73
2.3.4	Non-local binding by inanimate antecedents	76
2.3.5	Locally bound <i>sig</i> : <i>skamme, vaske, brænde</i>	80
2.4	Local binding of singular possessive non-reflexives - Han tog hans hat	82
2.5	Plural antecedent <i>sin</i>	88
2.5.1	The basic structural position of <i>sin</i> in the clause	88
2.5.2	Number restriction on Danish <i>sin</i> - the basics	90
2.5.3	Binding, feature matching, and feature impoverishment	94
2.5.4	Agreement variation also regularly present in the standard language	101
2.5.5	Distributivity	117
3	Reflexive pronouns with plural antecedents from runes to LANCHART	123
3.1	Introduction	123
3.2	The reflexives in Runic Danish and Middle Danish	126
3.2.1	Runic Danish (800-1100)	127
3.2.2	Middle Danish (1100-1500)	130
3.3	The reflexives in Older Modern Danish	139
3.3.1	<i>sig/dem</i> in Chr. III's Bible (1550)	141
3.3.2	<i>Sin</i> and <i>deres</i> from Chr. Pedersen to H.C. Andersen (1500-1900)	148
3.3.3	<i>sig/dem</i> in the dialects in younger modern Danish (1750-now) .	156
3.4	LANCHART: Present-day spoken Danish	158
3.5	Conclusion	163
4	<i>Sin</i> with plural antecedents in KorpusDK	165
4.1	Description of the online version of KorpusDK	167

4.2	Description of the offline version of KorpusDK	168
4.3	Scripts and technical considerations	173
4.4	Search methodology for plural antecedent <i>sin</i>	174
4.5	Data sets for comparison: plural antecedent <i>deres</i> , any antecedent <i>sin</i> (random 5 %)	181
4.5.1	Search strategy for plural antecedent <i>deres</i>	181
4.5.2	Search strategy for all <i>sin</i>	182
4.5.3	Number validation for plural antecedent <i>sin</i> : Checking a ran- dom 5 % from each file	182
4.6	Inclusion and exclusion criteria for examples with plural antecedent <i>sin</i>	182
4.6.1	Exclusion criteria	183
4.7	Coding	190
4.8	Results	196
4.8.1	Type of subject (full noun or pronoun)	196
4.8.2	Number of the nominal that contains the reflexive	198
4.8.3	Sentence type (simple or complex)	203
4.8.4	Argument or adjunct	208
4.8.5	Animate or inanimate	209
4.8.6	Antecedents in coordinations	212
4.8.7	Distributivity	214
4.8.8	Distance between reflexive and antecedent	220
4.9	Discussion	223
4.9.1	RQ1: How frequently do instances of plural antecedent <i>sin</i> oc- cur in the corpus?	224
4.9.2	RQ2: Do the instances of plural antecedent <i>sin</i> pattern similarly to regular instances of plural antecedent <i>deres</i> ? or <i>sin</i> with any kind of antecedent?	224
4.9.3	RQ3: Which syntactic factors favour plural antecedent <i>sin</i> ? . . .	229
4.9.4	RQ4: Which semantic factors favour plural antecedent <i>sin</i> ? . . .	229
4.9.5	Final note	229
5	Acceptability judgment data	230
5.1	Introduction	230
5.2	Experimental design and methodology	231
5.2.1	Distribution of the questionnaire	232

5.2.2	Materials	234
5.3	The final data set	244
5.3.1	Background information provided by the study participants . . .	246
5.3.2	The full data set, background information summarized	254
5.3.3	Data exclusion criteria and outliers	255
5.4	Statistical models and model control	259
5.4.1	Research question 1: Statistical model and model control	260
5.4.2	Research question 2: Statistical model and model control	263
5.4.3	Research questions 3 and 4: Statistical model and model control	266
5.5	Results and discussion	270
5.5.1	Research question 1: Paired t-test	270
5.5.2	Research question 2: ANOVA	272
5.5.3	Research questions 3 and 4	282
5.5.4	Some remarks on the statistics and design of the study	303
5.6	Conclusion	305
6	Conclusion	308
	Bibliography	314
A	Instructions given to participants and teachers about the acceptability study	329
A.1	Instructions given to participants	329
A.1.1	Danish original	329
A.1.2	English translation (only used in this appendix)	330
A.2	Instructions given to the teachers who run the experiment in the class- room	330
A.2.1	Danish original	330
A.2.2	English translation (only used in this appendix)	333
B	Scripts	336
B.1	R scripts	336
B.2	Python scripts	339
C	Acceptability judgment test items	341
C.1	Pilot study	341
C.2	Main study	345

C.2.1	All 32 primary test items	358
C.3	All 30 secondary sentences with ratings	391
D	Examples of plural antecedent <i>sin</i> from e.g. news media – a very informal collection	395

Chapter 1

Introduction: Generative grammar, variation, and binding theory

The possessive reflexive *sin* in modern Danish presumably only allows singular subjects, as opposed to all the other Scandinavian languages where *sin* may take both singular and plural subjects. This is exemplified in the contrast between Danish in (1a) and Swedish in (1b) where Swedish allows *sin* both with singular *hun* (En. *she*) and plural *de* (En. *they*), but Danish only allows *sin* with the singular subject. Danish uses the possessive third person plural *deres* instead of the reflexive when the subject is plural. (This is indicated by subscripts: a * indicates ungrammaticality. The elements that refer to the same entity are indicated with **boldface** and matching indices: $X_1 - X_1$.)¹

- (1) a. **Hun**₁ / **de**₂ elsker **sin**_{1/*2} nye bil. (Danish)
she /they love REFL's new car
'She/they love(s) her/their new car.'
- b. **Hon**₁ / **de**₂ älskar **sin**_{1/2} nya bil. (Swedish)
she /they love REFL's new car
'She/they love(s) her/their new car.'

This is not a new feature of Danish and it is attested in the earliest non-runic sources of Danish, going back to the 13th century (this is the topic of chapter 3 of this thesis). This is interesting as a linguistic phenomenon in and of itself as it exem-

¹I follow the Leipzig Glossing Rules (<https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>) and their standard abbreviations. One abbreviation that I use that is not included in the list of standard abbreviations is C for COMMON gender which is one of the two grammatical genders in Danish. The other gender is N, NEUTER.

plifies an internally inconsistent system that Danish speakers nonetheless succeed in acquiring and maintaining over a long period of time. On top of this, the Danish system has been claimed to be undergoing a linguistic change that would bring it closer to the other Scandinavian languages. This possible change is evident in the many instances of speakers using *sin* with plural antecedents against the norm of prescriptively correct Danish. A recent example is from an official campaign from the Danish governmental body Sikkerhedsstyrelsen (the Danish Safety Technology Authority) about safety and fireworks. They write that "Alle har en idiot i sit liv" which in prescriptively correct Danish should have been "Alle har en idiot i deres liv" (En. *Everyone has an idiot in their life*).



Figure 1.1: Screenshot from X (previously Twitter) from the Danish safety authority Sikkerhedsstyrelsen. They use *sin* with the plural antecedent *alle* in the official communication about their 2022 firework safety campaign.

Plural antecedent *sin* in Danish is mentioned here in there in the literature but mainly as a curiosity or as one among several factors in a description of the Danish reflexive system as a whole. Rask (1820) argues that *sin* in Danish really should be used with plural as well as singular subjects (and then proceeds to do so himself in his texts). Falk and Torp (1900) argue that *sin* was restricted to singular antecedents around the time of the Reformation (but see chapter 3 for an argument against this claim). Mikkelsen (1911) gives several examples of plural antecedent *sin* from the literature. Diderichsen (1939) investigates the topic in greater detail and I work with his data in chapter 3. A. Hansen (1965, 115) describes that *deres* is used as the bound form (rather than *sin*) when the subject is plural. He also adds that *sin* is sometimes used if the plural antecedent is reminiscent of or represents something that could be viewed a singular. Haugen (1976) seems to follow Falk & Torp in claiming that Danish diverged from the other Scandinavian languages in the use of *sin* and *deres* in the 15th century. E. Hansen (1993, 179) notes that the use of *sin* with plural antecedents seems to be increasing greatly in modern Danish and speculates that it could be regaining the option of occurring with antecedents of both singular and plural number. Juel Jensen (2009b) includes plural antecedent *sin* in a corpus study of spoken Danish and finds that 7 % of the examples with plural antecedents have *sin* as the bound form (rather than *deres*) and also that his data does not back up the hypothesis that *sin* is increasing in use with plural antecedents over time. Hansen and Heltoft (2011, 590) provide examples showing how the use of *deres* (as an alternative to *sin*) goes back at least to sources from the 13th century, and that *sin* can indeed be found with plural antecedents in modern Danish texts.

I investigate some of the properties and distribution of *sin* with plural subjects (antecedents) in modern Danish in this thesis. I compare the possessive reflexive *sin* to its non-possessive counterpart *sig* which, as I argue in chapter 3, has undergone a very similar change recently, going from a stage where it mainly allowed non-plural antecedents to the present stage where *sig* allows plural and singular antecedents for (almost) all speakers. I sum up the diachronic developments for *sin* and *sig* in figure 1.2. The natural question, then, is whether we will see a future stage where both *sin* and *sig* have regained the option of occurring with both singular and plural antecedents.

I will not be able to say for sure that this is where the language is heading. I do, however, find a number of morphosyntactic and semantic factors that favour the

	Singular antecedent	Plural antecedent
Danish, before 1000 AD antecedent 3rd person, all genders	sig sin	sig sin
Danish, 1200(?)–1900(?) AD antecedent 3rd person, all genders	sig sin	dem deres
(Standard) Danish, 2023 antecedent 3rd person, all genders	sig sin	sig deres

Figure 1.2: The very short summary of how *sin* and *sig* have developed over the last millennium in Danish.

use of *sin* with plural subjects and I do find that speakers from all over the country are fairly happy to accept these types of sentences. I also, informally, find that plural antecedent *sin* is used frequently both in informal and formal communication (see appendix D for a selection from the last five years), and that both linguists and laypersons who I have shown these examples of plural antecedent *sin* do not immediately (or at all) identify it as grammatically strange. Overall, I find it reasonable to say that *sin* with plural antecedents is an emergent form in Danish and that we clearly have some amount of morphosyntactically or semantically conditioned optionality between using *deres* or *sin* in Danish with plural antecedents. Syntactic change is a slow process and the attested variation that can be found in the modern language could simply remain variation and not become a language-wide change. It also could go the same way that *sig* did, given enough time. Time will tell.

Outline of the chapters in the thesis

This first chapter is an introduction to the framework, lines of thought, and concepts of generative grammar which is the approach to language that I work within. I initially outline how I approach (morphosyntactic) variation and change within this framework. The subsequent section is an introduction to the trees that are typically used to analyze morphosyntactic structures within generative grammar. Following this introduction, I take a deeper dive into some of the guiding principles behind generative grammar: Competence and performance, Universal Grammar, the framework Princi-

ples and Parameters, and Minimalism, its (perhaps) leaner sibling. The last part of the first chapter is a review of some of the theoretical approaches that have been used to capture the fact that sometimes, two items in a clause refer to the same entity: **She**₁ buys **herself**₁ a Christmas present.

The second chapter in the thesis, chapter 2 which starts on page 54, is about the Danish language and the Danish reflexive system in particular. I initially situate the Danish reflexives within the linguistic history of the Germanic languages. The following section is a general introduction to Danish grammar where I outline the clausal, verbal, and nominal systems. Danish is a rather lightly inflected language in general and it is almost exclusively in the nominal system of Danish that we find the interesting parts of Danish morphology. I follow this grammatical outline with a review of the Danish reflexive system, which is couched in an analysis first suggested in Vikner 1985. The last two sections of chapter 2 go into detail with some of the variation that we find in the use of Danish reflexives. The first topic is the widespread usage of locally bound non-reflexive pronouns in Danish, even with singular antecedents. This is a fairly stigmatised feature of Danish, and one which I hypothesize could also play into the variation that I find with plural antecedent *sin*. The second topic is plural antecedent *sin* in Danish, including where I place *sin* in a generative tree, a discussion on whether Danish *sin* is best characterized as specified as singular, and a number of instances of agreement variation where variation in *sin* usage (almost) mirrors variation in adjectival inflection.

Chapter 3, which starts on page 123 is a diachronic study on the use of the Danish reflexive *sig* and possessive reflexive *sin* from the earliest Runic sources of Danish to the 20th century. I build on previous findings from other researchers to support the argument that *sin* and *sig* have been restricted to (mainly, with a certain degree of variation particularly with *sig*) singular antecedents at least for the last 800 years. *Sig* seems to have changed cyclically and speakers have fully adopted a system where *sig* now appears with both singular and plural antecedents. *Sin* is still lagging behind but the argument can also be made that Danish speakers have had less reason to reanalyze *sin*.

Chapter 4, which starts on page 165, is a large-scale empirical study of the occurrence of *sin* with plural antecedents in the 56 million word Danish written corpus KorpusDK. I have found almost every instance of plural antecedent *sin* in the corpus. The chapter contains a detailed walk-through of the process towards finding these –

fairly sparse – examples first in the online version of the corpus and subsequently in an offline version. This final approach also required an amount of programming in R and Python in order to produce a useable search interface for the large amounts of text data. I compare the distribution of plural antecedent *sin* in the corpus to the distribution of *sin* in general and bound *deres*. This allows me to pinpoint the areas where plural antecedent *sin* is similar (or dissimilar) to the standard forms.

Chapter 5, which starts on page 230, is an acceptability judgment study on the use of plural antecedent *sin* in a sample of more than 500 young speakers of Danish. I used the results from the corpus study in chapter 4 to build the acceptability judgment study. The study complements the corpus study in approaching the same grammatical factors in a different, more experimentally controlled way while supplementing it with access to sociolinguistic factors, which are not available in the corpus.

The final chapter, chapter 6 from page 308 and onwards, is a conclusion that sums up the results from the thesis and outlines the many roads that could be taken from here. The appendices that follow the concluding chapter both show some of the scripts that I have written (with and without assistance) for the project, as well as the full list of experimental items from the acceptability judgment study.

1.1 Variation and change in a generative framework

There is no such thing as syntactic change, and this is a book about it.

(Biberauer and Walkden 2015, 1)

Before I delve into the specifics of syntactic change and variation, I need to establish the machinery needed for talking about change and variation within a theoretical setting where it makes sense to write a book about a topic which there presumably is "no such thing as". The quote is, of course, intended as a joke of sorts but then again not completely. One issue is that the object of study within generative grammar is (intended to be) the competence, the internalized grammar (see page 25), of individual speakers and that convincing arguments have been made to the effect that a person's (syntactic) competence largely does not change after the earliest years of first language acquisition. Rather, the way that syntax changes is through change in other domains, such as in the lexicon (word inventory), the phonology, or the semantics of a language ((Biberauer and Walkden 2015, 3-5). This insight has set a change in motion in terms of what is the interesting subject of study for variation and change

within generative grammar: The tradition of the 1980's and 1990's had researchers looking for change and variation in the *macroparameters* or *mesoparameters* of languages, e.g. change in the word order of a language compared to another related language, or an earlier stage of the same language (Biberauer and Roberts 2017). The shift towards viewing change as dependent on the (functional, featural inventory of the) lexicon (dubbed the Borer-Chomsky Conjecture by Baker 2008) has inspired research into the *micro-* or even *nanoparameters* of languages. These are parameters associated with a limited set of functional heads (or as little as one linguistic item, in the case of nanoparameters). Microparameters are also more susceptible to change precisely because a change in a microparameter is not as far-reaching as a change in a macro- or mesoparameter (Biberauer and Roberts 2017, 26). The properties of *sin* (and *sig*) in Danish is an example of microparametric variation and this study writes itself into this type of investigation of language change.

1.1.1 Language change is driven by language acquisition

Lightfoot (2010) argues for a view of language change where change is driven by (first) language acquisition. According to Lightfoot, children acquire their native language(s) through an innate machinery (Universal Grammar or UG, see page 25) and language from the environment (primary linguistic data, PLD). UG and PLD combined with general cognitive abilities let the child acquire an internal grammar (I-language, see page 27). An important point about PLD (or more generally about E-language, the language that is externalized as e.g. speech, see page 27) is that it is imperfect and includes e.g. slips of the tongue, phonetic factors, and memory limitations. This has the practical consequence that some linguistic errors may be adopted by language learners and become part of the I-language of that speaker. "Small changes in E-language sometimes trigger new I-languages, with more far-reaching consequences. [...] Once that happens and some children have new I-languages, E-language changes further, because the new I-language entails that people speak differently. As a result, the new I-language may propagate through the population rather quickly" (Lightfoot 2010, 681).

Grammars change when the data available to the learner changes. If we assume that children are "perfect learners", i.e. that they acquire language perfectly based on the input that they have been exposed to, the input must have changed. Input change could happen due to language contact (with other languages or other dialects). It

could also happen to due second language acquisition: Adults are imperfect learners and introduce errors. A type of second language acquisition that should also be considered here is acquisition of another dialect, e.g. if a speaker of a Danish dialect is tasked with acquiring specific, normative features of standard Danish at school. Even within a homogenous speech community the input will differ slightly from the previous generation pushed by changes in other components of the grammar. This will let some learners acquire a slightly different grammar than that of their parents, which, given enough speakers with a slightly different grammar, will lead to an overall change in the language at a smaller or larger scale.

1.2 Generative tree structures

1.2.1 CP – IP – VP

This section is a brief introduction to the terminology and basic operations that I use in this thesis to describe language. The general theoretical background is generative grammar as taught at the English Department at Aarhus University. The approach to grammar that I use in this thesis is generative, comparative morphosyntax and I tend to use Danish and English (and at times other relevant languages) to illustrate the points under discussion. The primary textbook that is used in our local introductory syntax courses is Haegeman and Guéron (2000) and I refer the interested reader to this for a more detailed introduction. It is not obvious that language should be described in this way and particularly in a Danish context, generative grammar is not the most widely used approach. For the same reason, I choose to have this initial introduction. The text here and below is a modified version of a similar chapter from my Master's thesis, Ehlers (2017).

The illustration in figure 1.3 is borrowed from Christensen (2005, 30) and serves as a very nice summary of the three primary domains in the syntactic tree. CP, or Complementizer Phrase, is the topmost level. It "'looks' outwards (upwards) into the universe of discourse" (Christensen 2005, 28). The CP may e.g. host topicalised elements, question words (wh-elements), and its head node is the typical position for the tensed verb in Danish main clauses. The IP, or Inflectional Phrase, is embedded within the CP. IP is the prototypical space for the subject of a clause as well as e.g. finiteness-marking (Inflection) and sentential negation. VP (Verb Phrase, labelled vP in the illustration) is the base position for the main verb, which makes up the scaf-

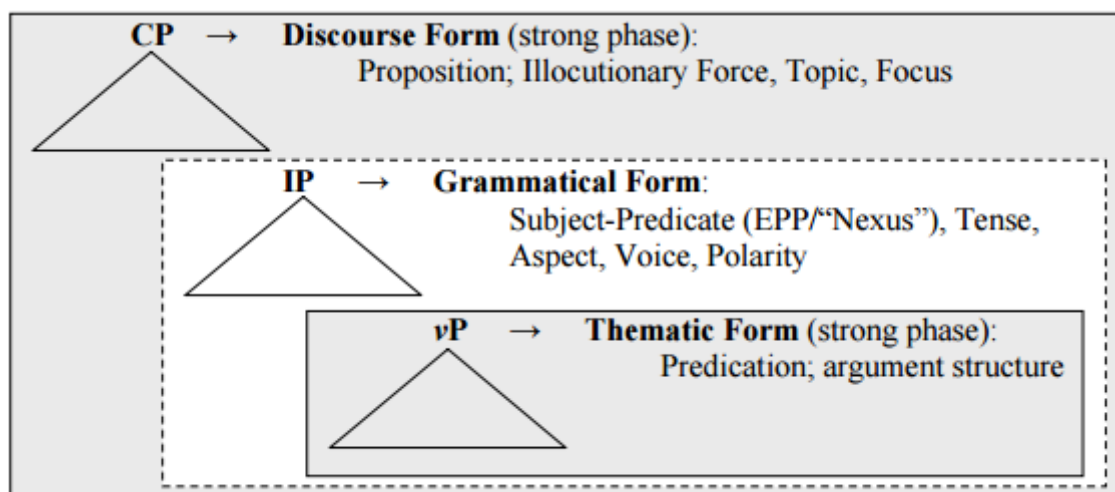
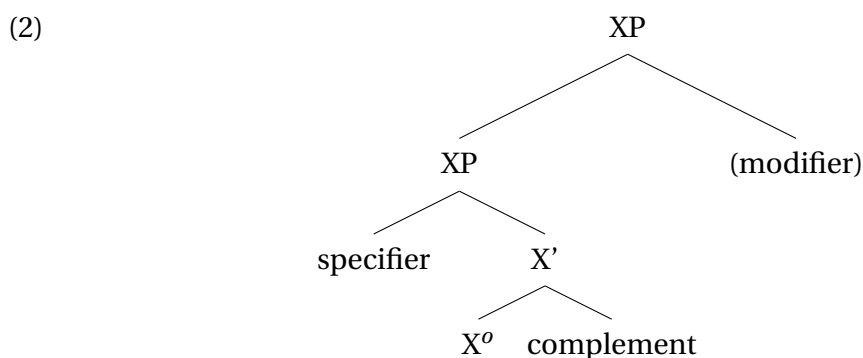


Figure 1.3: The clause and its main contents in very broad strokes. Illustration borrowed from Christensen (2005, 30).

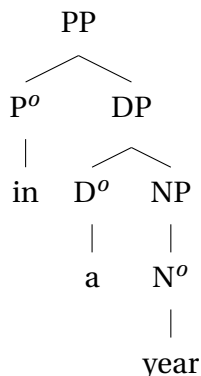
folding of the sentence. VPs also encode argument structure and thematic roles. VP is in a sense the thematic core of the clause and IP and CP build functional material on top of it.

Generative tree structures illustrate the analysis that language structure is best described as hierarchical rather than linear. The basic structure of a node in a tree is shown in (2) and every node is assumed to have the same basic structure. Commonly used XPs are e.g. CP (Complementizer Phrase), IP (Inflectional Phrase), VP (Verb Phrase), DP (Determiner Phrase), NP (Noun Phrase), AdjP (Adjectival Phrase), AdvP (Adverb Phrase), and PP (Prepositional Phrase).



Phrases are embedded in each other in order to build the tree structure. In the prepositional phrase *in a year*, the nominal *year* is embedded within the indefinite determiner phrase *a year* which is further embedded inside the prepositional phrase *in a year*. The nested structure is illustrated as the tree in (3).

(3)

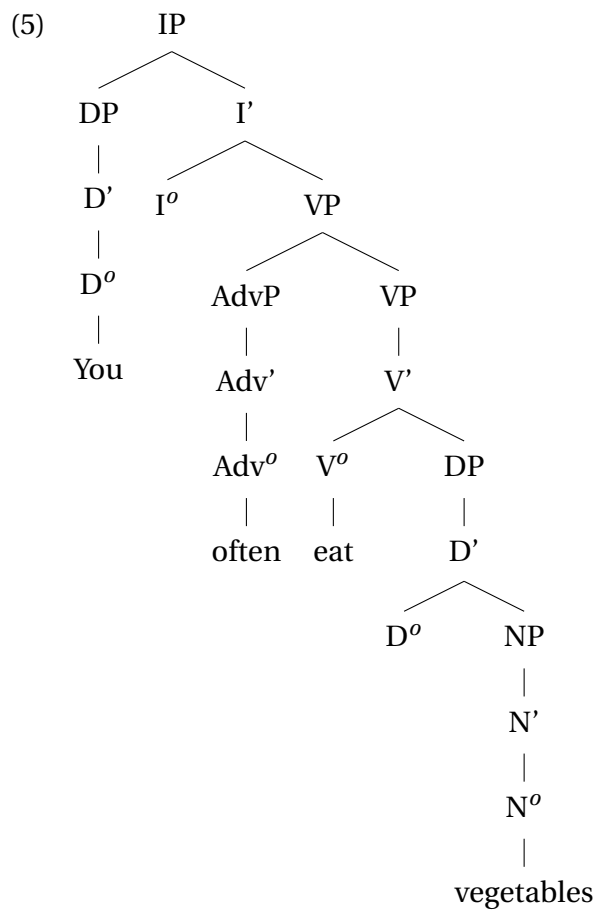


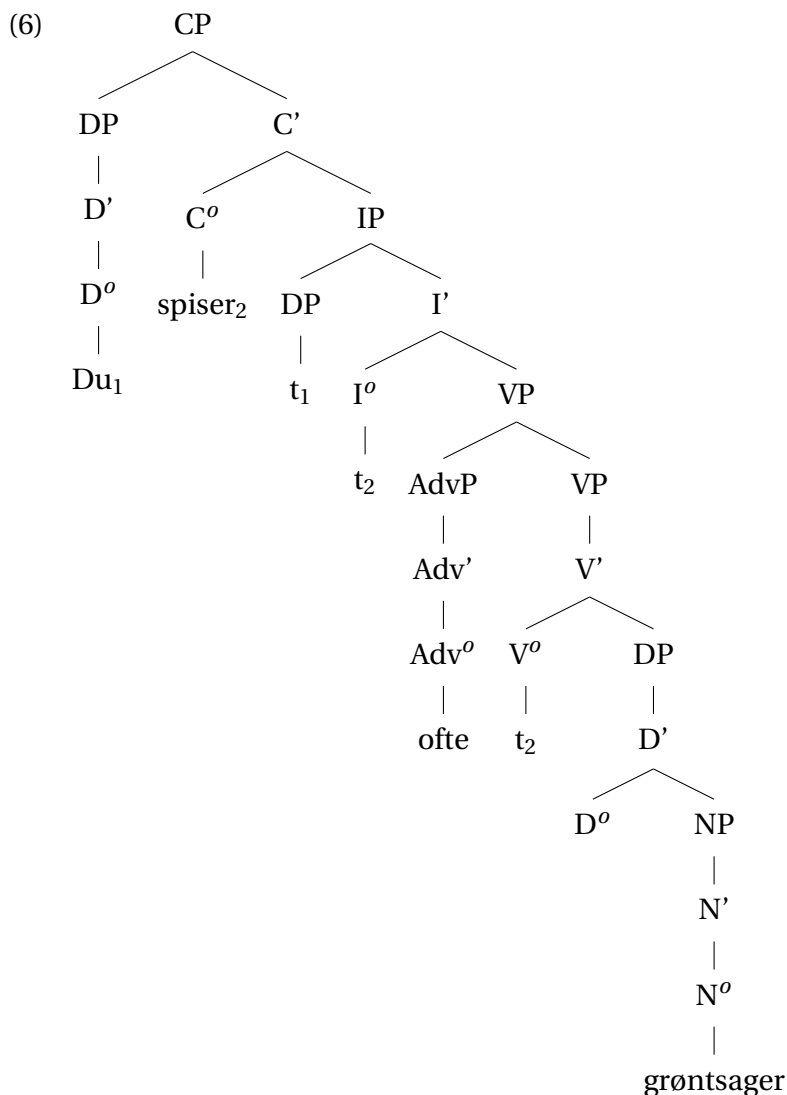
When a clause (IP or CP) is embedded within another clause, the higher clause is called the *matrix clause* and the lower clause is called the *embedded clause*.

- (4) a. [IP_{non-matrix} I know John.]
 b. [IP_{matrix} I saw [IP_{embedded} John eat apples.]]
 c. [IP_{matrix} I know [CP_{embedded} that John likes apples.]]

1.2.2 Movement

In (5) and (6), I illustrate an analysis of the same simple sentence in Danish and English. The great difference between the two is that Danish is a V2-language. In Danish and other V2-languages, the second constituent in the main clause is the tensed verb. This requires there to be two positions in all main clauses above the canonical subject and verb positions in IP and VP. These two positions are CP-spec and C°. CP-spec is where e.g. subjects or topicalized elements go. C° is the position to which the finite verb moves (see more in e.g. Vikner 1989). English is not a V2-language, which accounts for a bigger tree for the Danish sentence.





The Danish tree contains two instances of traces of movement, one for the verb *spiser* and one for the subject, *du*. They are indicated by t's with numbered indices. These illustrate the observation that items in the tree, and in sentences, may appear in a position distinct from their initial position. In English, the verb *eat* is at its initial position but in Danish the verb has been moved to the second position in the clause. The initial position of the verb can be seen in sentences where the moved verb is an auxiliary. The examples in (7) illustrate this with progressively longer auxiliary chains. The highest verb moves to the second position in the clause in Danish main clauses and any other verbs in the same minimal clause stay, unmoved, in their initial position.

- (7) a. Du spiser₁ ofte ___₁ grøntsager.
 b. Du har₂ ofte ___₂ spist₁ grøntsager.

- c. Du kunne₃ ofte ___₃ have₂ spist₁ grøntsager.
- d. Du ville₄ ofte ___₄ kunne₃ have₂ spist₁ grøntsager.
- e. 'You (would) often (could) (have) eaten vegetables.'

Verb movement is a specific case of *head movement* where a head moves upwards in the tree to another head position. In the examples of V2-related head movement above, the lexical item in the verbal head V° moves first to I° (we assume that this is also the case for Danish, even if the Danish verb cannot ever be found in I°) and further into the head of CP, C°, which is its final position. Movement of the nominal head into the D° (which is how e.g. Danish forms definite nominals with postnominal definiteness) is another example of head movement (discussed further in section 1.2.6).

Entire phrases may also move. A specific example of phrasal movement (also known as XP-movement) is the *wh*-movement in (8a) where the interrogative element moves into (typically) the highest position of the clause (CP-spec). The base position of the moved element *what* can be seen in the declarative version of the interrogative sentence, (8b), where the *wh*-element is replaced with a non-interrogative nominal.

- (8) a. What₁ did Peter often eat ___₁?
- b. Peter often ate vegetables₁.

Other examples of XP-movement are *passivisation* and *raising* where nominal phrases move from their base position in order to e.g. be assigned Case in another phrasal position.

As discussed on page 31, traces and indices were argued to not be primitive elements in the newer versions of generative grammar. I still use them for presentational purposes but do not, in doing so, make any claim about their ontology. Movement may also be illustrated with arrows with the same reservations, i.e. that arrows do not have ontological status in the theory but are useful for presentational purposes.

1.2.3 Arguments, adjuncts, and theta-role assignment

I distinguish two basic types of elements in the clause: arguments and adjuncts. Specifiers and complements are both arguments, and modifiers are adjuncts. Adjuncts are more or less optional elements in the sentence and often provide extra information about time, place, or manner. Arguments are in some way required by the another element in the clause (e.g. main verb) of the clause. The direct object, for instance,

is an argument required by the transitive verb. The examples in (9) illustrate some of the differences between arguments and adjuncts in a clause.

- (9) a. Harry drinks butterbeer.
 ARG ARG
- b. Harry drinks butterbeer in Hogsmoade.
 ARG ARG ADJ
- c. Harry drinks blood in Hogsmoade.
 ARG ARG ADJ

The content of the arguments typically defines the content of the clause in a more meaningful way than the content of the adjuncts. The sentence in (9a) *Harry drinks butterbeer* is not very different from the sentence in (9b) where the place adjunct *in Hogsmoade* is added. This also illustrates that the adjunct can be left out without changing the meaning of the sentence much. The sentence, however, may change quite dramatically if the argument *butterbeer* is changed to *blood* as in *Harry drinks blood in Hogsmoade* in (9c). Arguments may generally not be left out of the sentence: *Harry drinks butterbeer*, where the subject argument *Harry* is deleted, is not a grammatical sentence in English. Similarly, *Harry drinks butterbeer* is grammatical but the meaning of the sentence changes quite a bit and becomes a general statement about Harry's (perhaps problematic) drinking habits.

I follow the analysis from e.g. Chomsky (1995, 289-90) and Kratzer (1996) where the predicate selects a number of (internal) arguments and assigns θ -roles (theta-roles, thematic roles, semantic roles) to them in a local relationship. According to the θ -criterion from Chomsky (1981), all arguments must be assigned a θ -role and each θ -role is assigned to only one argument. The number of required arguments and corresponding theta roles depends on the verb. All clauses in Danish and English have a subject and this subject also starts in a local relationship with the predicate, according to the VP-Internal Subject Hypothesis (whose modern form is due to e.g. Fukui (1986), Sportiche (1988), Koopman and Sportiche (1991), and McCloskey (1997)). The subject is typically also the *external argument* of the predicate while one or more objects are the *internal arguments*.

Working downwards from a maximally large argument structure, a verb like *give* can be ditransitive and require two internal arguments: Something that is given (typically the direct object with a theta role as THEME) and the one who receives (typically the indirect object with a theta role as BENEFICIARY). The giver, the subject, is the ex-

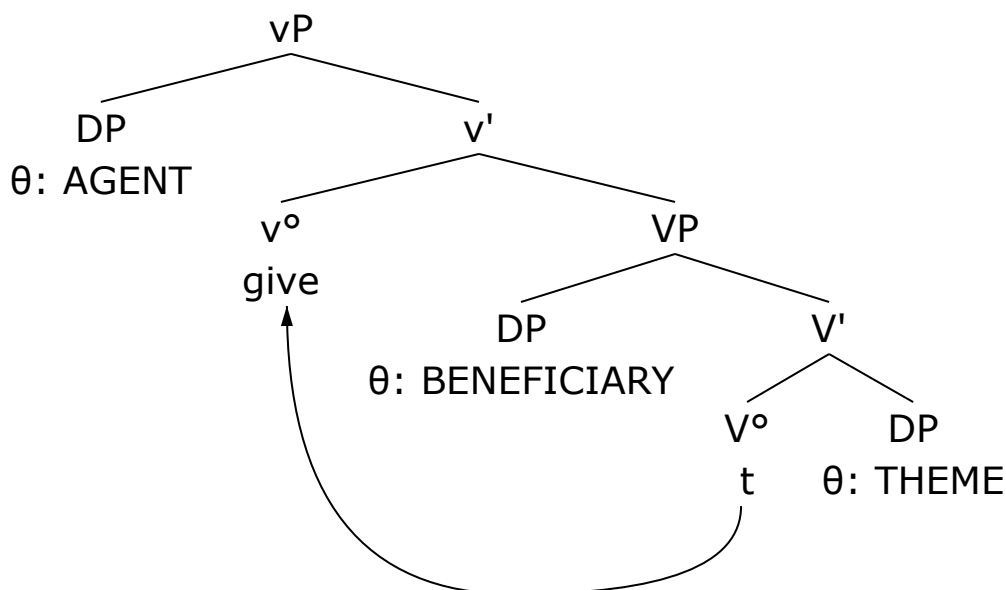


Figure 1.4: Ditransitive verb *give*: Theta role assignment with two internal arguments (direct object and indirect object) and one external argument (subject).

ternal argument and is base-generated in the specifier position of the so-called little-*v* where it receives the AGENT theta role. Little-*v* is described by Chomsky as "a light verb to which *V* overtly raises" (Chomsky (1995, 290)) in the constructions where it is present. Consequently, all three theta roles are assigned locally from the verbal head to its complement (THEME), to its initial specifier (BENEFICIARY), or to the specifier of *v* to which the verb raises and assigns the third theta role (AGENT) to the external argument. The positions and θ -roles in the *vP*-*VP* tree are illustrated in figure 1.4.

A verb like *drink* can be transitive and require one internal argument (the drinker, AGENT) and one external argument (the drinker, AGENT). It may also be intransitive and require just the external argument. I again follow the analysis in Chomsky (1995) and assume that the little *v* layer is also present in transitive and intransitive (unergative) structures². The theta roles are assigned locally by the verb as the trees in figures 1.5 and 1.6 show.

Another theta role configuration is known as the *unaccusative* which covers a class of intransitive predicates where the subject is not the AGENT and does not actively play a part in performing the action described by the verb. An example of an unaccusative is the sentence *The tree fell*. The tree does not play an agentive part in falling, which

²Little *v* is strictly speaking only necessary in the ditransitive construction and an alternative is that the external argument is base-generated in *VP*-spec with no added little *v* layer.

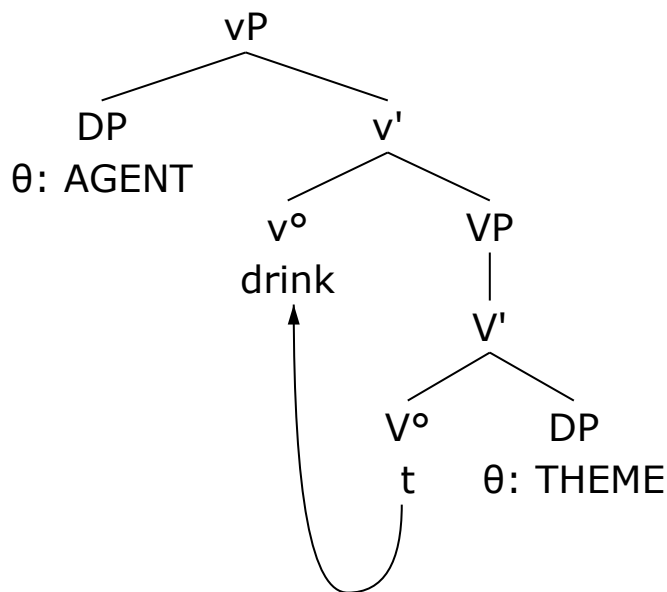


Figure 1.5: Transitive verb *drink*: Theta role assignment with one internal argument (direct object which is assigned THEME by the verb) and one external argument (subject which is assigned AGENT by the verb which has moved to little *v*).

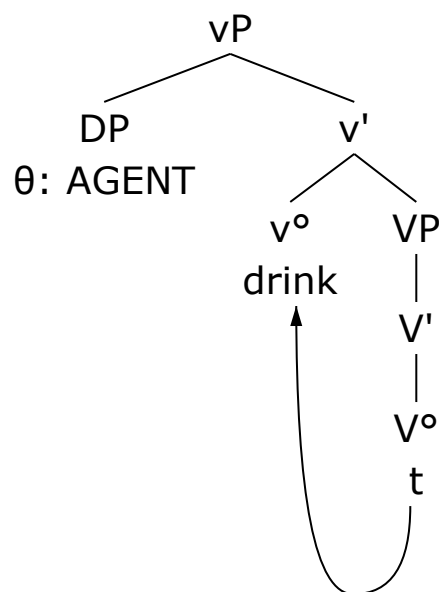


Figure 1.6: Intransitive (unergative) verb *drink*: Theta role assignment with one external argument (subject which is assigned AGENT by the verb which has moved to little *v*).

is further evident from the fact that you cannot add an agentive modifier such as *on purpose* or *in order to* (without coercing an AGENT reading, at least). The unaccusative verb assigns the THEME θ -role to its complement, which then moves up into VP-spec and becomes the de-agentive subject, as illustrated in 1.7. Unaccusatives do not have a little *v* layer, according to Chomsky (1995, 290), which makes sense under the assumption that AGENT is assigned to the vP-spec position.

1.2.4 Case-assignment

Both Danish and English noun phrases are assigned Case, either nominative or accusative. Nominative Case is assigned to the specifier position of the finite IP by I^0 . Accusative Case is assigned by lexical V^0 or preposition P^0 to its nominal complement. For English, the complementizer *for*, which is in C^0 , may also assign accusative Case to its nominal complement, usually the subject in the specifier position of IP. A special

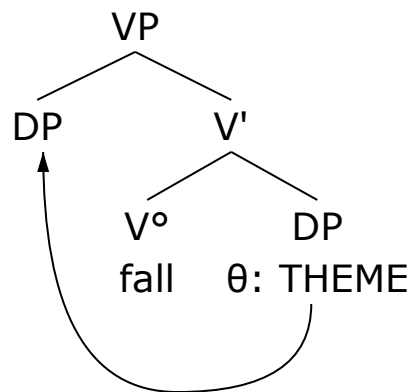


Figure 1.7: Unaccusative *fall*: Theta role assignment with one internal argument and no external argument.

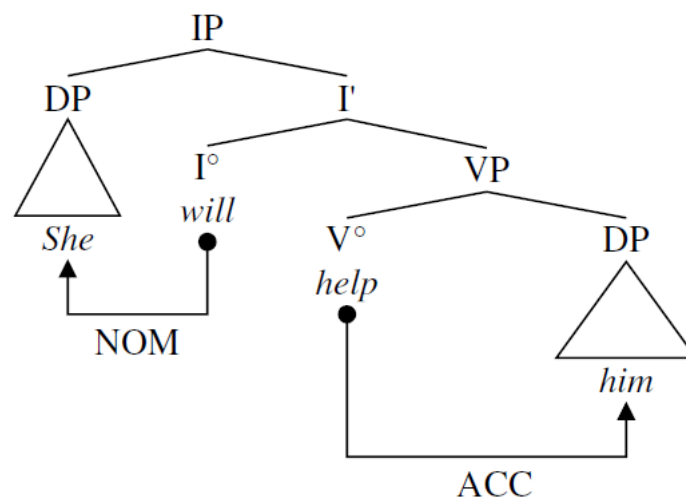


Figure 1.8: Case assignment in finite clauses. The main verb assigns accusative to its complement and the finite (feature of) I° assigns nominative to the subject in IP-spec.

case of Case assignment is Exceptional Case Marking, or ECM. This is the situation in which a lexical verb in the matrix clause assigns accusative Case to the subject position in the embedded clause. A non-exhaustive list of options for Case assignment are illustrated in the trees in figures 1.8 to 1.11. The trees themselves are kindly borrowed from Sten Vikner.

1.2.5 Subjects – overt and non-overt

All English and Danish clauses are assumed to have subjects, also in the cases where that subject is not visible or audible.

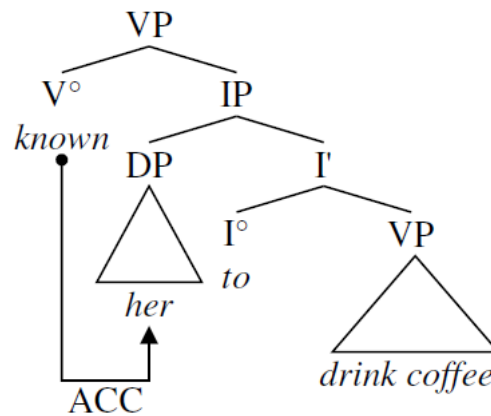


Figure 1.9: ECM case assignment: The main verb in the matrix clause assigns accusative Case to the subject in the embedded infinitival clause.

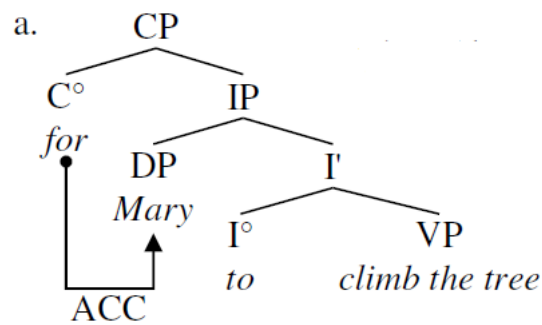


Figure 1.10: Case assignment by the complementiser *for* in C° which assigns accusative to the subject in IP-spec.

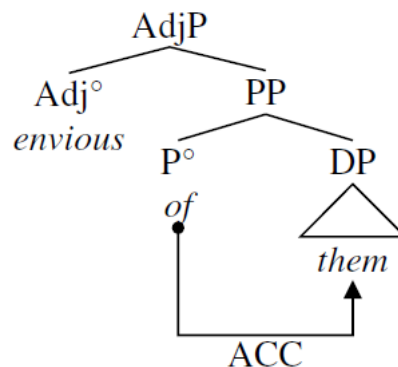


Figure 1.11: Case assignment by prepositions: The P° assigns accusative to its nominal complement.

- (10) a. John wanted [her to go for walks]
 b. John wanted [to go for walks]
 c. John₁ wanted [PRO₁ to go for walks]
- (11) PRO_{arb} to go for walks is probably healthy.

In (10a), *her* is the overt subject of the infinitival embedded clause [her to go for walks]. It is assigned accusative Case by the matrix verb *wanted*, rather than the nominative Case that subjects of clauses usually receive. In (10b), the embedded infinitival clause is [to go for walks]. There is no overt subject but I assume that the clause still has a subject, just a non-overt one. This is indicated in (10c) where PRO is the non-overt subject of the embedded infinitival clause. This non-overt subject is co-indexed with *John*, the matrix clause subject. This captures the fact that *John* in (10c) still seems to be the one going for walks. The claim that seemingly subjectless clauses still have subjects is desirable for a number of reasons. One advantage is that it allows the lexical verb to assign its AGENT (OR EXPERIENCER) θ -role to PRO. Another is that it allows us to account for the distribution of reflexive pronouns in embedded clauses in a consistent way, which will be relevant in chapter 2. In (11), PRO is not co-indexed with another element in the clause but has arbitrary reference. This also reflects the general nature of the statement.

Many of the tree structures in the thesis lack a number of details for reasons of transparent presentation. One of the primary omissions is that I do not necessarily include vP (the outer shell of the VP) and movement of subject from the verbal domain. Additionally, there is a general redundancy in the trees in that I generally illustrate all nodes as having both phrase (XP), bar (X') and head (X) levels. I also assume that all noun phrases are embedded in a DP-layer even in the absence of definiteness or determiners (see section 1.2.6). Lastly, I go against what seems to be modern convention by using IP rather than TP for the topmost phrase below CP (this convention follows work in the 1990s on a more fine-grained clausal spine, which began with Pollock (1989)).

1.2.6 DP: Nominal structure

I assume, following analyses first suggested by Abney (1987), that nominals such as *horses*, *those wonderful horses*, *Alanna's horse*, and *her* are determiner phrases (DP) and not only noun phrases (NP). Some researchers (e.g. Boskovic (2005) et seq. who

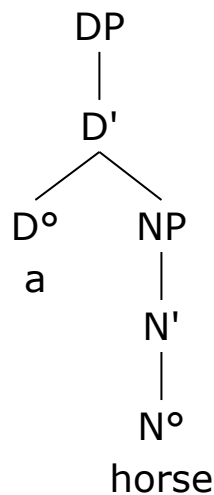


Figure 1.12: English indefinite nominal *a horse* with the indefinite article *a* as the head of the DP and the noun *horse* as the head of the NP.

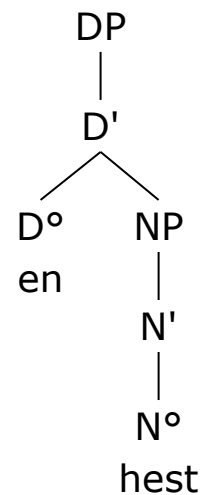


Figure 1.13: Danish indefinite nominal *en hest* with the indefinite article *en* as the head of the DP and the noun *hest* as the head of the NP.

argues that nominals in some languages do not project a DP) have made the argument that the DP is not a universal category and that there are languages where nominals are not embedded inside DPs. However, this is argued to be the case for languages without articles, which both Danish and English uncontroversially possess. This puts both languages squarely in the category of DP-languages (as opposed to NP-languages which may or may not have DP as a category).

I illustrate my chosen analysis of various kinds of English and Danish nominals below with relevant comments in the captions. The Danish examples are translations of the English ones. I contrast the two languages in all the example sets in order to clearly show where Danish is different from English, which is particularly clear with the definite, unmodified noun phrases in figures 1.16 and 1.17 where the Danish definite article is postnominal. Another difference is the fact that Danish adjectives have *weak* and *strong* agreement, depending on the definiteness of the DP. This is visible in the difference between *en sort hest* in the indefinite nominal in figure 1.15 and *den sorte hest* in the definite nominal in figure 1.19. *Sort* is an instance of strong agreement and *sorte* is weak agreement. This is explicated further in section 2.2.2.

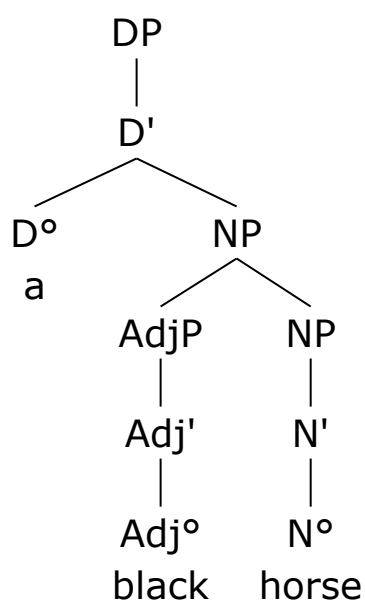


Figure 1.14: English modified indefinite nominal *a black horse* with the indefinite article *a* as the head of the DP, an adjectival modifier *black*, and the noun *horse* as the head of the NP.

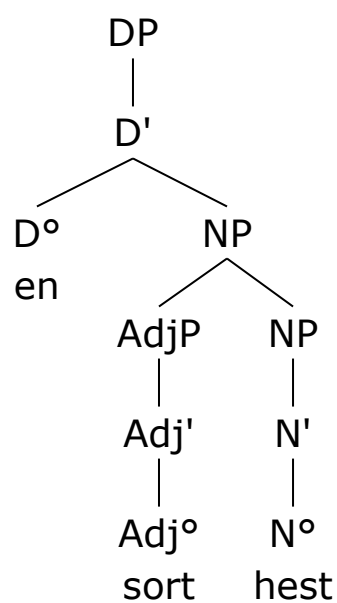


Figure 1.15: Danish modified indefinite nominal *en sort hest* with the indefinite article *en* as the head of the DP, an adjectival modifier *sort*, and the noun *hest* as the head of the NP.

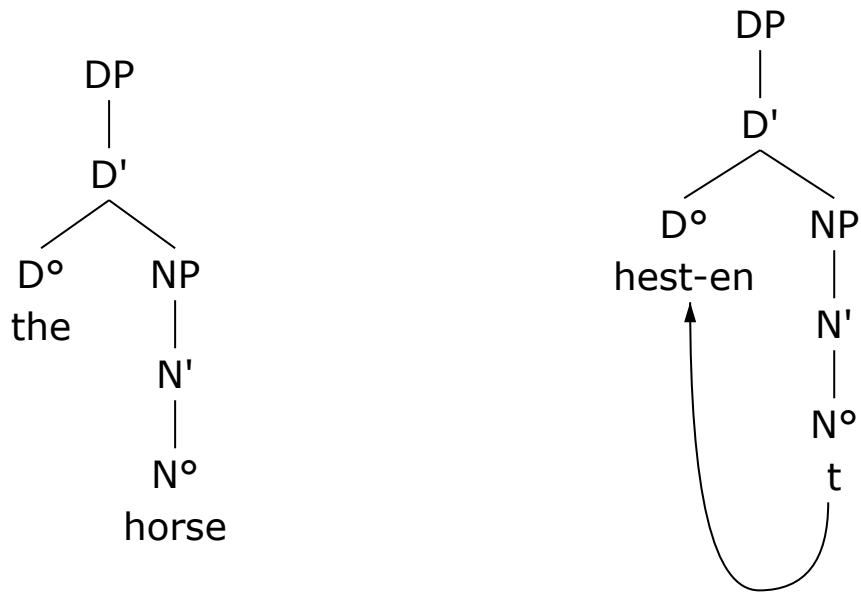


Figure 1.16: English definite nominal *the horse* with the definite article *the* as the head of the DP and the noun *horse* as the head of the NP.

Figure 1.17: Danish definite nominal *hest-en* with the postnominal definite article *-en* as the head of the DP and the noun *horse* as the head of the NP. The noun is moved up to join the definite article.

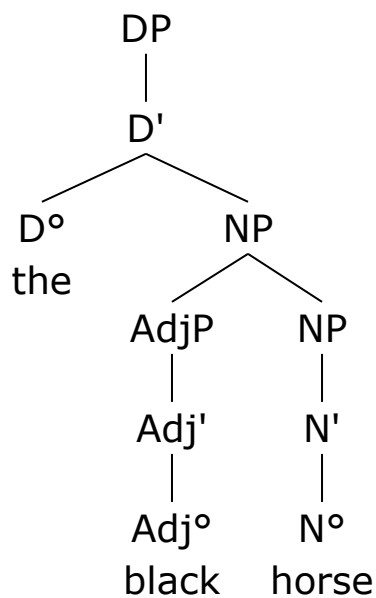


Figure 1.18: English modified definite nominal *the black horse* with the definite article *the* as the head of the DP, an adjectival modifier *black*, and the noun *horse* as the head of the NP.

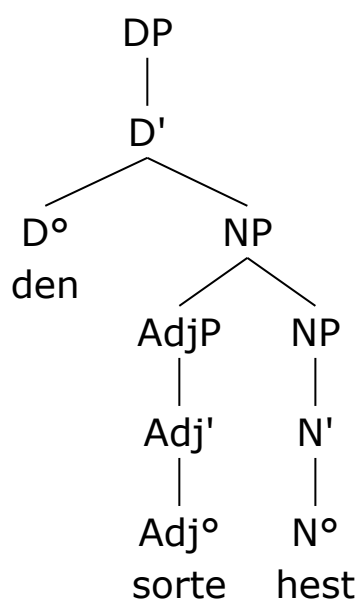


Figure 1.19: Danish modified definite nominal *den sorte hest* with the definite article *den* as the head of the DP, an adjectival modifier *sorte*, and the noun *hest* as the head of the NP. The noun stays in situ in the NP when it is modified by an adjective.

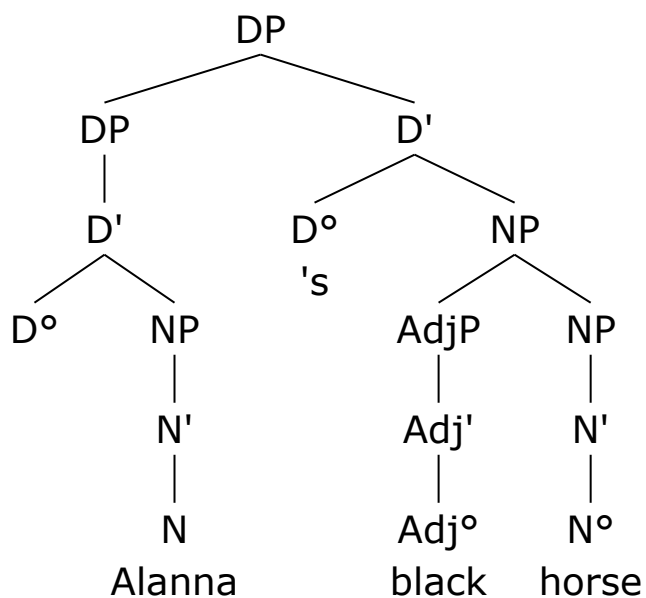


Figure 1.20: English modified definite nominal *black horse* with the possessor clitic *'s* as the head of the DP, an adjectival modifier *black*, and the noun *horse* as the head of the lower NP. The specifier position of the DP is filled by the possessor *Alanna* which is itself a DP.

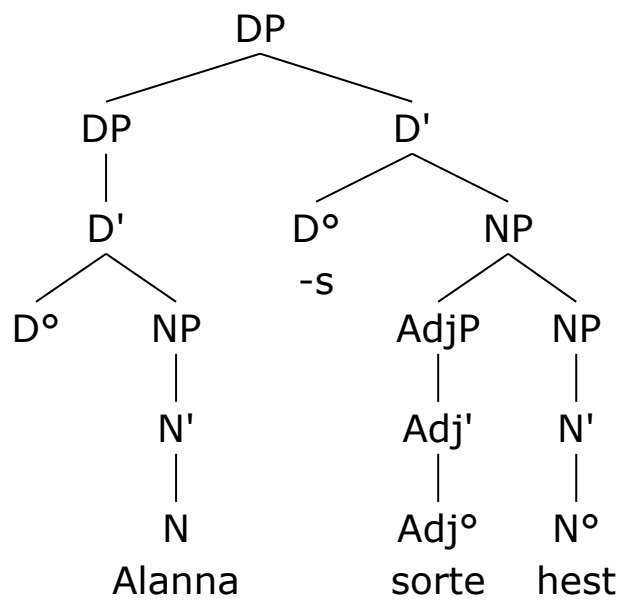


Figure 1.21: Danish modified definite nominal *sorte hest* with the possessor clitic *-s* as the head of the DP, an adjectival modifier *sorte*, and the noun *hest* as the head of the lower NP. The specifier position of the DP is filled by the possessor *Alanna* which is itself a DP.

1.3 Theoretical framework: Principles and Parameters

1.3.1 A general generative view of language

In this section I briefly introduce a few concepts that were introduced in earlier work by e.g., and most famously, Noam Chomsky. These are still indispensable in the discussion of language in terms of generative grammar, of linguistic theory and the scientific study of language in general.

A key concept here is Universal Grammar (UG) which refers to the innate (species-general, present in every human being) knowledge or skills that

- (12) allows every typical human being to learn their native language(s) in a seemingly effortless way if given species-appropriate input (acquisition)
- (13) makes human language(s) look the way they do (the subject matter of linguistics)

These two points necessarily intertwine: Humans are born with a certain mental machinery (which we call UG) that lets us acquire human language, and only human language. This in turn means that the language we acquire must be structured in a way that lets us acquire it. No kids grow up to speak Whale even if they spend a lot of time at the beach, and no human language expresses negation by inserting a "No" as the fourth word in a sentence. Humans are wired to learn human language, and human language is wired to be learnt by humans. This does not, crucially, mean that generativists claim that there is a Universal Grammar *language*, i.e. that all languages are the same or something to that effect. It only means that humans learn human language because we have something in our brains that enables us to do so.

Competence and performance

First of all is the question of whose language generative grammar attempts to describe:

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogenous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language

in actual performance.

(Chomsky 1965, 3-4)

This ideal speaker-listener as described by Chomsky, in his 1965 book *Aspects of the Theory of Syntax* (henceforth *Aspects*) does not exist. No speech-community is completely homogenous, and even if it can (and probably should) be argued that native speakers by definition know their native language perfectly, no speaker-listener is ever unaffected by the limits set by having a finite brain. This should be clear from the humorously overstated tone of the passage. The quote has still been used, rather unfairly, to criticize the generative grammar enterprise for viewing language in an unrealistic way (see e.g. Newmeyer (1983, 74), who criticizes the critics or Cobley (2005, 127) who *is* a critic, 40 years after *Aspects*).

The point of the passage is that only in a counter-factual universe where this ideal speaker actually exists, would the actual performed language of this person be the object of study for linguistics. In the real world, because we do not have access to an ideal speaker-listener, the subject matter for linguistics should not be the performance of single individuals but rather the underlying system of knowledge and ability that allows speakers to perform as they do. However, the only thing that we have access to is the performance of speakers with all the errors, slips of the tongue and processing limitations that this entails (Chomsky 1965, 4).

Grammaticality and acceptability

The concepts of grammaticality and acceptability add to the distinction between competence and performance. In Chomsky's words:

For the purposes of this discussion, let us use the term "acceptable" to refer to utterances that are perfectly natural and immediately comprehensible without paper-and-pencil analysis, and in no way bizarre or outlandish. Obviously, acceptability will be a matter of degree, along various dimensions. [...]

The notion "acceptable" is not to be confused with "grammatical". Acceptability is a concept that belongs to the study of performance, whereas grammaticalness belongs to the study of competence.

(Chomsky 1965, 10-11)

Crucially, acceptability and grammaticality do not necessarily overlap. Steven Pinker uses the sentence "Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo" to illustrate a grammatical sentence that is also unacceptable (it certainly is rather bizarre and outlandish, and does require some paper-and-pencil analysis to unpack) (Pinker 1995, 210). The even more classic example of a phrase that is both unacceptable and grammatical is Chomsky's own "colorless green ideas sleep furiously" (Chomsky 1957, 17).

The reverse situation, ungrammatical but acceptable, is possible as well. When Paul McCartney sings "But in this ever-changing world in which we live in" in *Live and let die*, the doubled preposition *in* makes the utterance ungrammatical but still acceptable enough to be sung. Note that it is then the job of the linguist to determine whether a given utterance is grammatical, independently of how speakers would rate a sentence for acceptability (Haegeman 1995, 7-8).

I-language and E-language

A further distinction is that between I-language and E-language, both introduced by Chomsky in 1986 in his *Knowledge of Language*. E-language, in short, is "everything about language which is external to the mind" (Terje Lohndal, p.c.). This includes (political) notions like *Norwegian* and *Japanese*, as well as the many outputs of linguistic production (performance).

The E in E-language and I in I-language are abbreviations of a number of things. The *Stanford Encyclopedia of Philosophy* gives two different E-words and three different I-words: **Extensional** and **External** versus **Intensional**, **Internal** and **Individual** (Scholz, Pelletier, and Pullum 2016). I(ndividual)-language indicates a view that sees language as a system in the mind of individual speakers, not e.g. as a shared cultural construct that has an existence outside the minds of speakers. This is not to say that language is independent of culture or community but rather to emphasize that the object of study for my branch of linguistics resides inside the brains of individual speakers.

This same point can be made with regard to the Internal-External dichotomy. E(xternal)-language is performance, I(internal)-language is competence. Intensional and Extensional are concepts from logic and can be illustrated in terms of a mathematical function. Given two functions f and g , we can get the same output by conveniently choosing a clever input value:

$$f(x) = x + 2, f(4) = 6$$

$$g(x) = 10 - x, g(4) = 6$$

The extensional part of the function is the output, which for both functions amounts to 6. Without further knowledge, this could lead the naïve mathematician to propose that f and g are the same function. The intensional part of the function is the function itself. Here, f and g are quite different and given a different input value would yield very different outputs. Returning the analogy to linguistics, E-language can be described as the (extensional) output of a (complicated) function and the aim of the linguist is to uncover the underlying (intensional) mechanism that yielded the output. This, however, is made trickier by the fact that two different functions can give the same output under certain circumstances (like f and g above) – and, of course, that human language is a bit more complicated than a linear function.

Lightfoot (2010, 677) describes external language as "the mass of unanalyzed utterances that a child might hear" and internal language as "the systems that grow in children on exposure to external language".

1.3.2 Government and Binding – rich UG and levels of representation

The titles Principles and Parameters (PP) and Government and Binding (GB) are often used more or less interchangeably as a descriptor of a specific kind of or approach to generative grammar which is followed (or supplemented by) an approach known as Minimalism. Lohndal and Lasnik write that "Principles and Parameters Theory comes in two incarnations: as Government and Binding Theory (1980s) and as the Minimalist Program (late 1980s until today)" (Lasnik and Lohndal 2010, 40). This quote exemplifies the idea that the Minimalist Program is not so much a departure from earlier work as a development and refinement of the same ideas that were around in the 1980s.

The basic idea behind describing grammar in terms of principles and parameters is that some parts of grammar come pre-defined (the principles) and some depend on linguistic input in infancy and childhood (the parameters).

What we expect to find, then, is a highly structured theory of UG based on a number of fundamental principles that sharply restrict the class of attainable grammars and narrowly constrain their form, but with parameters that have to be fixed by experience. If these parameters are embedded in a theory of UG that is sufficiently rich in structure, then the languages that are determined by fixing their values one way or another will appear to be quite diverse, since the consequences of one set of choices may be very different from the consequences of another set; yet at the same time, limited evidence, just sufficient to fix the parameters of UG, will determine a grammar that may be very intricate and will in general lack grounding in experience in the sense of an inductive basis.

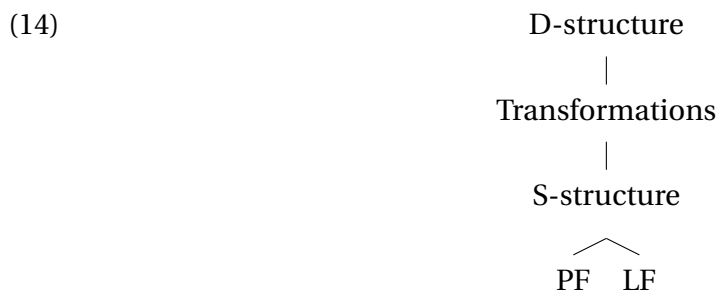
(Chomsky 1981, 3-4)

The view of language expressed by Chomsky here is that of a rich innate system (Universal Grammar, UG) that helps the child eventually acquire its native language(s) by restricting the set of possible attainable human grammars. Chomsky describes UG as consisting of several interacting subsystems: lexicon and syntax as the "base" together with PF (Phonetic Form) and LF (Logical Form) components. The structures generated by combining lexicon and syntax "are assigned PF- and LF-representations". The PF representation has to do with auditory or visual output (in case of e.g. sign language), while the LF representation carries the semantics of the utterance. Added to this are several more specific subsystems (bounding, government, θ -theory, binding, Case and control) which Chomsky describes as "subsystems of principles", which presumably means that these are part of UG, as well (Chomsky 1981, 5).

Added to the principles in UG are a number of parameters which delimitate "[t]he (limited) ways in which languages can differ syntactically[...]. The child then only has to set the correct value (mostly thought to be a choice between two options—like a switchbox as Jim Higginbotham aptly put it) based on the primary linguistic data" (Lasnik and Lohndal 2010, 43). Again turning to Chomsky, he argues that it is probably a reasonable assumption that UG only allows "a finite number of core grammars" (Chomsky 1981, 13), a consequence of his assumption that UG supplies "a finite set of parameters, each with a finite number of values, apart from the trivial matter of the morpheme or word list, which must surely be learned by direct exposure for the most part". I point these statements out because core elements of latter approaches to generative grammar crucially revolve around *minimizing* UG and possibly even

having parameters be emergent rather than an inbuilt fixed, finite set.

The model in (14), here adapted from Lasnik and Lohndal (2010, 42), sums up the core machinery of the GB era transformation and representation of syntactic structures.



The syntax component generates structures into which lexical items are inserted. This is the D-structure (deep structure) representation. This representation is transformed as appropriate and necessary to the S(urface)-structure representation with the rule Move- α (assumed to be present both in the syntax component and at LF and PF (Chomsky 1981, 18)). Move- α is a generalized movement rule "where α is some category" (Chomsky 1981, 18), i.e. with very few restrictions. Finally, the S-structure representation maps onto PF and LF representations where further instances of movement may take place.

1.3.3 Minimalist framework and terminology

"Minimalism advances the hypothesis that language is a 'perfect' solution for meeting the requirements imposed by the external systems. It seeks principled explanations instead of purely technical accounts" (Lasnik and Lohndal 2010, 46). A part of this venture is a move away from the richly structured UG of the earlier theory, towards as small a Universal Grammar as possible and necessary. This also entails that many of the principles and possibly also parameters that in GB times were thought to be innate, could perhaps be acquired rather than learned. I constrain myself to introducing the technology and theoretical objects that have been introduced under the heading of Minimalism rather than going into the background of why it was introduced in the first place. This introduction will be helpful in order to follow some of the specific proposals put forth to explain reflexive binding *post* Government and Binding times.

First of all, many conceptions of Minimalist syntax are *derivational* rather than *representational*. This means that the process of building the tree (or the sentence,

as it were) in itself takes care of whatever requirements that arise in such a process. Even in a seemingly simple sentence like "John eats", there are several steps before the final output. These, presumably, involve nominative Case assignment, movement of "John" to a proper subject position, the match-up of the verb and its finite -s ending, to mention a few superficial ones.

In a representational approach in line of GB, these steps would be taken care of by positing the different levels of representation (D- and S-structure, as previously discussed). At one level of representation, D-structure, the initial state of the sentence is assembled structurally all at once with subsequent lexical insertion, then transformed into S-structure, another level of representation. Finally the sentence is sent to PF and LF where even more transformations might take place (covert movement). The levels of representation, S and D, crucially, were shown to be empirically unwanted and unnecessary and were abandoned in Chomsky (1995) (see e.g. page 190 where he concludes that "Condition A *cannot* apply at S-Structure" and that this in turn is an argument in favour of abandoning S-Structure entirely). The only mechanisms and levels left, then, are the mechanism for assembling the sentence and the two interface components of phonology (PF, sound or sign output) and semantics (LF). The reason for keeping the PF and LF is the fairly evident insight that language has externalization (sound or sign) as well as meaning, and that these two are disconnected to some extent.

Numeration, Agree, and Merge

In the Minimalist derivational approach, the sentence assembly (the derivation) starts out with an unordered selection from the Lexicon, a Lexical Array. The Lexical Array becomes a Numeration that specifies how many instances of the particular item from the Lexicon that will be included in the derivation. Citko (2014, 8-9) describes this process (or, rather, metaphor for a process) in more detail. Following the *Inclusiveness Condition* first introduced in chapter four of Chomsky (1995), nothing else can be added during the course of the derivation. This has the consequence that indices and traces are eliminated as theoretical constructs since they cannot in a meaningful way be said to be part of an initial Numeration. Another consequence is that the concept of X-bar structure loses its place in the theory as it is not in itself a primitive of the lexicon.

The items in the Numeration are assembled into a hierarchical structure through

the operations Merge and Agree. Merge is a simple and all-powerful tool that creates bigger structures from smaller ones by merging them together in pairs. That Merge is a binary operation is the simplest option and therefore the most Minimalistically desirable: a binary operation can assemble ternary structures but a ternary operation cannot assemble binary ones. For Merge to not be binary, then, would require a more complicated system than assuming binariness.

Merge comes in different flavours: External and Internal Merge. External Merge is Merge of an item from the Numeration with a branch in the tree (or of two items from the Numeration). Internal Merge is perhaps better known as Move and is a movement process imagined as the Merger of two items already in the tree. This has the eventual consequence that one item moves to the other.

Agree is a relationship "between an LI [Linguistic Item] and a feature F in some restricted search space" (Chomsky 2000, 101). Wurmbrand describes Agree as "an abstract feature matching relation between a functional head and a 'goal' in situ" (Wurmbrand 2003, 229). The functional head is sometimes called a Probe, and the goal capital Goal (Citko 2014, 21). In a later article, Wurmbrand gives the more specific definition of Agree given in (15) which I adopt as well. The specifics of valuation of features, which the definition presupposes, are discussed in the upcoming section.

- (15) A feature F:___ on α is valued by a feature F:val on β iff
- a. β c-commands α AND
 - b. α is *accessible* to β .
 - c. α does not value [a feature of β]/[a feature F of β]

(Wurmbrand 2014, 132)

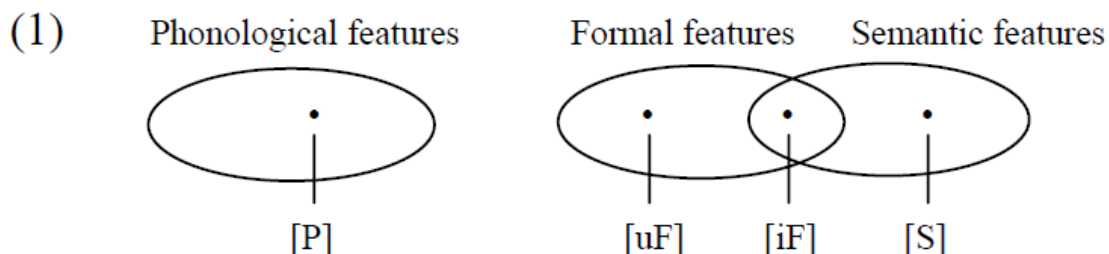
Features and valuation

Words have properties that determine their use and content, and these properties may well be distinct from how the word looks or sounds. These properties are called (morphosyntactic) features and they have become an important element of generative theory. I will not be able to provide an exhaustive review of the very many different approaches to features in generative grammar but rather limit myself to a few newer ones that are both fairly accessible and academically popular. Features are assumed to be theoretical objects in minimalism, rather than purely descriptive tools, "a feature [plural] for example is used analogously to chemists' use of H for the real-

world thing hydrogen" (Adger and Svenonius 2011, 28). In this view, a description of the feature inventory of a language can be seen as analogous to the description of the language itself, and differences in feature inventories as differences between languages (ibid.).

"A morphosyntactic feature is a property of words that the syntax is sensitive to and which may determine the particular shape that a word has. Features seem to be the core elements of languages that relate sound and meaning" (Adger 2003, 19). A concrete example of this, also borrowed from Adger, is the difference between "The sheep bleat" and "The sheep bleats". In the first case, the word *sheep* is plural and in the second it is singular. There is no difference in phonological realization on the nouns but the number difference is visible on the agreeing verb. The idea is that the two *sheep* differ in which features they have: one has a feature that indicates plurality, the other does not.

Number is a formal feature that in the *sheep* example has both semantic interpretation (there being either one or several sheep) and syntactic consequences (in that it triggers verbal agreement). Zeijlstra makes the distinction between phonological, formal and semantic features illustrated in the diagram below and definitions in (16) below.



- (16) a. A feature [uF] is semantically vacuous.
 b. A feature [uF] triggers syntactic operations Move and Agree in order to be deleted
 (Zeijlstra 2008)

Number is an *interpretable* feature (iF) on the noun *sheep* in the example above. The number feature on the verb only has a function in terms of syntactic agreement with the noun. It is an *uninterpretable* feature (uF) on the verb since it has no semantic content apart from that provided by the noun. Uninterpretable, purely formal, features need to be deleted in the course of the derivation. The underlying logic is that the purpose of the sentence assembly is to send instructions to the PF (sound) and

LF (meaning) interfaces. If these instructions contain material that cannot be fully processed at the interfaces, the derivation *crashes* (Chomsky 1995, 201) (Chomsky's condition of *Full Interpretation*). Uninterpretable features by definition have neither semantic (readable at LF) nor phonological (readable at PF) content and consequently must be removed before being sent to the interfaces.

Words can have features that are purely semantic and which consequently do not have any syntactic consequences. "An example of a proper semantic feature is sex (as opposed to gender), which does not trigger any syntactic operation. No feature has to be deleted, as sex can always be interpreted" (Zeijlstra 2008).

In current terminology, features can be *valued* or *unvalued* in addition to being interpretable or uninterpretable.

Let us consider valuation first. Certain features on lexical items appear to come from the lexicon unvalued, and receive their value from a valued instance of the same feature, present on another lexical item. The fact that D, N and A in [example] all bear the value *feminine* for the feature *gender* is due to a property of N - namely, the fact that the noun *puella* is listed in the lexicon as feminine.

(Pesetsky and Torrego 2007, 263)

I use the conception, terminology and notation of Pesetsky and Torrego (2007) since it seems to have become more or less standard. Their important contribution to the enterprise is that they unlink interpretability from valuation and also provide a nicely transparent notational system for features. For Chomsky (2001) (5), items with uninterpretable features are also necessarily unvalued, so Pesetsky and Torrego's suggestion is a departure from the strict one-to-one relationship between (un)valued and (un)interpretable. In their framework, both *uninterpretable* and *interpretable* unvalued features may act as Probes, whereas in Chomsky's conception only unvalued (and therefore uninterpretable, since these two are preconditions for each other) features can act as Probes. Note for the sake of consistency that it is still presumably the need for uninterpretable features to be deleted that drives the mechanism of Agree. However, unvalued features become the superficial drivers "because of the plausible consideration that the syntax has no direct access to information about interpretability, but can inspect a feature to determine whether it is valued" (Pesetsky and Torrego 2007, 269-70). Consequently, valuation and interpretability are still highly co-

dependent in Pesetsky and Torrego's framework; they just include the possibility that both interpretable and uninterpretable unvalued features can be Probes.

One remaining question, among many, is to which degree the feature system is a part of Universal Grammar. Theresa Biberauer and Ian Roberts propose a concept of emergent, rather than in-built parameters. In this conception, UG contains a feature template rather than a pre-specified set of features. The process of (first) language acquisition in this view becomes a process of creating a feature inventory for one's native language. The parameters of GB, in turn, are conceptualized as the result of bundles of features that together can look like a parameter. Parameters, then, are emergent because features are (Biberauer and Roberts 2017).

1.4 Binding theories

Language provides different strategies for establishing identity (*X and Y are the same*) between two entities. The classic example is that where *the morning star* and *the evening star* both refer to the planet Venus (and by empirical fact, then, to each other). In general terms, this is called coreference. There are no structural limitations on this sort of (accidental) reference, and speakers may not even be aware that there is coreference at all.

It is possible to establish an identity relation between two entities even when they do not semantically corefer (in a sense, the opposite situation as *the morning star = the evening star* example).

(17) No one believes they are guilty.

No one is a non-entity, not an individual, and thus cannot be the same person as *they* (in a real-life sense). However, it is completely unproblematic to read (17) in the way where the interpretation of *they* depends on *no one*. That relation is then purely linguistic, and we call it binding.

The kind of coreference that I will be concerned with is the one that is established not by empirical fact but by speaker intention (and I will also classify these as binding relations with coreference, or just binding). I use indices (together with **boldface**) to indicate that two entities in a clause are supposed to be coreferent. Just like the indices used to indicate movement in a tree (as mentioned on page 31), these indices are purely for presentational purposes and I do not commit myself to saying that they actually exist. However, indices and indexing play a meaningful part in much of the

original GB-style binding theory and great effort seems to be put into creating a modern binding theory without indices as a necessary theoretical construct. Preminger puts it succinctly and humourously in a LingBuzz manuscript when he writes that "[c]oncretely, 'X and Y share a binding index' should be understood here and throughout as shorthand for 'X and Y behave, binding-theoretically, the way that index-based theories of binding predict they would behave if they shared an index'" (Preminger 2019). I will adhere to the spirit of that quote and use the terms "share an index" or "are co-indexed" as shorthands in this thesis without commitment to the linguistic existence of indices.

The examples in (18) showcase sentences with various configurations of binding and coreference.

- (18) a. **Stephen King**₁ sometimes scares **himself**₁.
 b. After **the baroness**₁ had visited the lord, **she**₁ left the house.
 c. After the baroness₁ had visited the lord, she₂ left the house.

(18a) contains the reflexive pronoun *himself*, which, as I will establish, must be bound by and corefer with another element. This other element is called *the antecedent* and is the DP *Stephen King*. The antecedent binds the reflexive. The reflexive pronouns (such as *himself*) will be contrasted with non-reflexive pronouns, such as *she* in (18b) and (18c). *The baroness* and *she* corefer in (18b): The baroness visits the lord, and then leaves the house. There is no formal binding, presumably because neither element c-commands the other (c-command will be discussed further on page 40). In (18c), I change the index on *she* to specify a lack of coreference: The baroness visits the lord, and then some other woman leaves the house. Non-reflexive pronouns (sometimes referred to as pronominals) may corefer (as in (18b)) and they may not (as in (18c)). Just as with the reflexive pronouns, there are syntactic constraints on where and how non-reflexive pronouns may or may not corefer, and I will discuss this in detail, too.

1.4.1 Traditional binding theory a la Chomsky (*Lectures on government and binding* (LGB) 1981)

The three principles of binding in Government and Binding (GB) are defined as follows:

- (19) a. An anaphor is bound in its governing category (Principle A)
 b. A pronominal is free in its governing category (Principle B)
 c. An R-expression is free (Principle C)

(Chomsky 1981, 188)

This definition requires some unpacking as well as a few more definitions. We need definitions of *anaphor*, *pronominal* and *R-expression*. Further, we need a definition of what it means to be *bound* and *free* and finally what constitutes a governing category (which I will call the *binding domain*).

The term *anaphor* is often used in generative grammar as an umbrella term that covers both reflexive pronouns (*herself*) and reciprocal pronouns (*each other*). Anaphors depend on other sentence elements for their reference and content, which is essentially what Principle A formalizes. However, as explored as early as Lebeaux (1983), reciprocal and reflexive pronouns do not pattern in precisely the same ways. A striking difference is that a reciprocal may occur as the subject of a finite clause both in Danish and English. A reflexive used in the same context is usually ungrammatical. Compare the ungrammatical examples in (20)-(21) with reflexive subjects and the grammatical equivalents in (22)-(23) with reciprocal subjects. I gloss Danish simplex reflexive *sig* as *REFL* (shorthand for *reflexive*) since English has no direct equivalent. For the same reason, I gloss Danish possessive reflexive *sin* as *REFL's* and complex reflexive *sig selv* as *REFL self*. I follow standard conventions in using an asterisk * to indicate that a sentence is ungrammatical.

- (20) * **Johan**₁ vidste ikke hvad **sig** **selv**₁ havde gjort.
 Johan knew not what REFL self had done
 'Johan didn't know what himself had done.'
- (21) * **John**₁ didn't know what **himself**₁ had done.
 (Lebeaux 1983, 724)
- (22) **Johan og Marie**₁ vidste ikke hvad **hinanden**₁ havde gjort.
 Johan and Marie knew not what each-other had done
 'Johan and Marie didn't know what each other had done.'
- (23) **John and Mary**₁ didn't know what **each other**₁ had done.
 (Lebeaux 1983, 724)

As a consequence, I use the term *reflexive* rather than *anaphor* in order to explicitly exclude reciprocals.

Pronominals are non-reflexive pronouns, e.g. English *him* and Danish *ham*. They *may* still get their reference from other sentence elements under specific circumstances. R-expressions cover all the left-over non-pronominal, non-reflexive nouns. These can be e.g. proper names like *Harry Potter* or a noun phrase like *the gifted orphan*, which are both fully referential in themselves and usually do not get their reference from other noun phrases. Also note that Principle C mentions nothing about a governing category. It applies everywhere and R-expressions must never be bound, regardless of the syntactical configuration. I will not discuss Principle C in much detail, but just point out that sentences like "Mary loves Mary" and similar counter-examples to Principle C have been discussed in great detail in the literature and explained away or accepted as problematic ad nauseum. One way of dealing with them is to say that the two Marys are not completely the same entity and that the second Mary does not (fully) get its reference from the first Mary. Another way of dealing with them is to say that Principle C should be done away with.

I illustrate some basic distributional differences between reflexives, non-reflexives, and R-expressions in English in the examples in (24)-(26). Co-reference is indicated by means of indices and **boldface**. The reflexive pronoun is given in the a. examples, the non-reflexive in the b. examples and full DPs in the c. examples.

In (24), only the reflexive pronoun *himself* is grammatical in a sentence with a local antecedent and intended coreference. Neither non-reflexive *him* nor full DP *Harry* permit this.

- (24) a. **Harry**₁ likes **himself**₁.
b. ***Harry**₁ likes **him**₁.
c. ***Harry**₁ likes **Harry**₁.

This contrasts with (25), where the reflexive is the only one that is not grammatical. *Himself/him/Harry* is the only referential DP in the sentences, since the subject is expletive, non-referential, *it*. Consequently, there is nothing in the sentence for *himself* (or *him* or *Harry*) to corefer with and be bound by. This is clearly an issue for the reflexive pronoun, whereas the non-reflexive and the full DP work perfectly well without being bound.

- (25) a. *That it rains bothers himself.
b. That it rains bothers him.
c. That it rains bothers Harry.

In (26), the DPs are bound by a non-local antecedent (by non-local I initially just mean an antecedent that is not the closest subject). This is not possible with the reflexive and with the full DP, but fine with the non-reflexive *him*.

- (26) a. * **Harry**₁ thinks that I like **himself**₁.
 b. **Harry**₁ thinks that I like **him**₁.
 c. * **Harry**₁ thinks that I like **Harry**₁.

At a first glance, we see that both reflexives and non-reflexives can be bound, but in different ways (contrast (24) and (26)). Non-reflexive pronouns and full DPs can be free and reflexive pronouns cannot, (25). Finally, full DPs can only be free and never bound (bound full DPs disallowed in both (24c) and (26c)).

From a superficial point of view, (27) seems to contradict the conclusions that we reached above. In (27)a, reflexive *himself* corefers with *Harry*, a local antecedent, but the sentence is ungrammatical. Similarly, non-reflexive *him* in (27)b and full DP *Harry* in (27)c seem to corefer with local antecedents and the sentences are still grammatical. *Him* should not be able to be bound by a local antecedent, and *Harry* should not be able to be bound at all).

- (27) a. * **Harry**₁'s mum saw **himself**₁ in the mirror.
 b. **Harry**₁'s mum saw **him**₁ in the mirror.
 c. **Harry**₁'s mum saw **Harry**₁ in the mirror.

To explain these data, I need a definition of what exactly constitutes binding, and Chomsky provides one: " α is X-bound by β if and only if α and β are coindexed, β c-commands α , and β is in an X-position" (Chomsky 1981, 184). He works with concepts of both A-binding and A'-binding, which is the reason for calling it X-bound (as opposed to X'-bound). Binding, then, requires coindexation, c-command and that both antecedent (binder) and bindee are in the same type of position (A or A'). The last point will not be explored further. Additionally, if and only if something is not bound, it is free (Chomsky 1981, 185). This definition of binding is from before indices were removed from the theory, notably.

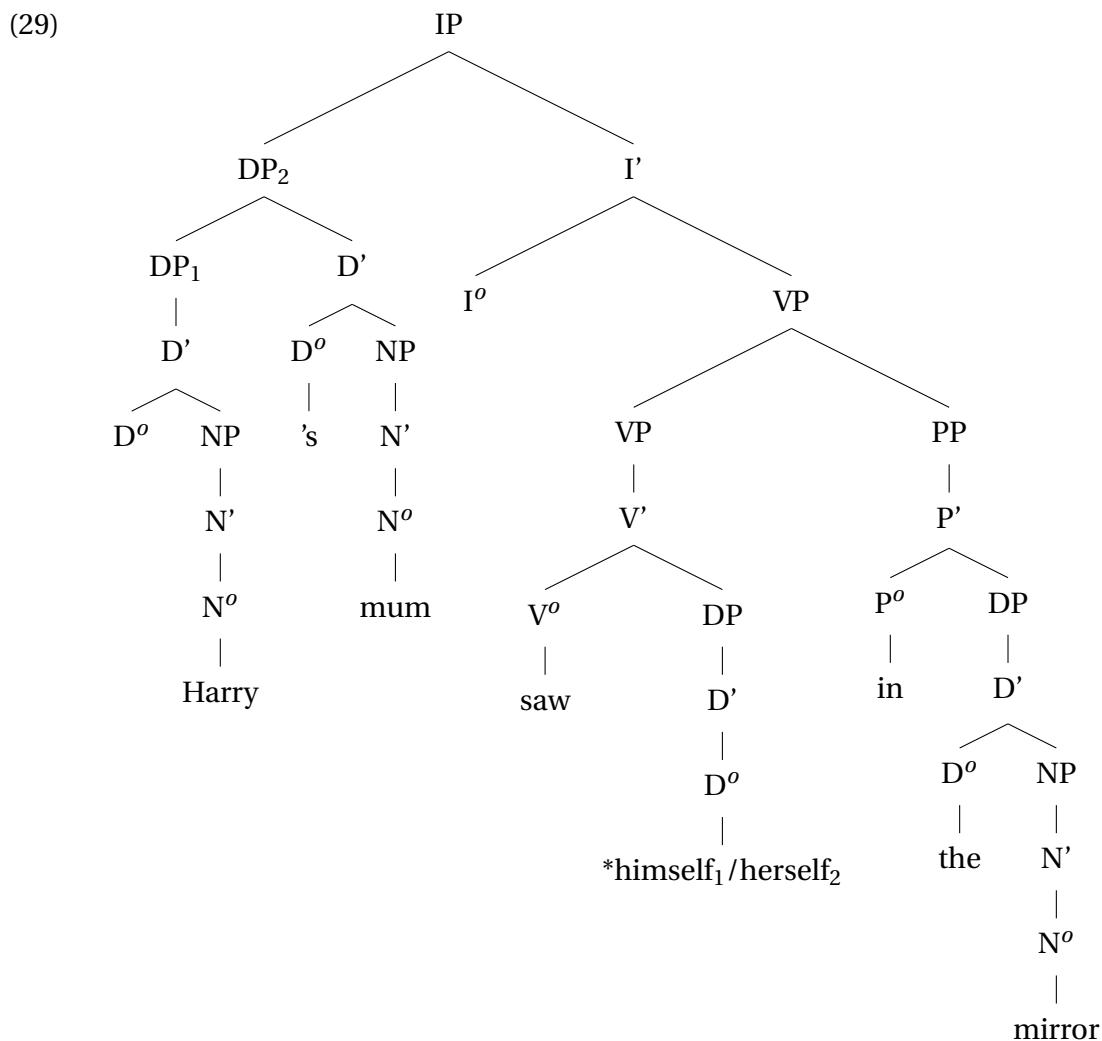
C-command is a specific structural relationship between two nodes in a tree. In terms of binding, it seems that the reflexive must be c-commanded by its antecedent in order for binding to be possible. I adopt Vikner's definition of c-command and reproduce it in (28).

- (28) X c-commands Y if and only if
- a. all nodes that dominate X also dominate Y,
 - b. X does not dominate Y, and
 - c. Y does not dominate X

(see e.g. Vikner (2011, 42))

A more accessible way of expressing the same relation is that "if you can get from X to Y in the tree by taking one step upwards and then climbing downwards the rest of the way, then X c-commands Y" (Vikner 2011, 42).

The problem that prevents proper binding in (27a) is that *Harry* does not c-command *himself*. You cannot get from *Harry* to *himself* by taking one step upwards since *Harry* is contained in the larger DP *Harry's mum*. The tree in (29) illustrates the situation. It is not possible to go one step up from the DP that only contains *Harry* and reach the reflexive by going downwards. It *is* possible to do so with the entire DP, *Harry's mum*, however, and the sentence would have been grammatical with reflexive *herself* bound by *Harry's mum*.



The last outstanding definition is *governing category*. Chomsky's definition of governing category in *Lectures on government and binding* is as follows: " β is a *governing category* for α if and only if β is the minimal category containing α , a governor of α , and a SUBJECT accessible to α " (Chomsky 1981, 211). I will use the term *binding domain* as an alternative to *governing category*. It is not necessarily the case that the specific definition of the binding domain (or governing category) should be the same across languages. In fact, there is ample evidence that e.g. Scandinavian and English reflexives have different constraints in terms of where they can and cannot be bound (e.g. Vikner (1985), Ehlers and Vikner (2016), Vikner and Ehlers (2017), and Ehlers (2017) for Danish and Hellan (1988) for Norwegian).

I repeat Principles A to C with my revised choice of wording below, give a hopefully also slightly more accessible definition of the binding domain (see Ehlers (2017, 72-75) for the argument for that specific definition), and a definition of binding in terms

of coindexation.

- (30) a. **Principle A:** A reflexive pronoun must be bound inside its binding domain.
 b. **Principle B:** A non-reflexive pronoun must be free inside its binding domain.
 c. **Principle C:** A full DP must be free.
- (31) The **binding domain** for X is the minimal IP/DP containing:
 a. X and
 i. a finite verb or
 ii. a subject that c-commands X
- (32) X **binds** Y if and only if
 a. X and Y are coindexed
 b. X c-commands Y

Intensifiers: English *herself* and Danish *selv*

After having established that reflexives such as *herself* must be bound, I have a note on another use of *herself* that looks like a reflexive but does not pattern as a reflexive.

- (33) a. **She**₁ buys **herself**₁ a house.
 b. She buys a house herself.

In (33a), *herself* is a reflexive pronoun. *Herself* is coreferent with a preceding, c-commanding DP (*she*) within the same clause, and is an argument of the verb *buys*. In (33b), *herself* is not a reflexive pronoun but an intensifier with the same phonological form as the reflexive. It is an adjunct rather than an argument of the verb and its function is to emphasize a contrast (she buys a house on her own, rather than e.g. with someone else). Only (33a) says anything about who receives the house (she does); in (33a), she could in principle be buying the house for someone else.

- (34) a. She smiles herself.
 b. ***She**₁ smiles **herself**₁.
 c. *She smiles a rocket.
 d. She smiles on the train.

The verb *smile* is normally intransitive (in the British National Corpus (BNC), *smiles+noun* gives 7 hits, and the only hit that could possibly be transitive is *smiles encouragement*). In spite of this, it is possible to have *herself* in what looks like the direct object position of *smile* in (34a). *Smile* with a regular direct object is ungrammatical (or at least requires some imagination to make sense of), as in (34c), and just as ungrammatical in (34b) where the intended reading is that of reflexive argument *herself*. (34d) is perfectly fine, with *smiles* accompanied by the adjunct *on a train*, and this would be the same interpretation that I will give to (34a), i.e. that *herself* in (34a) is an adjunct and an intensifier rather than an argument and a reflexive pronoun.

Translating the *herself* sentences to Danish makes the difference even more visible.

- (35) a. Hun smiler selv.
 she smiles self
 'She smiles herself.' (intensifier *herself*)
- b. *Hun smiler sig selv.
 she smiles REFL self
 'She smiles herself.' (reflexive *herself*)

Intensifier *herself* is the intensifier *selv* in Danish, and reflexive pronoun *herself* is reflexive pronoun *sig selv* (or *hende selv*, given a different syntactic configuration). English intensifier *herself* and reflexive *herself* are very different in terms of function and distribution, and I will view them as two separate lexical entries (if homophonous) rather than try to incorporate intensifier *herself* in a binding theory. Sørensen, Ehlers, and Vikner (2020) discusses overrides in English and Danish further.

1.4.2 Newer approaches to binding

Consider first that fact that under the standard theory the definition of 'bound' involves two nominal expressions in a c-command relation that are coindexed. Under minimalist assumptions, however, indices and similar devices are not available. [...] Obviously if we eliminate indices as a grammatical device, then binding theory must be recast in some other way since the standard theory is to a large extent a theory about the assignment of indices.

(Freidin 1997, 147)

In his *The minimalist program*, Chomsky speculates briefly that "Condition A may be dispensable if the approach based upon cliticization_{LF} is correct and the effects of Condition A follow from the theory of movement" (Chomsky 1995, 194), where cliticization_{LF} according to Norbert Hornstein also essentially just refers to movement (Hornstein 2007, 351). This idea has turned out to be quite popular. A large number of linguists have tried to come up with different ways of either altering the original Binding Theory to fit better with the framework provided by Minimalism, or to get rid of Principle A and B entirely. Reasons for wanting to do so are many. One is to get rid of the notions of government and governing category which are both "suspect notion[s] in MP" (Hornstein 2000, 153). Another reason is the wish to remove indices and traces as objects of the theory, as in the quote from Freidin cited above. Yet another is a wish to remove the special significance given to reflexives in the theory (Hornstein 2000, 154). I discuss some of these proposals here.

Binding in terms of movement

The first group of proposals falls under the general heading of *Movement*, which is in line with how Chomsky suggests Principle A can be dispensed with. I use work by Richard Kayne and Norbert Hornstein who have both made specific suggestions to this effect. An attractive feature of viewing binding as movement is that movement has long been established as a local relationship that has to apply in small steps in order to reach further. This is to some extent the same thing that characterizes binding. Movement is independently needed in the theory and a reduction of binding to movement would fit well with a Minimalist ideal.

Hornstein's arguments against traditional Binding Theory are collected under three headings: that Binding Theory is "suspect on methodological grounds given minimalist commitments" (Hornstein 2000, 153); that Principle A is empirically and theoretically unnecessary and should be dispensed with in favour of anaphors as "the residues of overt A-movement" (ibid.); and that Principle B should go the same way (ibid. and Hornstein 2007). His alternative suggestion is that local reflexive pronouns are the result of overt movement of the antecedent. The reflexive itself is the lowest copy in a movement chain. This also means that the reflexive is not included in the initial Numeration but rather formed in the course of the derivation. The general idea is illustrated in (36) where the lower copy of *John* is realized as the reflexive *himself*. This, according to Hornstein, makes Principle A unnecessary.

- (36) a. Intended clause: *John likes himself*
 b. A possible Numeration: {John, likes}
 c. (Parts of) the proposed derivation:
- i. likes John
 John likes ~~John~~
 - ii. John likes himself
- (Hornstein 2007, 359 and Hornstein 2000, 159)

Hornstein's second objective is to remove Principle B as well. He reasons that "if principle A is eliminated then principle B should be as well. Why so? The reason lies with the central empirical fact about pronouns; they, when bound, are in complementary distribution with reflexives" (Hornstein 2007, 354). Interestingly, this "central empirical fact" is not true for Danish (this will be elaborated further in chapter 2). The article that Hornstein cites for this central empirical fact even gives a few examples that show this non-complementarity (Safir 1997, 351).

I ignore this issue for the moment to pursue Hornstein's argument. His suggestion is that Principle B and (some) complementarity between reflexives and pronominals can be attributed to "the effects of derivational economy" (Hornstein 2007, 382) where movement (i.e. reflexives) is cheap and pronominal insertion costly. This explains complementarity between reflexives and pronominals since pronominals are only inserted when reflexives cannot be. From this follows that pronominals must be formed derivationally and not inserted from the Numeration, just like he suggests for reflexives. The alternative view would violate Chomsky's *Inclusiveness Conditions* that prevents insertion of lexical material that is not contained in the Numeration. Additionally, if pronouns and reflexives enter the derivation in different ways, they would be unable to compete and there would be no economy-based reason for preferring one over the other. See Safir (2004) for a greatly expanded version of this point. See Ackema and Neeleman (2013) for a specific instantiation of a system of derivationally created pronouns.

Kayne's proposal addresses Hornstein's directly. For Kayne, as opposed to Hornstein, the Numeration includes the pronoun. In entering the derivation, the pronoun "derivationally form[s] a constituent with its antecedent" (Kayne 2002, 134). The antecedent moves out of this constituent, leaving behind a trace and the pronominal or reflexive that makes up the other half of this constituent. According to Kayne, this can cap-

ture all three of Principles A, B, C. These can then all be eliminated from UG by being reduced to movement. I summarize the suggested movement patterns in (37).

- (37) a. Intended clause: *John thinks he is smart*
 b. A possible numeration: {John, he, thinks, is, smart}
 c. (Parts of) the proposed derivation:
 i. thinks [John he] is smart
 ii. John_i thinks [t_i he] is smart
 (Kayne 2002, 135)

Kayne calls [John he] a doubling constituent. He follows the analysis from Uriagereka (1995) where (clitic) doubling phrases are DPs with the double in DP-spec and the clitic in D^o. *John*, then, is in the specifier position and *he* perhaps in the head. This establishes a very local permanent relationship between antecedent (trace in specifier position) and pronoun (head), which then is what Kayne uses to account for binding. In (38), I summarize Kayne's proposal for how this approach to binding will get rid of Principles A-C.

- (38) a. Condition A: John likes himself
 i. Why? "*self* makes available an intermediate position for the pronoun that is not available in the absence of *self*" (Kayne 2002, 147). The pronoun part of the *self* reflexive then presumably stays in this intermediate position and somehow merges with *self* which then explains why binding of *self* reflexives behaves differently than binding of pronouns.
 (Kayne 2002, 147)
- b. Condition B: a) John thinks he is smart and b) *John likes him
 i. Why? The unstressed pronoun in the doubling constituent moves and pied-pipes the entire constituent with it. In a), *John* then moves up further and gets a theta role. In b), there is nowhere for *John* to move that would get him a theta role and the derivation crashes.
 (Kayne 2002, 145-46)
- c. Condition C: He thinks John is smart
 i. Why? The pronoun cannot move out of the doubling constituent
 (Kayne 2002, 137)

Kayne briefly mentions what he calls *zich* type reflexives, i.e. reflexives like Dutch *zich* and Danish *sig* that cannot be bound locally (except for inherently reflexive contexts like *skamme sig*) (Kayne 2002, 148-50). His solution to these kinds of reflexives is less clear than the rest of the article in that he introduces an unspecified item *DB* to be the double for the reflexive where the *DB* is then "in turn related to a DP [...] in ways that I will not explore any further here" (Kayne 2002, 150). In short, Kayne's suggestions for how non-local binding could take place within his own framework are rather limited in scope. Hornstein, similarly, mentions briefly in a footnote that it would be interesting and necessary to extend the analysis to e.g. Scandinavian possessive reflexives (like Danish *sin*) but does not do so himself (Hornstein 2007, 352).

It does seem like both *sin* and non-local reflexives would present an issue for the Movement-based approaches argued here. I only speculate briefly as I cannot hope to do the topic justice here. For *sin*, it cannot be said to just be a lower copy of its antecedent (Hornstein's approach): it must have some added *possessiveness* in addition to the features of the antecedent. In the terms of Kayne, the doubling constituent with *sin* would presumably need to contain *sin*'s complement, which at least makes the relationship between antecedent and reflexive less direct than for e.g. *sig* which does not take a complement. For non-local reflexives, we would presumably need to propose a specifier position (in addition to IP-spec of the embedded clause) that the antecedent can use to move out of the embedded clause to the matrix clause.

Binding in terms of features and Agree

The second group of approaches uses the assumption that reflexives are somehow feature-deficient, and the mechanism of Agree and feature valuation, to explain binding. As exponents for this approach, I use Reuland (2001, 2005, 2011, 2014) and Rooryck and Vanden Wyngaerd (2011, 2015). Note that e.g. Ken Safir argues against a binding theory only in terms of Agree: "In Safir 2010:97, I make an explicit argument for distinguishing Agree from binding on the basis of certain Icelandic constructions in which an oblique argument is bound by a nonnominative subject, while the verb in the same clause agrees with a nominative nonsubject" (Safir 2014, 111).

I begin with Reuland. "As is uncontroversial, SE-anaphors such as Dutch *zich*, Icelandic *sig*, etc., are not specified for the features number and gender. They are specified for person. Reuland (to appear) discusses various types of underspecification and concludes that SE-anaphors carry number and gender features that are unvalued in

the sense of Chomsky (2001, 2005) and P&T" (Reuland 2005, 510). Reuland's specific proposal for binding is that the binding relation is established through Agree (in the terms of Pesetsky and Torrego, described on page 34). Antecedent and reflexive both Agree (feature sharing and valuation) with a functional head and this process is what drives binding.

It may not be entirely uncontroversial that SE-anaphors are underspecified in the way that Reuland claims. Danish possessive reflexive *sin*, which I assume should be counted as an SE anaphor, usually only takes singular antecedents, which could indicate that it *is* specified for number. In the 2014 paper, Reuland writes that the difference between pronominals and reflexives is that "they [pronominals] are specified for grammatical number" (Reuland 2014, 12). If this is the only difference between anaphors and pronominals, Danish *sin* being specified for number could possibly constitute a counter example to Reuland's description.

Rooryck and Vanden Wyngaerd also use possessive pronouns in their argument against Reuland's approach. They point out that the issue with Agree-based theories of binding a la Reuland's "is that they will not work for cases involving possessive pronouns, since they are embedded in the object DP and do not enter into an Agree relation with ν " (Rooryck and Wyngaerd 2011, 50). Their own version of a binding theory is also Agree-based but with different mechanisms than Reuland's. Their idea is that reflexive pronouns are DPs that "enter the derivation with unvalued ϕ -features, which they need to value under Agree with an antecedent. This leads to feature sharing, which is semantically interpreted as referential dependence. Pronouns are minimally different in that they enter the derivation with valued ϕ -features, leading to an interpretation of disjoint reference at the interface" (Rooryck and Wyngaerd 2011, 2). Their specific proposal has several legs and part of this is that they view simple reflexives like Dutch *zich* (which they use as their main expository example) as being fundamentally different from complex reflexives like *zichzelf* as "[c]omplex reflexives are merged as the internal argument of a transitive verb" (Rooryck and Wyngaerd 2011, 54-5).

Conversely, "the simplex reflexive *zich* is a DP that is merged as the possessum in a possessive constituent that also hosts its antecedent, represented as the possessor" (Rooryck and Wyngaerd 2011, 54). The Agree relation, then, is between possessor (the specifier) and possessum (the complement) where the unvalued features of *zich* are valued and shared by the valued features on the antecedent. Non-local reflexives "constitute a more recalcitrant problem, which the literature provides no adequate ac-

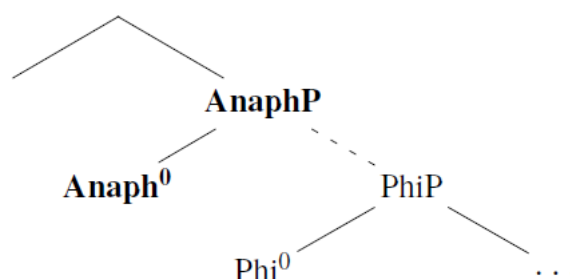
THE φ -ENCAPSULATION HYPOTHESIS

Figure 1.22: Preminger's suggestion for a universal structure of anaphors. They are anaphoric precisely because they contain an anaphoric layer on top of the ϕ -bearing (person, number, gender) layer. Illustration from Preminger (2019, 2).

counts of, and which we shall also have to leave at least partially unresolved" (Rooryck and Wyngaerd 2011, 154). Finally, they define a regularity, *Absence of Principle B Effects* where "Pronouns behave like anaphors when a dedicated class of reflexive pronouns is lacking" (Rooryck and Wyngaerd 2011, 19). This would also account for the first and second person accusative pronouns in Danish (e.g. *mig*) that can be either reflexive or pronominal, as well as reflexive or pronominal *his* in English.

A newer take *against* binding as Agreement is provided by Omer Preminger (e.g. in the still-unpublished Preminger (2019)). He argues that a reduction to binding as Agreement cannot be the right solution. Part of his argument is the observation that feature-matching between antecedent and anaphor is neither necessary nor sufficient for binding and further that feature-matching does not necessarily have to entail a syntactic agreement process. He proposes an approach to binding through *encapsulation* where the a pronoun becomes anaphoric because it is contained within an additional anaphoric layer. I mention Preminger because the binding variation that this thesis deals with precisely seems to be a case of featural mismatch between antecedent and anaphor.

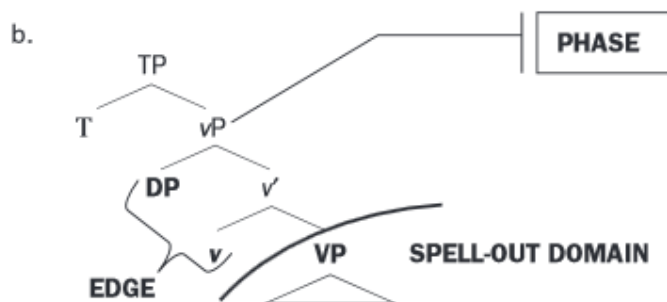
Chomsky 2008 uses Reuland's Agree-based approach to binding (Reuland 2001) to develop a first attempt at a phase-based theory of binding and also agrees with Reuland in speculating that possibly c-command is not necessary for a binding relationship to exist (Chomsky 2008, 142).

Binding in terms of phases

Another newer approach to binding is to cast it in terms of phases. This approach, just like the movement one, aims to reduce the locality condition on binding to other, independently motivated locality conditions. Phases, as introduced in Chomsky (2000) and Chomsky (2001), precisely split the derivation into smaller bits. This seems to fit well with the locality conditions posed by binding theory.

These smaller bits start out as partitions of the Numeration, called Subarrays. Instead of having access to the entire Numeration, the derivation incrementally builds smaller, self-contained structures from the Subarrays. These structures are the *phases*. One motivation for introducing phases is that they presumably reduce the processing load: "[o]perative complexity in some natural sense is reduced, with each stage of the derivation accessing only part of LA" (Chomsky 2008, 106). There is in principle no limit to how long and complex a grammatical sentence can be, and a phase-based derivation reduces this potentially infinite complexity.

When the phase is finished (what this means is in itself contested), parts of it are sent to the PF and LF interfaces. This process is called Spell-Out. After Spell-Out takes place, the Spelled-Out parts of the phase are no longer available (the Phase Impenetrability Condition, which also comes in different flavours, weak and strong. See Citko (2014) for discussion). This means that all parts of a phase that need to be available to the syntax later in the derivation must leave the Spell-Out domain before Spell-Out takes place. I illustrate the general make-up of a phase below, borrowed from Citko.



(Citko 2014, 32)

Opinions differ as to which phrases that may be phases. Chomsky himself originally suggested that phases may be "the closest syntactic counterpart to a proposition: either a verb phrase in which all [theta]-roles are assigned or a full clause including tense and force" (Chomsky 2000, 106). This would correspond to CP and (most) vPs as phases. Later, he opens the door for possibly including DPs as phases, too (Chomsky 2008, 143).

The general idea of the phase-based approaches is that phases are the binding domains. There are several different published attempts at making binding phase-based and I discuss a number of these in this section. They agree on the status of phases as binding domains but otherwise differ quite a bit, also in their proposed mechanism for taking care of binding without indices. They have different ways of dealing (or not dealing) with non-local binding. Finally, they disagree in their definitions of which phrases that can function as phases. In the original proposal from Chomsky, CP and vP (termed nP in the particular paper) are suggested as phases but "neither finite TP nor unaccusative/passive verbal phrase is a phase" (Chomsky 2000, 107).

Canac-Marquis (2005) is to my knowledge the first attempt to capture binding theory in terms of phases. The paper is from a conference proceedings, appears somewhat unfinished and only a five-page abstract version of it was peer-reviewed. He chooses to define (potential) phasehood through Case assignment. This opens up for several more potential phase types than what Chomsky suggests (TP, DP and PP, to name a few) and conversely does not necessarily see CP as a phase.

Canac-Marquis' final definition of a phase-based revision of Principle A is that "A reflexive must be bound in its phase only if there is an antecedent in the phase" (Canac-Marquis 2005, 500). This both under- and overgenerates. In principle, it predicts a sentence like **I know that I like himself* to be grammatical which is not desirable. The paper does not address how binding would take place, i.e. what the motivation for binding would be in the first place. Finally, the only language investigated is English where non-local anaphors do not occur.

Lee-Schoenfeld (2008) suggests that CP, agentive vP, complex DP (DP with a possessor in DP-spec) and θ -assigning PP are phases precisely because these appear to constitute binding domains. This makes her argument somewhat circular. The intention of the paper also seems more in line of arguing for binding (and possessor raising) as diagnostic tools rather than for arguing for a specific instantiation of phase-driven binding.

She argues on the basis of non-complementarity between non-locally bound reflexives and pronominals in German that reflexives may move (covertly) to the phase edge and be bound in the higher phase (Lee-Schoenfeld 2008, 290-91). This is driven by what looks like a binding-specific checking mechanism: "condition A checks whether any reflexive contained in the domain that is being evaluated has a potential antecedent" (Lee-Schoenfeld 2008, 291). The reflexive may move to the phase edge if

it does not have a potential antecedent, perhaps driven by the need to check features.

Quicoli (2008) adopts the definition of binding domain developed in Chomsky (1981) and later works. The binding domain is defined in terms of accessible SUBJECT (I give the precise definition on page 41). SUBJECT is either the normal definition of subject, i.e. whatever is in IP-spec or DP-spec, or, alternatively, finite verbal agreement. He notes in a footnote that this definition could just as well be defined in terms of phases, where vP (or DP, presumably, although he does not mention it) contains a regular lower-case subject, and CP constitutes a domain in terms of finiteness (Quicoli 2008, 301). Arguably, this definition could just as well include IP as a binding domain as well. However, in Chomsky (2008) it is convincingly argued that TPs (IPs) should not be seen as phases, and consequently not as binding domains (143-44).

Quicoli's suggestion for how binding ties up with phases "is that condition A applies cyclically at the end of each phase" (304), i.e. that binding is an independent procedure that now just applies at phase ends rather than at specific levels of representation for the finished sentence. He illustrates the process with the example in (39).

- (39) a. [_{VP} They v [_{VP} respect each other_i]]
 b. [_{VP} They_i v [_{VP} appeared to each other_i [_{TP} to [_{VP} ~~they~~_i [_{VP} respect each other_i]]]]]
 (Quicoli 2008, 305)

Antecedent *they* binds both reciprocals in (39b). Principle A applies cyclically at the vP phases, presumably while the antecedent is still available to the phase (i.e. before Spell-Out of the non-edge domain). He suggests that principles B and C can be stated in the same terms.

Despić (2015), in contrast to the other phase-based proposals, argues for including DPs as phases and binding domains and, most interestingly for my purposes, uses the Danish possessive reflexive *sin* as a case study for the argument.

1.4.3 Summary

In essence, the topic of binding has gone from being a reasonably unified and broadly explanatory theory to the current, very much not unified and less broadly explanatory field of study. The aim of this thesis is not to choose a side or to argue that one binding theory is better than the other. Rather, I focus on a small part of a small lan-

guage in order to hopefully further advance our understanding of what is possible and impossible in terms of binding and morphosyntactic (micro)variation.

Chapter 2

Danish language and Danish reflexives

2.1 The *sig* and *sin* reflexives in the Germanic languages

I start with a very quick recap of the main points about reflexive binding which were explored in more detail in chapter 1.

Some languages like English and Danish have a set of reflexive pronouns. These can be characterized in terms of binding, a relationship that holds between an antecedent and the reflexive or non-reflexive pronoun. I established on page 40 that the antecedent must c-command and be coindexed with the reflexive (or non-reflexive) pronoun in order for binding to occur.

In English, the reflexive pronouns have the form *X-self*, and they differ from regular pronouns in that they must be bound in a local relation with e.g. a subject (40a, Principle A). Non-reflexive pronouns must be free in this same relation (40b, Principle B).

- (40) a. He₁ loves himself₁.
b. *He₁ loves him₁.

Danish has a rather more complicated system of reflexives than English does, with more divergent elements (reflexives *sig* and *sig selv*, non-reflexive *hende*, *hende selv*, reflexive possessive *sin*, and non-reflexive possessive *hendes* to mention a representative few) and more constraints. I introduce the Danish (and more generally Germanic) reflexives specifically in the next subchapters.

2.1.1 Reflexive possessive *sin* in the Germanic languages

In addition to the English-type *-self* reflexives, Danish has a reflexive possessive pronoun *sin* that allows speakers to distinguish between the two possible readings of the Danish equivalent to English *she loves her cat*. In English, this sentence, as given in (41), can in principle either refer to her own cat, the bound reading in (41a), or someone else's cat, the unbound reading in (41b). The bound reading, *her own cat*, requires possessive reflexive *sin* in Danish, as in (42a). The non-bound reading, another person's cat, is not possible with *sin*, (42b). Conversely, the bound reading is not possible with non-reflexive *hendes*, (42c), but the unbound version is perfect, (42d).

- (41) a. **She**₁ pets **her**₁ cat.
 b. She₁ pets her₂ cat.
- (42) a. **Hun**₁ aer **sin**₁ kat.
 she pets REFL's kat
 'She pets her (own) cat.'
- b. ***Hun**₁ aer **sin**₂ kat.
 she pets REFL's kat
 'She pets her (another person's) cat.'
- c. ***Hun**₁ aer **hendes**₁ kat.
 she pets her kat
 'She pets her (own) cat.'
- d. **Hun**₁ aer **hendes**₂ kat.
 she pets her kat
 'She pets her (another person's) cat.'

The possessive reflexive Danish *sin* is a reflex of a similar form which can be traced back to (at least) Common Germanic. Both Danish and English are part of the language subfamily that we call the Germanic languages (the language family is Indo-European). Icelandic, Faroese, Swedish, Norwegian, and Danish are grouped together as the Scandinavian (= North Germanic) languages. Danish, Swedish, and Norwegian are further subgrouped as the Mainland Scandinavian languages (as opposed to Insular Scandinavian which covers Icelandic and Faroese). English, Frisian, Dutch, Afrikaans, Low German, German, and Yiddish are the West Germanic languages. Gothic (and a few others, all extinct) is an East Germanic language. In terms of timing, the Goths may have migrated from Scandinavia into Eastern Europe in the beginning of the first

century AD. The main source of the original Gothic language is a 6th century copy of a 4th century Bible translation allegedly made by Wulfila (literally: "Little Wolf"). The last reports of speakers of a language related to Gothic are from the 17th century. The text corpus from Gothic is the earliest textual evidence of a Germanic language, and we use Gothic as a way of approximating the earlier Common Germanic stage that predates the individual Germanic languages.

Most of the Germanic languages, except for English, have retained a cognate of the Gothic possessive reflexive *seins*. In Gothic, *seins* (glossed as REFL's in (43a)) allowed third person antecedents of any gender and number and was inflected like an adjective, just like the Latin possessive reflexive *suus*, in (43b)¹.

(43) a. **Gothic**

hairdeis	sa	goda	saiwala
shepherd.NOM.SG	the.MASC.NOM.SG	good.MASC.NOM.SG	life-FEM.ACC.SG
seina	lagjiþ	faur lamba	
REFL's-FEM.ACC.SG	lays-down for	sheep	

'the good shepherd lays down his life for the sheep' (wulfila.be, John 10:11)

b. **Latin**

Bonus pastor	animam	suam	dat pro
good shepherd	life.FEM.ACC.SG	REFL's.FEM.ACC.SG	gives for
ovibus	suis.		
sheep.FEM.DAT.PL	REFL's.FEM.DAT.PL		

'the good shepherd lays down his life for his sheep'

Sin has been retained as a possessive reflexive in Danish and the other Scandinavian languages but has either been lost (in e.g. English) or repurposed in the other Germanic languages.

In the continental West Germanic languages, with German *sein* as the specific case, *sein* is a reflex of the Common Germanic possessive reflexive but it is not a reflexive pronoun. It has been repurposed as a possessive personal pronoun of masculine and neuter gender. The corresponding pronouns for feminine and plural are variations of the possessive pronoun *ihr* and none of the languages have preserved or

¹However, according to Politzer 1952, 67, there was a tendency to use possessive reflexive *suus* (and its inflectional variants) mainly with singulars (and *eorum*, non-reflexive plural pronoun, with plurals) which can be seen in Latin texts back to the 7th century. In the later Romance languages, some languages extended *suus* to both singular and plural, and some languages limited *suus* to the singular only.

7 se froda swa þeah befæste þ rice
 heahþunenum menn Haroldde sylfum
 æþelum eorle se in ealle tid
 hyrde holdlice hærran sinum.

Figure 2.1: Literary translation of the piece of text from the Anglo-Saxon Chronicle: "But the prudent king had settled the realm on high-born men, on Harold himself, the noble earl, who in every season faithfully heard and obeyed his lord" (accessed through Project Gutenberg through <https://archive.org/stream/theanglosaxonchr00657gut/angsx10.txt>)

innovated another reflexive possessive form (see Allen 2008, 41 who cites Lockwood 1968 for the specifics on German). The examples in (44) show how cognates of *sein* show up in the continental West Germanic languages as non-reflexive possessives.

- (44) a. Er isst sein Essen. (German)
 Hij eet zijn voedsel. (Dutch)
 Hy eet sy kos. (Afrikaans)
 Er est zeyn esnvarg. (Yiddish)
 Hy yt syn iten. (Frisian)

'He eats his food.'

There are remnants of a *sin* in a few Old English texts but the form "occurs mostly in poetry, rarely in prose, and it does not survive into ME" (Mustanoja 1960, 156), although Jane Roberts (Roberts 2016, 54) speculates that *sin* will have been part of the vocabulary of the 10th century English scribe Aldred. The example in figure 2.1 is as late as 1065 (from Mitchell 1985, 119).

A common denominator for the languages that have lost or repurposed possessive reflexive *sin* is that they all have prenominal definiteness markers, whereas the languages that have preserved *sin* as a reflexive all have postnominal definiteness markers (or no definite articles at all). This is the case for the Scandinavian languages where *sin* is preserved as a possessive reflexive. *Sin* must be bound by a subject and cannot occur unbound in the Scandinavian languages. In all the languages, *sin* can be bound by third person antecedents of any gender, but only in Danish is *sin* limited to third person singular antecedents. This point is visible in the parallel examples in (45) where Danish is the only language which does not permit the plural third person pronoun to bind *sin*.

- (45) a. **Han₁/hun₂/de₃** elsker **sin_{1/2/*3}** nye bil. (Danish)
Han₁/hon₂/de₃ älskar **sin_{1/2/3}** nya bil. (Swedish)
Han₁/hun₂/de₃ elsker nye bilen **sin_{1/2/3}**. (Norwegian)
Hann₁/hun₂/tey₃ elska(r) **sín_{1/2/3}** nýggja bil. (Faroese)
Hann₁/hún₂/þau₃ elska(r) nýja bílinn **sinn_{1/2/3}**. (Icelandic)
 he/she/they love REFL's new car
 'He/she/they love(s) his/her/their new car.'

The third person plural possessive *deres* (En. *their*) is used instead of *sin* with Danish plural antecedents. The contrast can be seen in the examples in (46) and I describe this *sin* being number-restricted.

- (46) a. ***De₁** elsker **sin₁** nye bil. (Danish)
 They love REFL's new car
 'They love their new car.'
- b. **De₁** elsker **deres₁** nye bil. (Danish)
 They love their new car
 'They love their new car.'

The observation that this thesis revolves about is that this may be changing, or at least that *sin* seems to be used fairly frequently with plural antecedents in Danish. Some relatively recently heard and read examples are reproduced in (47).

- (47) a. **Rigtig mange₁** har ikke mulighed for at passe **sine₁** børn
 very many have not opportunity for to care-for REFL's children
 hjemme. (facebook, 11/4-19)
 at-home
 'Lots of people don't have the opportunity to take care of their children at home.'
- b. **De der rådne Oxford-akademikere₁** der kun kan se **sin₁** egen
 those there rotten Oxford-academics who only can see REFL's own
 næsetip. (pers. comm., 12/3-19)
 nose-tip
 'Those rotten Oxford academics who can't see beyond the end of their nose.'
- c. Jer der har **børn₁** der har fået **sin₁** 12mdr vaccine blev
 you who have children who have got REFL's 12month vaccine became
 de syg? (facebook, 3/4-19)
 they ill?

'Those of you who have children who have received their 12 month vaccine did they fall ill?'

- d. **Nedskæringerne**₁ viser fortsat **sit**₁ grimme ansigt (facebook, 2/4-19)
cutbacks.DEF show still REFL's ugly face
'The budget cuts still show their ugly face'

2.1.2 Reflexive *sig* in Danish and other Germanic languages

As to the other half of the Germanic reflexive spectrum, Gothic had a reflexive pronoun *sik* that corresponds to *sig* in modern Danish and *sich* in modern German. The Ingvaenic² cognate of *sig* was lost at a time prior to Old English (Faltz 1977, 210). We see *sig* preserved in most of the other Germanic languages (e.g. German *sich*, Dutch *zich*, *sig/seg* in the Scandinavian languages), either with a direct path from the earlier Germanic dialects or through later borrowings from neighbouring dialects (as has been argued to be the case for Dutch, Postma 2011). Modern Frisian stands out among its neighbours in not having a *sich*. *Sich*, as opposed to *sein*, has not become limited to masculine and neuter and in all the languages that preserve or have adopted it later, *sich* may occur with third person antecedents of all genders and numbers. That is, except for earlier stages of Danish where *sig* mainly occurred with singular antecedents and the regular plural pronoun *dem* was used reflexively, just like *deres* is in modern Danish. A simplified sketch of this development is illustrated with the examples in (48). (48a) is a modern translation of (48b). (48b) is from the Danish poet Grundtvig, (48c) is from the 1550 translation of the Bible, and (48d) is from the area law Jyske Lov, the original text dated to 1241.

- (48) a. Da hvælver **rosentelte**₁ **sig** **selv**₁ ved havets bred. (2023)
then arch rose-tents REFL self by ocean.DEF's shore
- b. Da hvælve-Ø **Rosen-Telte**₁ **dem** **selv**₁ ved Havets Bred. (1850)
then arch-PL rose-tents them self by ocean.DEF's shore
- c. **Menniskane**₁ forundrede **dem**₁. (1550)
humans.DEF marvelled them
'The humans marvelled.'

²Ingvaenic or North Sea Germanic: The Anglo-Saxon tribes living around the North Sea, some of whom migrated into Britain. Old Frisian, Old Saxon and Old English all have in common that they do not have a *sig* reflexive — presumably because they all originate from a dialect area that lost *sig* prior to branching out into the three languages.

- d. Tha mughæ **skipær**₁ gøra **thæm**₁ skip sialf (1241, Jyske Lov book then must/can? shipmen make them ship self 3, ch. 5)

'Then the sailors must/can? make themselves a ship.'

Chapter 3 goes into more detail with the diachronic development of *sin* and *sig* in Danish.

2.2 A general introduction to Danish grammar

Danish is spoken as a first language (L1) by the majority of the approximately 6 million people living in Denmark and as a first or second language by speakers in particularly Greenland, the Faroe Islands, and parts of northern Germany. I provide a reasonably short overview of Danish clause structure and nominal structure in the sections below.

2.2.1 Clause structure and verbal system

Danish is an analytic language with very little inflectional richness, particularly in the verbal domain. Danish clauses have Subject – Verb – Object (SVO) order and all main clauses are verb second (V2). I illustrate various V2 configurations in (50) (note that the traces make it look like the verb only moves from I_o – this is a simplified picture as the verb originates in a VP further down in the tree).

- (49) **V2 at a glance:** A constituent – the finite verb – the rest (summary courtesy of Vikner (2021, 3))

(50) **Subject-initial V2**

- a. [_{CP} Sofie₁ købte₂ [_{IP} t₁ t₂ aldrig mælk.]
Sofie bought never milk
'Sofie never bought milk.'

Object-initial V2

- b. [_{CP} Mælk₁ købte₂ [_{IP} Sofie t₂ aldrig t₁.]
Milk bought Sofie never
'Sofie never bought milk.'

Adverbial-initial V2

- c. [_{CP} Måske købte₂ [_{IP} Sofie t₂ aldrig mælk.]
Perhaps bought Sofie never milk

'Sofie never bought milk.'

(Typically) no V2 in embedded clauses

- d. ... at Sofie aldrig købte mælk.
 that Sofie never bought milk
 '... that Sofie never bought milk.'

Danish finite verbs are inflected for past and present tense, (51a)-(51b). Past and present perfect are formed compositionally with an auxiliary verb and the relevant non-finite verb form.

- (51) a. Sofie fotografere-r. (present tense)
 Sofie photograph-PRS.
 'Sofie photographs.'
- b. Sofie fotografere-de. (past tense)
 Sofie photograph-PST
 'Sofie photographed.'
- c. Sofie har fotografere-t. (past perfect)
 Sofie has photograph-PTCP
 'Sofie has photographed.'
- d. Sofie kommer gå-ende. (present perfect)
 Sofie comes walk-PTCP
 'Sofie is walking.'

Danish verbs show no agreement with the ϕ -features (Person, Number, Gender) of the nominals in the clause. The examples in (52) show the invariant -r present tense inflection across all persons, numbers, and two gendered third person singular pronouns. The pattern would be the same with a past tense verb, i.e. invariant inflection on the verb.

- (52) a. Jeg fotograferer.
 1SG photograph
 'I photograph.'
- b. Du fotograferer.
 2SG photograph
 'You photograph.'
- c. Hun /han fotograferer.
 3F.SG /3M.SG photographs
 'She/he photographs.'

- d. Vi /I /de fotograferer.
 1PL /2PL /3PL photograph
 'We/you/they photograph.'

2.2.2 Nominal system

Grammatical gender

Danish has two grammatical genders: Common gender as on e.g. *hat*, (53a), and neuter gender as on e.g. *house*, (53b).

- (53) a. En hat
 SG.INDF.ART.C hat
 'A hat'
- b. Et hus
 SG.INDF.ART.N house
 'A house'

Gender is an inherent feature of Danish nouns and is marked overtly as agreement on the non-nominal elements in the nominal phrase. Gender is overtly present on demonstratives and articles as well as in inflection on various modifiers in the nominal phrase. -Ø means *null* inflection.

- (54) a. En rød-Ø hat
 SG.INDF.ART.C red-SG.C hat
 'A red hat'
- b. Et rød-t hus
 SG.INDF.ART.N red-SG.N house
 'A red house'

Gender marking is only present in the singular. *Katte* (En. *cats*) in (55a) is common gender plural and *egern* (En. *squirrels*) in (55b) is neuter gender plural. Both are modified with the attributive adjective *røde* which does not show the gender difference.

- (55) a. Røde katte spiser fisk. *katte, plural*
 red cats eat fish
 'Red cats eat fish.'
- b. Røde egern spiser nødder. *egern, plural*
 red squirrels eat nuts
 'Red squirrels eat nuts.'

Definiteness

Danish has postnominal definiteness marking when the nominal is unmodified, (56a). If the nominal is modified by e.g. an adjective, a prenominal article is used instead, (56b). The article corresponds to the gender of the head noun in the singular: Common gender postnominal *-en* and prenominal *den*, neuter gender postnominal *-et* and prenominal *det*. The plural article is postnominal *-ene* and prenominal *de*, irrespective of gender, (56c).

- (56) a. Katt-en spiser græss-et.
 cat-DEF.ART.COMMON eats grass-DEF.ART.NEUTER
 'The cat eats the grass.'
- b. Den sorte kat spiser det grønne græs.
 DEF.ART.COMMON black cat eats DEF.ART.NEUTER green grass
 'The black cat eats the green grass.'
- c. Katt-ene spiser de store mus.
 cat-DEF.ART.PL eat DEF.ART.PL large mice
 'The cats eat the large mice.'

Danish stands out among the Mainland Scandinavian languages in not having double definiteness marking. Swedish and Norwegian (and Faroese, but not Icelandic, to also mention Insular Scandinavian) marks definiteness twice in most contexts: With a prenominal definiteness marker and a postnominal suffix (see e.g. Delsing (1993, 116) where the Swedish examples, including the interlinear gloss, in (57) are from).

- (57) a. hus-et (Swedish)
 house-the
 'the house'
- b. det stora hus-et (Swedish)
 the large house-the
 'the large house'
- c. hus-et (Danish)
 house-the
 'the house'
- d. *det store hus-et (Danish)
 the large house-the
 'the large house'
- e. det store hus (Danish)
 the large house

'the large house'

Inflection in the nominal phrase

The Danish nominal system has quite a bit more inflectional richness than the verbal system. Both attributive and (most) predicative adjectives show some agreement in person, number, and gender (ϕ -agreement, in short), as do some of the possessive pronouns.

Danish has overt agreement in number and gender within the nominal phrase between the head noun and most other elements inside the phrase (adjectives, articles, demonstratives, numerals that can inflect). The shape of the agreement marker depends on whether the nominal phrase is definite or indefinite. When the nominal is definite the agreement is invariant -e, so-called *weak* agreement, irrespective of the ϕ -features of the head noun. The examples in (58) show definite noun phrases with both genders and the singular and in the plural. All modifying adjectives have the weak -e inflection.

- (58) a. Den gul-e bjørn elsker honning. (bjørn, common
DEF.ART.SG.C yellow-AGR bear loves honey
gender, singular)
 'The yellow bear loves honey.'
- b. Det grønne hus er mit. (hus, neuter gender, singular)
DEF.ART.SG.N green-AGR house is mine
 'The green house is mine.'
- c. De røde bukser er flotte. (bukser, plural)
DEF.ART.PL red-AGR pants are nice
 'The red pants are nice.'

When the nominal is indefinite the agreement suffix varies according to the number and gender of the head noun, so-called *strong* agreement. The agreement suffix is -e in plural nominals irrespective of whether the nominal is definite or indefinite. It is very possible that an argument could be made where the Danish plural agreement marker and weak agreement marker are in fact a kind of default agreement marker. This is outside the scope of this chapter, however, even if it is an interesting thought. If followed and shown to be reasonably correct, it could lead to the conclusion that Danish plurals in fact never agree except for default agreement.

- (59) a. En gul-Ø bi laver honning. (bi, common
INDEF.ART.SG.C yellow-AGR.SG.C bee makes honey
gender, singular, -Ø (null) ending)

'A yellow bee makes honey.'
- b. Et grøn-t træ er smukt. (træ, neuter gender,
INDEF.ART.SG.N green-AGR.SG.N tree is beautiful
singular, -t ending)

'A green tree is beautiful.'
- c. Grønn-e træer er sjove. (træer, plural, -e ending)
Green-AGR.PL trees are funny.
'Green trees are funny.'

Predicative adjectives are marked for agreement according to the strong agreement paradigm, i.e. an -Ø suffix on the adjective for singular common gender nouns, a -t suffix on the adjective for singular neuter gender nouns, and an -e suffix on the adjective for plural nouns.

- (60) a. En gul-Ø bi er normal-Ø. (bi, common
INDEF.ART.SG.C yellow-AGR.SG.C bee is normal-AGR.SG.C
gender, singular, -Ø ending)

'A yellow bee is normal.'
- b. Et grøn-t træ er smuk-t. (træ, neuter
INDEF.ART.SG.N green-AGR.SG.N tree is beautiful-AGR.SG.N
gender, singular, -t ending)

'A green tree is beautiful.'
- c. Grønn-e træer er sjov-e. (træer, plural, -e ending)
Green-AGR.PL trees are funny-AGR.PL
'Green trees are funny.'

Adjectives that are created from participles show less agreement in predicative use in modern Danish than regular adjectives (Retskrivningsordbogen §33-§35). The adjective *overrasket* in (61a) is perfectly acceptable with the singular -t inflection even though the subject is plural *de*. The same sentence with a plural-inflected adjective, (61b), is markedly less acceptable. This is especially true when the participial adjective takes a prepositional complement.

- (61) a. De blev overraske-t af uvejret.
they became surprise-SG.N by storm-DEF
'They were surprised by the storm.'
- b. ?De blev overraske-de af uvejret.
they became surprise-PL by storm-DEF
'They were surprised by the storm.'

Case marking

Case is overtly present only on pronouns, (62), where Danish distinguishes between NOMINATIVE and ACCUSATIVE (sometimes called OBLIQUE to indicate that the Case could technically be seen as an amalgam of accusative and dative).

- (62) a. Jeg ser hende.
1NOM.SG see 3ACC.SG
'I see her.'
- b. Hun ser mig.
3NOM.SG see 1ACC.SG
'She sees me.'
- c. Sofie-Ø ser Lars-Ø.
Sofie-NOM sees Lars-ACC
'Sofie sees Lars.'
- d. Lars-Ø sees Sofie-Ø.
Lars-NOM sees Sofie-ACC
'Lars sees Sofie.'

Possessive clitic -s

Danish marks possession with a clitic -s that cliticizes to the end of the possessor constituent. The -s is not a genitive marker as it may attach to any word class as long as the word is the last element in the possessor phrase.

- (63) a. Sofie-s hund løber stærkt.
Sofie-POSS runs quickly
'Sofie's dog runs quickly.'
- b. [Manden der kravlede højt op]-s hund løber stærkt.
Man.DEF who climbed high up-POSS dog runs quickly
'The man who climbed up high's dog runs quickly.'

Inflectional forms of the Danish possessive pronouns

Whenever I write *sin* in this thesis, I use *sin* as shorthand for all three inflectional forms: *sin*, *sit* and *sine*. Exceptions are *sin* used in examples or where I otherwise note that I only refer to the single form *sin*. In general, *sin* inflects (more or less) like the Danish adjectives in that it corresponds in gender and number to its complement, as shown in (64), just like adjectives correspond in gender and number to the nominal they modify, or definite articles do to their nominal complements.

- (64) a. **Morten**₁ krammer **sit**₁ gaml-e æbletræ. *sit*, neuter, singular
 Morten hugs REFL's old-WEAK apple-tree
 'Morten hugs his old apple tree.'
- b. **Marie**₁ bygger **sin**₁ rød-e husbåd færdig. *sin*, common, sg.
 Marie builds REFL's red-WEAK house-boat finished
 'Marie finishes building her funny house boat.'
- c. **Johan**₁ finder **sine**₁ sød-e venner udenfor. *sine*, no gender, pl.
 Johan finds REFL's sweet-WEAK friends outside
 'Johan finds his sweet friends outside.'

There are no *strong* and *weak* paradigms of inflectional forms for *sin* and *sin* only ever occurs in definite DPs. It is possible that *sin* itself has a definite feature (or that it inherits it from somewhere else) and that it consequently makes the nominal definite. This is the analysis put forward in Julien (2005). The examples in (64) all have a DP with a *sin* and an adjective, and the reflexive is overtly inflected to correspond to the features of the head noun and the adjectives all have the weak -e inflectional form. This very clearly indicates that there must be an additional mechanism that plays into the inflection on the reflexives, compared to that of the adjectives.

As can be seen in (64c), it is completely grammatical and unproblematic for *sin* (*sine*, to be precise) to occur as part of a plural nominal, *sine venner*. I describe *sin* as number-restricted in Danish but the number restriction is only with reference to the antecedent of *sin*, not the nominal that *sin* is a part of. Whenever *sin* has a singular antecedent, it makes no difference to the acceptability whether *sin/sit/sine* is part of a singular or a plural nominal, as e.g. shown by the parallel examples in (65). *Sin* inflects as *sin*, *sit*, or *sine* according to the number and gender features of its nominal complement.

- (65) a. * **Hun**₁ elsker **hendes**₁ barn.
 she loves her child

'She loves her child.'

- b. **Hun**₁ elsker **sit**₁ barn.
she loves REFL's child
'She loves her child.'
- c. ***Hun**₁ elsker **hendes**₁ børn.
she loves her children
'She loves her children.'
- d. **Hun**₁ elsker **sine**₁ børn.
she loves REFL's children
'She loves her children.'

In the other direction, whenever another third person possessive is used (e.g. plural possessive *deres* as a bound form in (66), or singular feminine possessive *hendes* as an unbound form in (67)), bound or unbound, the possessive form is invariant with respect to the gender and number of the nominal that it is a part of.

- (66) a. **De**₁ fotograferer **deres**₁ barn. (*barn*, neuter, singular)
they photograph their child
'They photograph their child.'
- b. **De**₁ fotograferer **deres**₁ kat. (*kat*, common, singular)
they photograph their cat
'They photograph their cat.'
- c. **De**₁ fotograferer **deres**₁ børn. (*børn*, plural)
they photograph their children
'They photograph their children.'
- (67) a. Han₁ fotograferer hendes₂ barn. (*barn*, neuter, singular)
he photographs her child
'He photographs her child.'
- b. Han₁ fotograferer hendes₂ kat. (*kat*, common, singular)
he photographs her cat
'He photographs her cat.'
- c. Han₁ fotograferer hendes₂ børn. (*børn*, plural)
he photographs her children
'He photographs her children.'

This is the case for most of the Danish possessives, discounting *min* (En. *my/mine*), *din* (En. singular *your/yours*) and *sin*. That is, the majority of the Danish possessive pronouns only agree overtly with their antecedent, not their complement.

2.3 The Danish reflexive system

The following section outlines the analysis of the Danish reflexives that was originally published as Vikner (1985), and which I essentially subscribe to. The analysis is situated within the Government and Binding (GB) framework of Chomsky (1981). The Danish reflexive system, in several respects, is not directly compatible with Chomsky's Principle A and B. Specific issues are e.g. the Danish reflexive *sig*, which cannot be locally bound, and the reflexive possessive *sin*, which may be bound either locally or non-locally.

2.3.1 +/- *selv* (1st and 2nd person examples)

The following sections, which lay out the Danish reflexive system, draw heavily on my Master's thesis, Ehlers (2017). The specific setup of the examples is from Vikner (1985) (and again in e.g. Ehlers and Vikner 2016).

Danish has a distinction between reflexive and non-reflexive pronouns just like English does. I follow Vikner (1985) (and Ehlers and Vikner (2016), Vikner and Ehlers (2017), and Ehlers (2017)) and call this the $\pm selv$ distinction. In English, $\pm selv$ is the primarily relevant constraint in all three persons (1st, 2nd, 3rd person as covered by Principle A and Principle B which are summarized on page 42). The Danish $\pm selv$ constraint derives the difference between the first and second person object pronouns, *mig* and *dig* (En. *me* and singular *you*), and the first and second person reflexive pronouns, *mig selv* and *dig selv* (En. *myself* and *yourself*). I use first person for illustration in (69) and (70) but second person can be substituted in without any other changes. I use indices and boldface to indicate binding relations. The *selv* in the examples is part of the complex pronoun *mig selv* and is not a separate intensifier modifying *mig* (see e.g. Jensen 2010 for a discussion of the various uses of *selv*).

The organisation of the examples is as outlined in (68).

- (68)
- a. antecedent is subject in the minimal IP
 - b. antecedent is non-subject in the minimal IP
 - c. antecedent is subject outside minimal infinitival IP
 - d. antecedent is non-subject outside the minimal infinitival IP
 - e. antecedent is subject outside the finite IP
 - f. antecedent is non-subject outside the finite IP

- g. no antecedent
- h. antecedent does not c-command the intended coreferential pronoun
(The examples and organisation are adapted from Vikner (1985) and Ehlers and Vikner (2016).)

The examples in (69) show the binding properties of reflexive *mig selv*. The examples in (70) show the directly complementary binding properties of non-reflexive *mig*.

- (69)
- a. **Jeg**₁ fotograferer ofte **mig selv**₁.
I photograph often myself
 - b. Harry og Ron viste **mig**₁ et billede af **mig selv**₁.
Harry and Ron showed me a picture of myself
 - c. ***Jeg**₁ overtalte Harry og Ron til [PRO at fotografere **mig selv**₁].
I convinced Harry and Ron to to photograph myself
 - d. *Harry og Ron lovede **mig**₁ [PRO at fotografere **mig selv**₁].
Harry and Ron promised me to photograph myself
 - e. ***Jeg**₁ tror at [Harry og Ron vil fotografere **mig selv**₁].
I think that Harry and Ron will photograph myself
 - f. *Harry og Ron lovede **mig**₁ [at de ville fotografere **mig selv**₁].
Harry and Ron promised me that they would photograph myself
 - g. *Harry og Ron fotograferede **mig selv**₁.
Harry and Ron photographed myself
 - h. *Fotografiet af **mig**₁ overraskede **mig selv**₁.
Photograph.DEF of me surprised myself
- (70)
- a. ***Jeg**₁ fotograferer ofte **mig**₁.
I photograph often me
 - b. *Harry og Ron viste **mig**₁ et billede af **mig**₁.
Harry and Ron showed me a picture of me
 - c. **Jeg**₁ overtalte Harry og Ron til [PRO at fotografere **mig**₁].
I convinced Harry and Ron to to photograph me
 - d. Harry og Ron lovede **mig**₁ [PRO at fotografere **mig**₁].
Harry and Ron promised me to photograph me
 - e. **Jeg**₁ tror at [Harry og Ron vil fotografere **mig**₁].
I think that Harry and Ron will photograph me

- f. Harry og Ron lovede **mig**₁ [at de ville fotografere **mig**₁].
Harry and Ron promised me that they would photograph me
- g. Harry og Ron fotograferede **mig**₁.
Harry and Ron photographed me
- h. Fotografiet af **mig**₁ overraskede **mig**₁.
Photograph.DEF of me surprised me

The important thing to notice in (69) and (70) is the lack of overlap in grammaticality in the examples. (69a) and (69b) are the only two grammatical constructions in (69). (70) is in direct complementary distribution with (69), meaning that (70a) and (70b) are the only two examples that are not grammatical. Vikner (1985) (and Ehlers and Vikner 2016) suggests that the two generalizations in (71) can account for the Danish data as illustrated above. These examples and the generalizations correspond to Chomsky's Principles A and B. Note, however, that this fit is only true in Danish first and second person. The third person pronouns in Danish are limited by other constraints, as will be discussed in the next two sections.

- (71) a. *mig selv* must be bound in the minimal IP containing its case assigner (Principle A).
- b. *mig* must be free in the minimal IP containing its case assigner (Principle B).
(Ehlers and Vikner 2016)

2.3.2 ± sig (*sin* examples)

The second dimension that is necessary for describing the Danish reflexive system is that which I follow Ehlers and Vikner (2016) in calling the ± *sig* requirements, and which, notably, do not apply to English because the English cognate of *sig* was lost prior to Old English. The generalizations that correspond to the generalizations given for ±*selv* in (71) are given in (72).

- (72) a. *sin* must be bound by a subject in the minimal finite IP.
- b. *hendes* must not be bound by a subject in the minimal IP.
(Vikner 1985)

These generalizations do not correspond to any of the binding principles A to C. They are not in complete complementary distribution given that (72a) requires that the antecedent is a subject in the minimal finite clause, whereas (72b) only demands that

the pronoun not be bound in the minimal clause. This system predicts that there will be cases where a possessive pronoun in an embedded non-finite clause may be bound by a subject in the finite matrix clause, which then leads to a situation of structurally conditioned optionality between the reflexive and the non-reflexive possessive. That this is the case is illustrated in (73) and (74) where both (73c) and (74c) are grammatical in Danish.

- (73) a. **Hun**₁ fotograferer ofte **sin**₁ ugle.
She photographs often REFL's owl
- b. * Harry og Ron viste **hende**₁ et billede af **sin**₁ ugle.
Harry and Ron showed her a picture of REFL's owl
- c. **Hun**₁ overtalte Harry og Ron til [PRO at fotografere **sin**₁ ugle].
She convinced Harry and Ron to to photograph REFL's owl
- d. * Harry og Ron lovede **hende**₁ [PRO at fotografere **sin**₁ ugle].
Harry and Ron promised her to photograph REFL's owl
- e. * **Hun**₁ tror at [Harry og Ron vil fotografere **sin**₁ ugle].
She thinks that Harry and Ron will photograph REFL's owl
- f. * Harry og Ron lovede **hende**₁ [at de ville fotografere **sin**₁ ugle].
Harry and Ron promised her that they would photograph REFL's owl
- g. * Harry og Ron fotograferede **sin** ugle.
Harry and Ron photographed REFL's owl
- h. * Fotografiet af **hende**₁ overraskede **sin**₁ ugle.
Photograph.DEF of her surprised REFL's owl
- (74) a. * **Hun**₁ fotograferer ofte **hendes**₁ ugle.
She photographs often her owl
- b. Harry og Ron viste **hende**₁ et billede af **hendes**₁ ugle.
Harry and Ron showed her a picture of her owl
- c. **Hun**₁ overtalte Harry og Ron til [PRO at fotografere **hendes**₁ ugle].
She convinced Harry and Ron to to photograph her owl
- d. Harry og Ron lovede **hende**₁ [PRO at fotografere **hendes**₁ ugle].
Harry and Ron promised her to photograph her owl
- e. **Hun**₁ tror at [Harry og Ron vil fotografere **hendes**₁ ugle].
She thinks that Harry and Ron will photograph her owl

- f. Harry og Ron lovede **hende**₁ [at de ville fotografere **hendes**₁ ugle].
Harry and Ron promised her that they would photograph her owl
- g. Harry og Ron fotograferede **hendes** ugle.
Harry and Ron photographed her owl
- h. Fotografiet af **hende**₁ overraskede **hendes**₁ ugle.
Photograph.DEF of her surprised her owl

2.3.3 Full system (*sig/sig selv* examples)

The four generalizations in (75) combine the requirements for *selv* and for *sig*, explored independently in the previous sections. The first dimension, $\pm selv$, describes whether the pronoun must or can have an antecedent in the minimal IP. This is analogous to Chomsky's Principle A in the original Binding Theory. $\pm selv$ captures the difference in distribution between *mig* (En. *me*) and *mig selv* (En. *myself*) (and the corresponding plural first person pronouns as well as the second person pronouns in both plural and singular). The second dimension, $\pm sig$, does not have an English counterpart. $\pm sig$ describes whether the pronoun needs an antecedent that is a subject in the minimal finite IP. This captures the distributional difference between *sin* (REFL's) and *hendes* (*her*). The combination of these two dimensions describes the distribution of the third person non-possessive pronouns and reflexives.

- (75) a. ***sig selv*** (+*sig*, +*selv*) must be bound by a subject in the minimal IP containing its case assigner.
- b. ***hende selv*** (–*sig*, +*selv*) must be bound in the minimal IP containing its case assigner and the antecedent must not be a subject.
- c. ***sig*** (+*sig*, –*selv*) must be bound by a subject in the minimal **finite** IP but must be free in the minimal IP containing its case assigner.
- d. ***hende*** (–*sig*, –*selv*) must be free in the minimal IP containing its case assigner. (Ehlers and Vikner 2016)

The four sets of examples below show the various possibilities and impossibilities in the third person non-possessive pronouns and reflexives. The first set shows that *sig selv* must be bound by the local subject.

- (76) a. **Hun**₁ fotograferer ofte **sig selv**₁.
She photographs often REFL self

- b. * Harry og Ron viste **hende**₁ et billede af **sig selv**₁
 Harry and Ron showed her a picture of REFL self
- c. * **Hun**₁ overtalte Harry og Ron til [PRO at fotografere **sig selv**₁].
 She convinced Harry and Ron to to photograph REFL self
- d. * Harry og Ron lovede **hende**₁ [PRO at fotografere **sig selv**₁].
 Harry and Ron promised her to photograph REFL self
- e. * **Hun**₁ tror at [Harry og Ron vil fotografere **sig selv**₁].
 She thinks that Harry and Ron will photograph REFL self
- f. * Harry og Ron lovede **hende**₁ [at de ville fotografere **sig selv**₁].
 Harry and Ron promised her that they would photograph REFL self
- g. * Harry og Ron fotograferede **sig selv**₁.
 Harry and Ron photographed REFL self
- h. * Fotografiet af **hende**₁ overraskede **sig selv**₁.
 Photograph.DEF of her surprised REFL self

The second set of examples shows the binding requirements of the Danish *PRON* + *selv* complex non-reflexive pronouns with *hende selv* as the example. They show that *hende selv* must be bound by a non-subject that is sufficiently local.

- (77) a. * **Hun**₁ fotograferer ofte **hende selv**₁.
 She photographs often her self
- b. Harry og Ron viste **hende**₁ et billede af **hende selv**₁
 Harry and Ron showed her a picture of her self
- c. * **Hun**₁ overtalte Harry og Ron til [PRO at fotografere **hende selv**₁].
 She convinced Harry and Ron to to photograph her self
- d. * Harry og Ron lovede **hende**₁ [PRO at fotografere **hende selv**₁].
 Harry and Ron promised her to photograph her self
- e. * **Hun**₁ tror at [Harry og Ron vil fotografere **hende selv**₁].
 She thinks that Harry and Ron will photograph her self
- f. * Harry og Ron lovede **hende**₁ [at de ville fotografere **hende selv**₁].
 Harry and Ron promised her that they would photograph her self
- g. * Harry og Ron fotograferede **hende selv**₁.
 Harry and Ron photographed her self

- h. * Fotografiet af **hende**₁ overraskede **hende selv**₁.
 Photograph.DEF of her surprised her self

The third set of examples shows the binding requirements of *sig*, namely that it must be bound by a subject that is sufficiently non-local.

- (78) a. * **Hun**₁ fotograferer ofte **sig**₁.
 She photographs often REFL
- b. * Harry og Ron viste **hende**₁ et billede af **sig**₁
 Harry and Ron showed her a picture of REFL
- c. **Hun**₁ overtalte Harry og Ron til [PRO at fotografere **sig**₁].
 She convinced Harry and Ron to to photograph REFL
- d. * Harry og Ron lovede **hende**₁ [PRO at fotografere **sig**₁].
 Harry and Ron promised her to photograph REFL
- e. * **Hun**₁ tror at [Harry og Ron vil fotografere **sig**₁].
 She thinks that Harry and Ron will photograph REFL
- f. * Harry og Ron lovede **hende**₁ [at de ville fotografere **sig**₁].
 Harry and Ron promised her that they would photograph REFL
- g. * Harry og Ron fotograferede **sig**₁.
 Harry and Ron photographed REFL
- h. * Fotografiet af **hende**₁ overraskede **sig**₁.
 Photograph.DEF of her surprised REFL

The fourth set of examples shows the binding requirements of the non-reflexive simple pronouns with *hende* as the example. They show that *hende* must be locally free but that it allows non-local binding. The same is true for the other personal pronouns, e.g. *ham* and *dem* (En. *him* and *dem*).

- (79) a. * **Hun**₁ fotograferer ofte **hende**₁.
 She photographs often her
- b. * Harry og Ron viste **hende**₁ et billede af **hende**₁
 Harry and Ron showed her a picture of her
- c. **Hun**₁ overtalte Harry og Ron til [PRO at fotografere **hende**₁].
 She convinced Harry and Ron to to photograph her
- d. Harry og Ron lovede **hende**₁ [PRO at fotografere **hende**₁].
 Harry and Ron promised her to photograph hende
- e. **Hun**₁ tror at [Harry og Ron vil fotografere **hende**₁].
 She thinks that Harry and Ron will photograph her

	Domain reflexives bound in the minimal IP containing its case assigner (= Principle A)	Domain non-reflexives not bound in the minimal IP containing its case assigner (= Principle B)	Neutralised (possessives)
Binder reflexives bound by a subject in the minimal finite IP	<i>sig selv</i>	<i>sig</i>	<i>sin</i>
Binder non-reflexives not bound by a subject in the minimal IP	<i>ham selv, hende selv, den selv, det selv, dem selv</i>	<i>ham, hende, den, det, dem</i>	<i>hans, hendes, dens, dets</i>
Neutralised (1st & 2nd person)	<i>mig selv, dig selv, os selv, jer selv</i>	<i>mig, dig, os, jer</i>	<i>min, din, vores, jeres, deres</i>

Figure 2.2: Vikner’s summary of the Danish reflexive system. First introduced in Vikner (1985) and further discussed and empirically supported in Ehlers and Vikner (2016), Vikner and Ehlers (2017), and Ehlers (2017).

- f. Harry og Ron lovede **hende**₁ [at de ville fotografere **hende**₁].
Harry and Ron promised her that they would photograph her
- g. Harry og Ron fotograferede **hende**₁.
Harry and Ron photographed her
- h. Fotografiet af **hende**₁ overraskede **hende**₁.
Photograph.DEF of her surprised her

Vikner (1985) argues that the Danish reflexive system that is shown in the three preceding sections can be summarized as in figure 2.2 and I essentially agree with this analysis. I argue in Ehlers (2017) that the binding domain should be revised to the definition in (80) for slightly better coverage but the essential distributional differences are there.

- (80) The **binding domain** for X is the minimal IP/DP containing:
- a. X and
 - i. a finite verb or
 - ii. a subject that c-commands X

2.3.4 Non-local binding by inanimate antecedents

The examples in the previous section all have animate antecedents of the reflexive and this may well hide a substantive point. Reuland and Zubkov 2022 argue that an inanimate antecedent cannot bind a reflexive across a possessor (their example (14), reproduced here as (81) in Russian. The index type and reflexive shorthand (REFL/REFL’s)

are changed slightly to fit the glossing conventions of this thesis. They also argue that an inanimate antecedent cannot bind a reflexive non-locally into an embedded infinitival clause (their example (24), reproduced here as (82)). This is contrasted with animate antecedents in the same constructions where binding is apparently perfectly fine, as shown in the last example in both example sets.

(81) **Russian: Non-local binding by inanimates not possible (infinitival clause)**

- a. Èta kniga₁ isportila ego₂ otnošenie k sebe_{*1/2/*3} /nej_{1/3}.
 this book.NOM damaged.FSG his attitude.ACC to REFL.DAT /it.DAT
 'This book damaged his attitude towards it/himself.'
- b. Èta kniga₁ isportila ego₂ otnošenie k svoemu_{*1} /ee₁
 this book.NOM damaged.FSG his attitude.ACC to REFL's /its
 avtoru.
 author.DAT
 'This book damaged his attitude towards its author.'
- c. Vanja₁ isportil ee₂ otnošenie k sebe_{1/2/*3} /[svoej_{1/2/*3}
 Vanya.NOM damaged.M.SG her attitude.ACC to REFL.DAT /REFL's
 sestre].
 sister.DAT
 'Vanya₁ damaged her₂ attitude towards him/herself/[his₁/her₂ sister].'

(82) **Russian: Non-local binding by inanimates not possible (across possessor)**

- a. Èta kniga₁ zastavila Anju₂ PRO₂ pereproverit' sebja_{*1/2/*3}
 this book.NOM forced.FSG Anya.ACC recheck.INF REFL.ACC
 /ee_{1/3}.
 /it.ACC
 'This study made Anya recheck it/herself.'
- b. Èta kniga₁ zastavila Anju₂ PRO₂ pereproverit' svoi_{*1/2/*3}
 this book.NOM forced.FSG Anya.ACC recheck.INF REFL's
 /ee_{1/3} vyvody.
 /its conclusions.ACC
 'This study made Anya recheck her/its conclusions.'
- c. Vanja₁ zastavil Anju₂ PRO₂ pereproverit' sebja_{1/2/*3}
 Vanya.NOM forced.M.SG Anya.ACC recheck.INF REFL.ACC
 /svoi_{1/2/*3} /ee_{1/3} vyvody.
 /REFL's /its conclusions.ACC
 'Vanya₁ made Anya₂ recheck him₁/herself₂/[his₁/her₂ conclusions].'

At a first glance, I find the same effects in Danish. (83) shows an inanimate antecedent that (fails to) bind a reflexive non-locally across an embedded animate sub-

ject into an infinitive clause. (84) shows binding across an animate possessor into a nominal. The (sometimes) ambiguity of the Danish sentences is also clear in the English translation in the glosses where *sin* is translated as its/her or his/her, depending on the gender of the possible antecedents.

(83) **Animacy effects in non-local binding: Infinitive**

- a. **Bogen**₁ fik Lise₂ til at genoverveje **sin**_{*1/2} konklusion.
book.DEF made Lise to to reconsider REFL's conclusion
'The book made Lise reconsider its/her conclusion.'
- b. **Lars**₁ fik Lise₂ til at genoverveje **sin**_{1/2} konklusion.
Lars made Lise to to reconsider REFL's conclusion
'Lars made Lise reconsider his/her conclusion.'
- c. **Bogen**₁ fik Lise₂ til at genoverveje **dens**_{1/*2} konklusion.
book.DEF made Lise to to reconsider REFL's conclusion
'The book made Lise reconsider its conclusion.'
- d. **Lars**₁ fik Lise₂ til at genoverveje **hans**_{1/*2} konklusion.
Lars made Lise to to reconsider REFL's conclusion
'Lars made Lise reconsider his conclusion.'

The possessive reflexive *sin* in (83b) is ambiguous as to its antecedent: Either Lars makes Lise reconsider his conclusion (non-local binding) or Lars makes Lise reconsider her own conclusion (local binding). Both readings are possible, just as predicted by Vikner's reflexive system. This is not the case when the matrix subject is inanimate, as in (83a). Here the sentence can only read as though the book makes Lise reconsider her own conclusion (local binding), and the non-local reading where the book makes Lise reconsider the book's conclusion is not available. It is possible to coerce this non-locally bound reading but only in a situation where the book is somehow alive and capable of active convincing, i.e. a situation where the book is in fact animate. (83c) shows that the reading where the (inanimate) book makes Lise reconsider the book's conclusion is actually possible, but only with a pronominal, not a reflexive. This option is also possible with an animate matrix antecedent, (83d).

(84) **Animacy effects in non-local binding: Across possessor**

- a. **Bogen**₁ ændrede Lises₂ holdning til **sin**_{*1/2} konklusion.
book.DEF changed Lise's opinion towards REFL's conclusion
'The book changed Lise's opinion of its/her conclusion.'

- b. **Lars**₁ ændrede Lises₂ holdning til **sin**_{1/2} konklusion.
Lars changed Lise's opinion towards REFL's conclusion
'Lars changed Lise's opinion of his/her conclusion.'
- c. **Bogen**₁ ændrede Lises₂ holdning til **dens**_{1/*2} konklusion.
book.DEF changed Lise's opinion towards REFL's conclusion
'The book changed Lise's opinion of its conclusion.'
- d. **Lars**₁ ændrede Lises₂ holdning til **hans**_{1/*2} konklusion.
Lars changed Lise's opinion towards REFL's conclusion
'The book changed Lise's opinion of his conclusion.'

The same effect of inanimate versus animate matrix subject seems to be true for binding into a nominal with an animate possessor, as indicated by the examples in (84). (84b) is ambiguous between the reading where Lise changes her opinion of Lars' conclusion (non-local binding) and the reading where Lise changes her opinion of her own conclusion (local binding). This apparent animacy effect is not predicted by Vikner's reflexive system which does not take animacy into account. To complicate matters further, however, I do find some examples in the Danish KorpusDK corpus of non-local binding of reflexive *sig* by an inanimate antecedent. I reproduce a few of these in (85). Similar examples are possible with *sin* instead of *sig*, as is evident from the constructed examples in (86). I have not found any examples with *sin* with a non-local inanimate antecedent in KorpusDK but this may not be too surprising since examples of non-local Danish reflexives are scant in the first place, and non-locally bound inanimate antecedents even more so. The examples in (85) are perfectly fine in Danish which could indicate that the apparent impossibility of inanimate non-local antecedents that shows up in (84) and (84) is not actually a full ban on non-local binding of inanimates, contra to what Reuland and Zubkov 2022 find for Russian. The examples all have the reflexive within a prepositional phrase (PP), and that may be a relevant factor. Most of the examples of non-locally bound *sig* found in Ehlers and Vikner (2016) are also contained within PPs, which is perhaps a somehow more permissive binding domain.

(85) **Animacy effects in non-local binding: Danish counter-examples from KorpusDK**

- a. **De fleste af planeterne**₁ har en eller flere måner kredsende
the most of planets.DEF have one or more moons
omkring **sig**₁.
orbiting around REFL

'Most of the planets have one or more moons orbiting around them.'

- b. ... mens **vognens jernhjul**₁ gungrer af sted med små
 while cart.DEF's iron-wheels thunder a long with small
 gnister flygende efter **sig**₁.
 sparks flying after REFL
 '... while the cart's iron wheels thunder along with small sparks flying af-
 ter them.'
- c. Hvis en ting er god, behøver **den**₁ jo ikke så mange ord til at
 if a thing is good needs it yes not so many words to to
 reklamere for **sig**₁.
 advertise for REFL
 'If a thing is good, it does not need as many words to advertise for it, after
 all.'

(86) **Animacy effects in non-local binding: Constructed Danish counter-examples with *sin***

- a. **Planet**₁ har flere måner kredsende omkring **sin**₁ overflade.
 planets.DEF has several moons orbiting around REFL's
 surface
 'The planet several moons orbiting around its surface.'
- b. Hvis en ting er god, behøver **den**₁ jo ikke så mange ord til at
 if a thing is good needs it yes not so many words to to
 reklamere for **sine**₁ kvaliteter.
 advertise for REFL's qualities
 'If a thing is good, it does not need as many words to advertise for its
 qualities, after all.'

2.3.5 Locally bound *sig*: *skamme, vaske, brænde*

The large majority of instances of *sig* found in the large Danish written corpus KorpusDK are examples of locally bound *sig*. This is perhaps surprising given the claim of the system in figure 2.2 that *sig cannot* be bound locally. In order to account for this fact, we need to distinguish between argument *sig* (which is what is captured by the requirements on *sig* in figure 2.2) and non-argument *sig*. Argument *sig* is assigned a theta-role and can typically be exchanged with another nominal. Non-argument *sig*

usually cannot be exchanged with another nominal without changing the meaning of the clause. Non-argument *sig* occurs with (at least) three types of verb classes which Ehlers and Vikner (2016) call *skamme*, *vaske*, and *brænde*. I give two representative examples from each verb class in (87).

- (87) a. Peter skammer sig.
Peter shames REFL
'Peter is ashamed.'
- b. *Peter skammer sig selv.
Peter shames REFL self
'Peter shames himself.'
- (88) a. Peter vasker sig.
Peter washes REFL
'Peter washes (himself).'
- b. Peter vasker sig selv.
Peter washes REFL self
'Peter washes himself.'
- (89) a. Peter brænder sig.
Peter burns REFL
'Peter burns himself (by accident, with little damage).'
- b. Peter brænder sig selv.
Peter burns REFL self
'Peter burns himself (on purpose, perhaps fatally).'

Ehlers (2017) found that the *brænde* verbs make up the majority of (a specific subset of) locally bound *sig* in KorpusDK (56.5 %), that the *vaske* verbs are reasonably infrequent (9.0 %), and that the *skamme* verbs are somewhat less frequent than the *brænde* verbs (34.5 %).

The *skamme* verbs are represented by *skamme* in (87a)-(87b). These are traditionally called the *inherently reflexive verbs*. They only occur with *sig* in Danish and are ungrammatical with e.g. *sig selv* or other nominals. The English equivalents usually occur with no verbal complements at all (e.g. *be ashamed*). The *vaske* verbs are represented by *vaske* in (88a)-(88b). These verbs may occur with either *sig* or *sig selv* without a change in meaning. The English equivalents are essentially the same: They may occur without a verbal complement (*Peter washes*) or with a reflexive verbal complement (*Peter washes himself*) without change in meaning. The class contains grooming

verbs (*wash, shave* and other verbs that are typically self-directed) but also a verb such as *forsvare* (En. *defend*) which is less obviously self-directed. The third class of verbs is represented by *brænde* in (89a)-(89b). These verbs may occur both with *sig, sig selv*, and other nominals but there is a clear difference in meaning between the two options. *Peter brænder sig* is non-volitional, an accidental burning, and *Peter brænder sig selv* is clearly volitional, e.g. an act of self-harm. One way of analyzing this difference is to say that the distribution of θ -roles is different in (89a) and (89b). Vikner (1985) argues that the sentences with locally bound *sig* have one less θ -role than the same sentence with locally bound *sig selv*. In (89b), Peter is the AGENT and *sig selv* is the THEME. The sentence reads as though his burning of himself is agentive and on purpose. In (89a), conversely, Peter is the THEME and *sig* has no θ -role and this configuration gives the sentence the non-agentive, non-volitional reading.

2.4 Local binding of singular possessive non-reflexives - Han tog hans hat

The final topic of this chapter is the occurrence of locally bound non-reflexive pronouns in specific parts of Jutland, but also the rest of Denmark. This issue with Danish reflexives is surprisingly salient in the public discourse about grammar. It is exemplified by the clause *Han tog hans hat og gik hans vej* (Eng. *He took his hat and walked away*). I would expect that a lot, if not all, of non-linguistically inclined adult Danish speakers would have a ready opinion when presented with that specific sentence (a sentiment that is mirrored by e.g. Juel Jensen in Juel Jensen (2009a)). It is even used as the topic starter in a great deal of popular writing about Danish grammar³.

The sentence and its grammatically correct twin are reproduced in (90). The issue is that standard Danish, as opposed to English where *He took his hat* without further context is unproblematically ambiguous, has *sin* as the canonical locally bound form.

³The first four results from a Google search on the topic are given below and similar results continue for several pages. All hits are either blog posts or newspaper articles that explain the difference between *hans* and *sin* in humorous ways:

Han tog hans hat og gik hans vej, Politiken, 7/11 2007 (<https://sproget.dk/raad-og-regler/artikler-mv/sprogligt-politikens-sprogklumme/7-november-2007/han-tog-hans-hat-og-gik-hans-vej>)

Han tog hans hat og gik hans vej, sprogkiosken.dk, 22/8 2017 (<http://sprogkiosken.dk/2017/08/22/han-tog-hans-hat-og-gik-hans-vej/>)

Hans eller sin?, sprogkontoret.dk, 22/5 2014 (<https://sprogkontoret.dk/hans-eller-sin/>)

Han tog hans hat og gik hans vej, Fyens Stiftstidende, 9/1 2006 (<https://fyens.dk/bagsiden/han-tog-hans-hat-og-gik-hans-vej>)

By using *hans* as the locally bound form instead of *sin*, the sentence *Han tog hans hat* becomes both a joke about grammar and about stealing other people's head gear.

- (90) a. % **Han**₁ tog **hans**₁ hat og gik **hans**₁ vej.
 he took his hat and went his way
 He took his hat and walked away.
- b. **Han**₁ tog **sin**₁ hat og gik **sin**₁ vej.
 he took REFL's hat and went REFL's way
 He took his (own) hat and walked away.

The issue presumably arises because speakers of primarily some of the Jutlandic Danish dialects (historically) have used the reflexive *sin* in a very different way compared to the standard Danish norm. This has evidently been salient enough to be picked up as a(n undesirable) difference, and the trait (using *sin* in a non-standard way) is undoubtedly a grammatically heavily stigmatised feature, which can e.g. be seen in the comment sections on online articles or blog posts where commentators are quick to point it out if there is a grammatical error (prescriptively speaking) involving *hans* or *sin*, or even in comic strips such as the one in figure 2.3 by Maren Uthaug.

Jul Nielsen (1986) is a very comprehensive study of the ways that *sin* has been used in a variety of Jutlandic dialects, mainly based on dialectal observations made by professional (and semi-professional) linguists and on questionnaires with speakers born between 1880 and 1920. I briefly summarize the main points here and refer the interested reader to that work for more details (as well as e.g. Hagedorn and Jørgensen (2009) and Diderichsen (1939) for more on the status of *sin* in Jutland). Jul Nielsen generally concludes on the use of *sin* in Jutland that it has a limited use compared to standard Danish in some respects but expanded use in other respects (Jul Nielsen 1986, 43).

The classic description of Jutlandic is the one where *hans* or *hendes* is used as the bound possessive form instead of the reflexive *sin*, perhaps even to the complete exclusion of *sin*. This is too simplified a picture, however. With subjects that are clearly male or female it seems to be the case, especially in Western Jutland, that *hans* and *hendes* are used as bound possessive forms. Hagedorn and Jørgensen (2009) refine the analysis further and argue that *hans* and *hendes* are in fact mainly used when the subject is male or female *and* human (and in some cases gendered animals may be treated as "humans" and require the same kind of reference). See examples (91a)-

Vidste ikke, hvad der var værst.

Det er, fordi Leo fra min
klasse havde tabt hans
overstregningstusch, og
så tog han min, og så
tog jeg din.



Argh, hvordan kan
jeg hverken have fået
lært dig forskellen på
dit og mit eller på
hans og sin?!



Maren Uthaug

Figure 2.3: Comic strip by artist and author Maren Uthaug. Text: "Not sure what the worst thing was." "It is because Leo from my class had lost his marker and then he took mine, and then I took yours." "Argh, how have I managed to neither teach you the difference between yours and mine or *hans* and *sin*."

(91b) (from Åby in Jammerbugt in Northern Jutland and Mandø in Southern Jutland, respectively). All examples are from Jul Nielsen (1986) unless noted otherwise.

- (91) a. han₁ vandede hans₁ køer
he watered his cows
'he watered his cows'
- b. hun₁ bor ved hendes₁ kusine
she lives by her cousin
'she lives with her cousin'

With subjects that are neuter or common gender *sin* is used as the bound possessive form in all Jutlandic dialects, (92a)-(92c) (from Torsted and Rødding in North-Western Jutland, and Oksbøl in Southern Jutland). All Jutlandic dialects, in accordance with standard Danish, use *deres* as the bound possessive form when the subject is plural, according to Jul Nielsen (1986, 56), and this is apparently common enough that he does not even provide an example of it.

- (92) a. æ kalv er ved sin moder
the calf is by REFL's mother
'the calf is with its mother'
- b. æ skytteforening var der med sin ny fane
the rifle-club was there with REFL's new banner
'the rifle club was there with their new banner.'
- c. hvordan Oksbøl by har fået sit navn
how Oksbøl village has got REFL's name
'how the village of Oksbøl received its name'

The southern parts of Jutland (and the observations made of Danish speakers south of the border in northern Germany) show a system which is more reminiscent of German, namely one where *sin* is used as a masculine possessive pronoun (as in masculine and neuter *sein* in German). This means that *sin* is used as the bound form when the subject is male but when the subject is female the form *hendes* (En. *her*) is used, just like the difference between possessive *sein* (German *his* and *its*) and *ihr* (German *her* and *their*) (Jul Nielsen 1986, 51-2). Jul Nielsen provides a list of examples of locally bound *hendes* which come from speakers from various parts of Denmark, not just Southern Jutland. The point here is that *sin* and *hendes* could both be used as bound forms with feminine antecedents in those parts of Denmark with the greatest amount of German influence, but only *sin* was possible as the bound form with masculine antecedents.

The above-mentioned are the cases where the Jutlandic dialects seem to use *sin* in more limited contexts than standard Danish. Hagedorn and Jørgensen (2009, 30) argue that "Western Jutland has developed the stem *sin* into a normal non-reflexive pronoun with one extra condition, that the antecedent may only be a specific non-human referent or a non-specific human referent." Their main reason for arguing this is that *sin* in Western Jutland can be used in various contexts where it occurs entirely unbound. Examples of this are shown in (93). (93a) is from Klim in North-Western Jutland and *sin* occurs without a clear antecedent in the example. (93b) is from Bredsten in Southern Jutland and this *sin* has an antecedent in a different clause, i.e. an antecedent that does not c-command *sin* within the standard binding domain. The examples are from Jul Nielsen (1986, 70) and also cited in Hagedorn and Jørgensen (2009, 16) as examples (32) and (36).

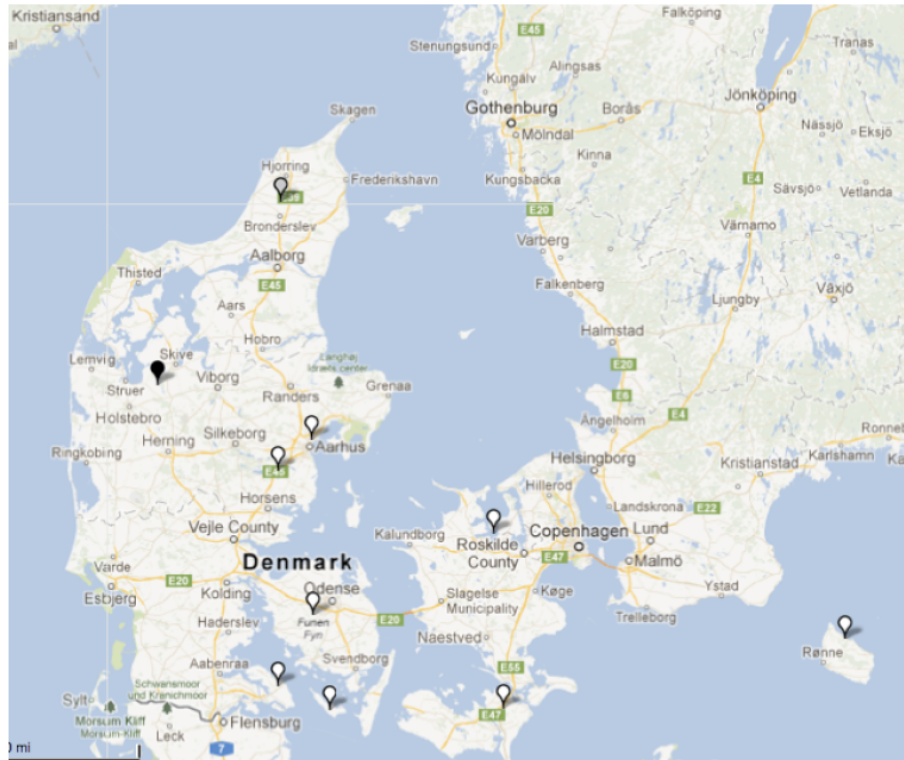
- (93) a. se sine øjne
se REFL's eyes

'Look at its eyes' (about a child's eyes)

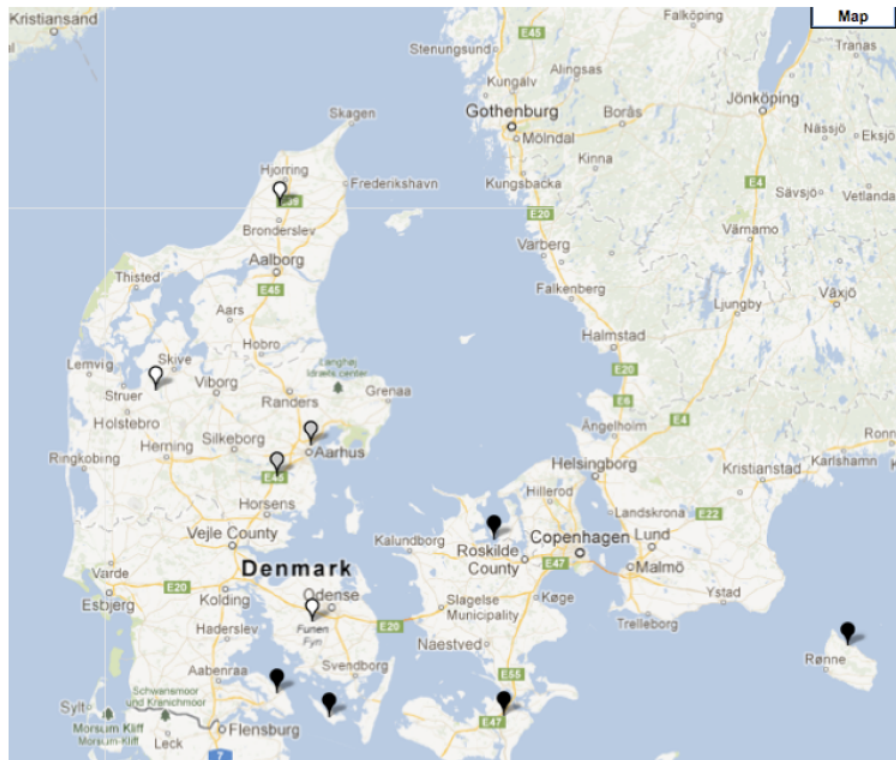
- b. den ville itte rigtig æde, og sin hale den hængte ned
 it would not really eat and REFL's tail it hung down
 'It did not really want to eat and its tail hung down.'

Lundquist (2014) discusses data from the Nordic Syntax Database (Lindstad et al. 2009b, also known as the Nordic Dialect Database and ScanDiaSyn). The maps reproduced in (2.4) show the distribution of acceptability of locally bound *sin* compared to locally bound *hans* across various Danish locations. The judgments from Western Jutland on Map 4 clearly stand out from the rest in Denmark, in that the Western Jutland respondents apparently do not accept locally bound *sin* at all (at least when *sin* has a person antecedent, as in the given example). Compare this to map 5 where speakers in most of the western parts of Denmark are reasonably happy to accept a locally bound *hans*.

Juel Jensen (2009b) investigates language data from the LANCHART database (see section 3.4) in order to describe the status of reflexive use in modern Danish. He finds that non-reflexive forms are used reflexively in 23 % of the clauses that have a person antecedent. This is most pronounced with speakers from Jutland where 47 % of clauses that contain a person antecedent contain the non-reflexive form, compared to speakers from Sjælland where 14 % of the clauses that contain a person antecedent contain the non-reflexive form. The dialect descriptions precisely find a difference in reflexive use with person antecedents. This difference is therefore in line with what could be expected if the dialectal difference still shows up to some extent even in speakers who mainly use standard Danish, which is the majority of the LANCHART speakers. Jensen's data show that the variation in reflexive use with person antecedents is indeed present in the standard language. He speculates that this is not so much a sign of language-internal grammatical changes but a general loosening of the standard language norm which has otherwise prescribed the use of *sin* in reflexive contexts. This is in line with what I find in chapter 3 where the data shows that there has been variation in reflexive use for as long as we have language data to show it. It is also in line with the data from Lundquist (2014) which show broad acceptability of locally bound *hans*. The general pattern in Danish is that there seems to be a nation-wide variability in reflexive use where a fair number of speakers accept or use locally bound singular non-reflexives with person antecedents. This variability is not new and probably, for the same reason, also not a sign of a substantive change in



Map 4: Local singular antecedent of *sin*. (#243: *Jeg tror at Jon holder af sin katt. 'I believe that John likes his cat.'*) (White = high score, grey = medium score, black = low score)



Map 5: Local singular antecedent of *hans*/short distance pronoun. (#244: *Jeg tror at Jon holder af hans katt. 'I believe that John likes his cat.'*) (White = high score, grey = medium score, black = low score)

Figure 2.4: Maps of bound *sin* and *hans* in multiple locations in Denmark. Map 4 shows judgments for the sentence "Jeg tror at Jon₁ holder af sin₁ kat." Map 5 shows judgments for the sentence "Jeg tror at Jon₁ holder af hans₁ kat." Discussed in Lundquist (2014) and originally drawn from Lindstad et al. (2009b).

progress in the grammar, contra the speculation in Lundquist (2014, 533). An interesting avenue of further research could be to investigate the phonetics of bound versus unbound *hans* (and the other relevant pronouns). I find, based very informally on the casual speech of closer acquaintances, that bound *hans* is more phonetically reduced than unbound *hans*. This could point to an analysis where the bound pronoun variation could in fact be seen as a phonetic (PF realisation) variant of *sin* with the reflexive realised by certain speakers as phonetically reduced *hans* (or *hendes*) rather than *sin* while the LF representation stays the same.

Howe (1996, 101) summarizes a development where the Continental West Germanic languages (Frisian, Afrikaans, Dutch, German, Old Saxon) replaced the masculine and neuter singular genitives *is* and *es* with *sin*. It surfaces in modern German as neuter and masculine *sein* and in modern Dutch as *zijn*- (and e.g. English preserved the original form in masculine possessive *his*). The various *sein* forms in these languages are regular pronouns, not reflexives, and just like the Jutlandic *sin* they can be used unbound. The (majority of) the Jutlandic dialects have in common with the Continental West Germanic languages both that the reflexive possessive has been repurposed to a regular, non-reflexive, pronoun and that definiteness is marked on prenominal freestanding articles rather than on postnominal suffixes (as in the rest of the Danish language area and the other Scandinavian languages). These two factors may well be related, as explored in Reuland (2011) and further in Despić (2015) who find a pattern where languages without definiteness marking as well as "[l]anguages with prenominal (article-like) definiteness marking [...] systematically lack reflexive possessives" (Despić 2015, 203). It is beyond the scope of this thesis to follow this track further but it could certainly be interesting to explore whether the article system and reflexive system of Western Jutland has followed the same developmental path as the Continental West Germanic languages mentioned above. See also 132 and onwards in this thesis for more speculation on the influence from German on the Danish reflexives.

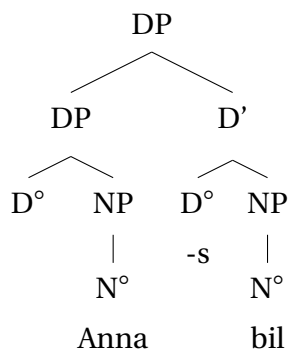
2.5 Plural antecedent *sin*

2.5.1 The basic structural position of *sin* in the clause

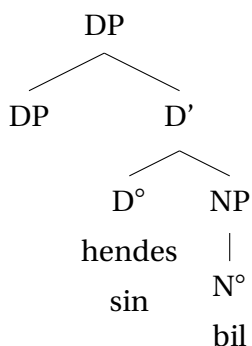
Vikner (2014, 202) argues for an analysis of the DP with possessors as illustrated in (94) and (95). The possessor DP *Anna* is in the specifier position and the possessive

clitic *-s* is in the head position of the main DP. Possessive pronouns such as *hendes* or *sin* occupy the same position in D° as the possessive clitic *-s*. This analysis originally comes from Fiva (1987).

(94)



(95)



I adopt this basic structure so that a full, albeit still simplified, syntactic tree with the possessive reflexive *sin* could look like the large tree in figure 2.5. The subject *Alanna* is born in VP-spec, moves through IP-spec, and up into CP-spec (because Danish is a V2 language). The finite main verb moves from V° , through I° and into C° . The reflexive possessive *sin* sits as the head of the DP complement of the verb. The subject, which is also the antecedent of *sin*, is born in a local relationship with *sin* and c-commands *sin* both from its initial position and its final position. The triangle above the subject *Alanna* is used to abbreviate a bit of the less directly relevant tree structure.

Several researchers have made other suggestions as to how to place *sin* in a tree structure. E.g. Delsing (1993) and, building on Delsing, Despić (2015) argue that *sin* is in fact born as the head of a dedicated Poss(essor)P(hrase) as the complement of D° . The noun phrase is then the complement of Poss°. The possessor head (*sin*) moves to D° in order to satisfy an Edge Feature (EF) of the DP, according to Despić (2015). This would also presumably bring about the desired consequence that DPs with *sin* are always definite. The specific analysis does not bear directly on the topic at hand,

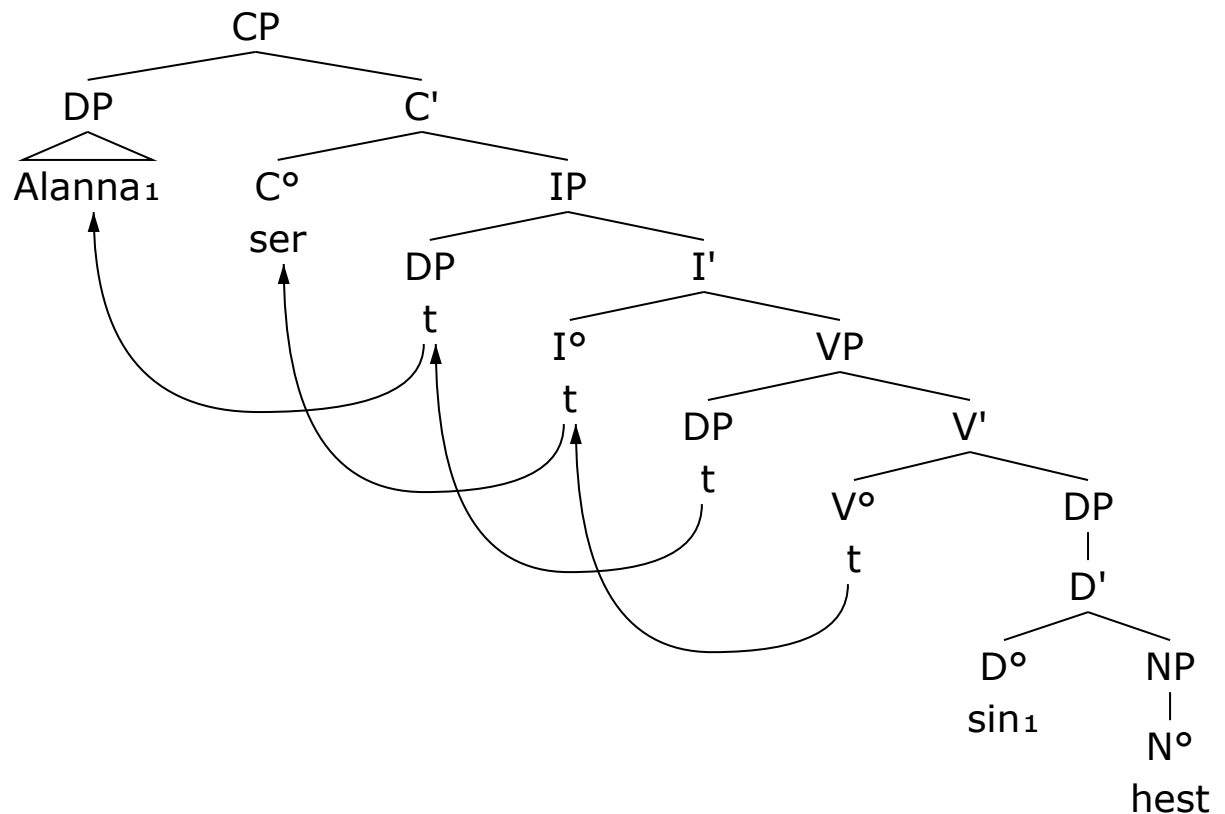


Figure 2.5: Syntactic tree depicting the Danish sentence *Alanna ser sin hest* (En. *Alanna sees REFL's horse*).

however, and I have not been able to build an argument for the Danish-specific restrictions on *sin* that follows in a non-ad hoc way from its position in the tree. Consequently, I stand by the simpler analysis from figure 2.5 but make note that there may be good arguments for expanding the structure further.

2.5.2 Number restriction on Danish *sin* - the basics

In the Danish reflexive system summarized in figure 2.2 on page 76, *deres* stands out in being the only 3rd person item with no *selv* or *sig* restrictions on it at all. The example sets in (96)-(102) illustrate the basic traits of *sin* and *deres* in standard Danish. With third person singular antecedents of any gender, the non-reflexive possessive cannot occur locally bound and *sin* must be used instead, (96).

- (96) a. ***Hun**₁ elsker **hendes**₁ barn.
 she loves her child
 'She loves her child.'

- b. **Hun**₁ elsker **sit**₁ barn.
 she loves REFL's child
 'She loves her child.'

The judgments are reversed when the antecedent is third person plural. That is, for third person plural antecedents, *sin* is not grammatical and the non-reflexive possessive is used instead, (97). *Sin* is *number-restricted* in Danish in the sense that it generally does not allow plural antecedents.

- (97) a. **De**₁ elsker **deres**₁ barn.
 they love their child
 'They love their child.'
- b. ***De**₁ elsker **sit**₁ barn.
 they love REFL's child
 'They love their child.'

The number restriction is only applied to the antecedent. A singular subject such as *hun* in (98) may be the antecedent of *sin* contained in a singular DP, (98a), as well as the antecedent of *sin* contained in a plural DP, (98b).

- (98) a. **Hun**₁ elsker **sit**₁ barn.
 she loves REFL's child
 'She loves her child.'
- b. **Hun**₁ elsker **sine**₁ børn.
 she loves REFL's children
 'She loves her children.'

A plural subject such as *de* (En. *they*) may be the antecedent of *deres* contained in a singular DP, (99a), as well as the antecedent of *deres* contained in a plural DP, (99b). Note also here that the morphological form of *deres* is invariant with respect to the ϕ -features of the containing DP while the morphological form of *sin* (*sit*, *sine*) depends on the number and gender of the nominal head in the DP that contains the reflexive possessive.

- (99) a. **De**₁ elsker **deres**₁ barn.
 they love their child
 'They love their child.'
- b. **De**₁ elsker **deres**₁ børn.
 they love their children
 'They love their children.'

Deres is a possessive pronoun and not a dedicated reflexive. This means that *deres* can be used in several ways that *sin* cannot, e.g. with a non-c-commanding antecedent as in (100a)-(100b) or entirely unbound as in (100c)-(100d).

- (100) a. **Peter og Marias₁** nabo kommer med gaver til **deres₁** barn.
Peter and Maria's neighbour comes with gifts to their child
'Peter and Maria's neighbour brings gifts for their child.'
- b. ***Peter og Marias₁** nabo kommer med gaver til **sit₁** barn.
Peter and Maria's neighbour comes with gifts to REFL's child
'Peter and Maria's neighbour brings gifts for their child.'
- c. **Vi₁** elsker **deres₂** have.
we love their garden
'We love their garden.'
- d. ***Vi₁** elsker **sin₂** have.
we love REFL's garden
'We love their garden.'

With singular antecedents and non-local binding, *sin* and the non-reflexive possessive are both possible, (101).

- (101) a. **Harry₁** overtalte **vennerne₂** til [**PRO₂** at fotografere **hans₁** ugle].
Harry persuaded friends.DEF to to photograph his owl
'Harry persuaded the friends to photograph his owl.'
- b. **Harry₁** overtalte **vennerne₂** til [**PRO₂** at fotografere **sin₁** ugle].
Harry persuaded friends.DEF to to photograph REFL's owl
'Harry persuaded the friends to photograph his owl.'

When the antecedent is third person plural and binding is non-local, *sin* is still not possible and *deres* is used instead, (102).

- (102) a. **De₁** overtalte **mig₂** til [**PRO₂** at fotografere **deres₁** ugle].
They persuaded me to to photograph their owl
'They persuaded me to photograph their owl.'
- b. ***De₁** overtalte **mig₂** til [**PRO₂** at fotografere **sin₁** ugle].
They persuaded friends.DEF to to photograph REFL's owl
'They persuaded me to photograph their owl.'

In modern standard Danish this number restriction is only true for *sin*, not for *sig* or *sig selv*, the other 3rd person reflexive pronouns. Both *sig* and *sig selv* may be locally

bound by 3rd person antecedents of any gender and number, (103a)-(103b), and the non-reflexive plural *dem* may not be locally bound, (103c).

- (103) a. **Hun**₁ underholder **sig** **selv**₁.
 she entertains REFL self
 'She entertains herself.'
- b. **De**₁ underholder **sig** **selv**₁.
 they entertain REFL self
 'They entertain themselves.'
- c. ***De**₁ underholder **dem** **selv**₁.
 they entertain them self
 'They entertain themselves.'

Sig and *dem* are both possible when the antecedent is non-local, (104).

- (104) a. **De**₁ overtalte **mig**₂ til [**PRO**₂ at fotografere **dem**₁].
 They persuaded me to to photograph them
 'They persuaded me to photograph them.'
- b. **De**₁ overtalte **mig**₂ til [**PRO**₂ at fotografere **sig**₁].
 They persuaded me to to photograph REFL
 'They persuaded me to photograph them.'

Table 2.1: Phi-features of some English pronouns

	Pronoun	Phi-features
Singular	I	[1st person, singular]
	you	[2nd person, singular]
	she	[3rd person, singular, feminine]
	he	[3rd person, singular, masculine]
	it	[3rd person, singular, neuter]
	they	[3rd person, singular]
Plural	we	[1st person, plural]
	you	[2nd person, plural]
	they	[3rd person, plural]

2.5.3 Binding, feature matching, and feature impoverishment

An intuitive way of understanding why reflexives must be bound is that they are somehow deficient on their own, so that they need another element that can provide them with meaning or reference. This deficiency can be described in terms of ϕ -features. A ϕ -feature, to reiterate a previous point, is an umbrella term for the morphological features **person**, **number**, and **gender**. The personal pronouns are to varying degrees specified for these features, as shown in table 2.1 which gives a simplified impression of the featural content of some of the English pronouns. (I do not commit myself to whether or not e.g. singular should be described as a meaningful feature or whether singular is in fact better described as the lack of plural, as suggested by e.g. Nevins (2007), Nevins (2011), Ackema and Neeleman (2018), and Ackema and Neeleman (2019)).

I provide a similar table for some of the Danish reflexive and non-reflexive pronouns in table 2.2. In this sense, the Danish reflexive pronoun *sig* is certainly rather feature impoverished, but *sin* is less so if we assume that its number-restriction is because *sin* is specified for singular.

In English, the reflexive pronouns do seem to have some inbuilt featural content, in that they match the regular personal pronouns in person and number (*him* – *himself*, *them* – *themselves*, to give two examples).

The examples in (105) (not an exhaustive list) show that the English reflexives cannot be bound if they have person, gender or number features that mismatch that of their antecedent.

(105) a. * We₁ photographed yourself₁.

Table 2.2: Phi-features of Danish reflexives *sig* and *sin*

Pronoun	Phi-features
sig	[3rd person]
sin	[3rd person, singular]
hans	[3rd person, singular, masculine]
hendes	[3rd person, singular, feminine]
deres	[3rd person, plural]

- b. We₁ photographed ourselves₁.
- c. You₁ photographed yourself₁.
- d. * You₁ photographed ourselves₁.
- e. * Mary₁ photographed himself₁.
- f. * [James and John]₁ photographed himself₁.

Antecedent and reflexive must be morphologically compatible. How compatible they must be is another question. It cannot be the case that they must match completely. Even for a clearly grammatical sentence like (105b), the antecedent *we* (NOMINAL) and the reflexive *ourselves* (ACCUSATIVE) do not match in terms of Case.

Burzio (1991, 87-89) defines anaphors (reflexives, in my terms) as nominals that have no ϕ -features. He assumes that antecedent and reflexive must agree in order for binding to occur. He defines a concept of *pseudo-agreement* in order to capture the fact that there are many cases where the antecedent of the (assumed featureless) reflexive does have features:

(106) Agreement = α agrees with β if:

- a. (Strict Agreement) α and β have identical Φ -features, or
- b. (Pseudo-Agreement, Italian)
 - i. β has no gender, no number, no person, and
 - ii. α is third person

He argues that there are language-specific differences in what type of antecedent that is allowed to pseudo-agree and provides a spectrum from most permissive to least permissive. I combine this with his table (21) which combines specific languages with the pseudo-agreement spectrum (Burzio 1991, 92):

(107) [Zero features], pseudo-agrees with:

- a. all (Russian: reflexives may take first and second person antecedents as well as third)
- b. 3rd (Romance: reflexives may take third person antecedents)
- c. 3rd sing (Danish possessives: *sin* may take third person singular antecedents)
- d. impersonals (French stressed object *soi* may take an impersonal antecedent)
- e. nothing (e.g. English possessives, West Flemish: no dedicated reflexive forms exist)

Franks and Schwartz (1994) argue that Burzio's pseudo-agreement is unmotivated and that it cannot be the case that anaphors should be defined as featureless. They provide an alternative definition where they incorporate NON-DISTINCTNESS rather than identity of features.

- (108) a. If α binds β , then β agrees with α .
 b. β agrees with α iff β is non-distinct from α in Φ -features.
 (Franks and Schwartz 1994, 234)

They suggest for Danish that *sin* should then be specified as [-1st pers, -2nd pers, -pl, **u**gender] (238), i.e. that the Danish *sin* may be bound by antecedents that are non-distinct from this specification. This would be an antecedent that is *not* first and second person and *not* overtly plural. (It must be noted that many of their Danish examples contain errors, but none that directly impact the point above.)

Newer works with a similar approach to anaphora as somehow featurally underspecified are e.g. Safir (2004) and Reuland (2005) et seq. Safir suggests that the Danish *sin* (and *sig*, which he also claims is restricted to singular antecedents, counter to fact with regard to modern Danish) is "specified for singular in Danish, but not in Norwegian and Swedish" (Safir 2004, 72). Reuland, contra (but not in response to) Safir, claims that it is "uncontroversial, [that] SE-anaphors such as Dutch *zich*, Icelandic *sig*, etc., are not specified for the features number and gender" (Reuland 2005, 510). In a later work, Reuland writes:

If gender were the sole factor responsible [for why 3rd person pronominals cannot be bound] one would expect that bound *him* in *John washes him* and its cognates is out, but bound *them* in *The girls are washing them* is well formed. There is no Germanic language where this is the case (in fact I do not know of any language that has this pattern, but for present

purposes the Germanic case is enough).

(Reuland 2011, 162)

This observation does not hold generally. Sundaresan (2020, 432) notes that "Korean *caki* and Dravidian *taan* are underspecified for gender alone: i.e. can take antecedents of any gender, but these must be 3SG; German *sich* (and its Germanic relatives) seem to be underspecified for both gender and number...".

This pattern of gender-underspecification is also, I would argue, precisely what we see with *sin* in present-day Danish and with both *sin* and *sig* in earlier stages of Danish. In older stages of Danish, at least from around 1200 to 1700-1900, both *sin* and *sig* were used primarily with singular antecedents and *deres* and *dem* with plural antecedents (see K.M. Pedersen 2017 and chapter 3 of this thesis), *pace* Reuland's description above. *Dem* as a local reflexive can still be found in speech in modern (dialectal) Danish, but it is extremely infrequent (again see chapter 3). The example pairs in (109) show a (necessarily simplified) general picture of the development in usage of *sig* in Danish over time.

(109) a. Present-day Danish (2023)

- i. Hun₁ skammer sig₁ over sit liv... (KorpusDK)
she shames REFL over REFL's life
'She is ashamed of her own life...'
- ii. Jo, de₁ skammer sig₁ nu over, at... (KorpusDK)
yes they shame REFL now over that
'Yes, they are now ashamed that...'

b. Danish in 1800s⁴

- i. *Bil* en Pige, som Maanen₁ tog til sig₁ (1808)
Bil a girl that moon.DEF took to REFL
'Bil a girl that the moon adopted'
- ii. saa kloge Folk₁ [...] maae skamme dem₁ ved, at de... (1832)
so clever people must shame them by that they
'people that clever must be ashamed that they...'

c. Danish in the 1500s⁵

⁴Examples from N.E.S. Grundtvig: *Nordens mytologi eller Udsigt over Eddalæren* (1808) and *Nordens mytologi eller Sindbilled-Sprog* (1832). Note that *sig* is the most frequent form used with plural antecedents at this point in time, so the usage in the example is old-fashioned.

⁵Examples from Chr. III's Bible (1550). *Sig* and *dem* are both used with plural antecedents in the text. The relative frequencies, i.e. whether *sig* or *dem* is the most frequent form, vary from text to text.

- i. Men Lamech₁ tog sig₁ to Hustruer (Genesis 4.19)
but Lamech took REFL two wives
'But Lamech took two wives...'
- ii. Menniskene₁ ville icke mere lade dem₁ straffe aff min Aand
people.DEF would not more let them punish by my spirit
(Genesis 6.3)

'The people would not let themselves be punished by my spirit any longer.'

d. Danish in the 1200s⁶

- i. ... hwæm sum hun₁ wil at giftæ sik₁ (Book 1, ch. 33)
who that she will to marry REFL
'... then she may ask whoever she wants to give her away in marriage.'
- ii. Tha mughæ skipær₁ gøræ thæm₁ skip sialf (Book 3, ch. 5)
then must/can? shipmen make them ship self
'Then the sailors must/can? make themselves a ship.'

Animacy and/or gender as an explanatory factor

I am not sure whether there is a meaningful difference between saying that *sin* is specified for singular (as Safir 2004) or saying that *sin* can occur with antecedents that are not plural and which have a gender specification (as would be the natural conclusion on Franks and Schwartz 1994, Reuland (2011), and Sundaresan (2020)). One possible point in favour of the gender-based analysis is the corpus data that will be discussed in chapter 4. I find a much higher occurrence of *sin* used with inanimate plural antecedents than with animate plural antecedents (compared to how *sin* normally occurs) and one way of explaining this difference is that the inanimates typically are not marked for natural gender.

Several researchers have made connections between binding and animacy in their work, both for Danish and other languages. Their focus is strictly on animacy rather than gender so it is not directly clear whether (natural) gender is a deciding factor in these cases, or whether (in)animacy itself is enough to make a difference. Reuland and Zubkov 2022 note that Russian shows non-complementarity in local binding by inanimates, i.e. that inanimate antecedents may either bind reflexives or non-reflexives locally. Hansen and Heltoft (2011) claim that something similar is the case for Danish:

⁶Examples from *Jyske Lov* (1241).

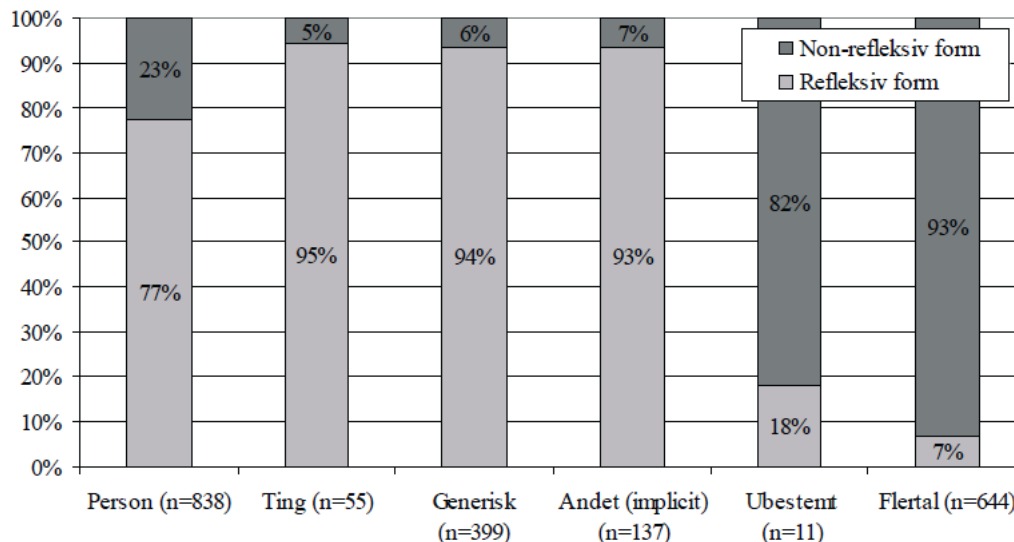
"I moderne usus synes der at være en tendens til at der henvises ikke-refleksivt, altså med personligt pronomen, til -HUM, også i situationer hvor refleksiv henvisning er nødvendig ved +HUM" (Hansen and Heltoft 2011[597]) (my own English translation: "Modern usage seems to show a tendency to refer non-reflexively, i.e. with a personal pronoun, to inanimates, also in cases where reflexive reference is necessary with animates"). They provide the examples in (110) which both seem slightly suspect. (110a)-(110b) presumably have no c-command between antecedent and bound form, which could indicate that the fact that a reflexive in the position still seems acceptable is due to hyper correction or something similarly extra-syntactic. Examples (110c) and (110d) sound equally good (with the reflexive) or equally bad (with the non-reflexive) to my Danish native speaker ears.

(110) **GDS: (Spurious) examples of non-complementarity of binding by inanimates**

- a. der går **fabrikant Emsig**₁ med **sine/*hans**₁ børn.
there walks industrialist Emsig with REFL's/his children
'Industrialist Emsig is walking there with his children.'
- b. der lå så kirken med sine/dens prægtige tårne.
there lay then church.DEF with REFL's/its magnificent towers
'That's where the church with its magnificent towers lay.'
- c. chaufføren er blevet lykønsket af sine/*hans passagerer.
driver.DEF is been congratulated by REFL's/his passengers
'The driver has been congratulated by his passengers.'
- d. bussen er blevet ramponeret af sine/dens passagerer.
bus.DEF is been by wrecked REFL's/its passengers
'The bus has been wrecked by its passengers.'

Juel Jensen (2009b) investigates reflexive use in the Danish spoken language corpus LANCHART. Interestingly, he finds that there is much more variation in the choice between reflexive and pronominal when the antecedent can be categorized as a Person (i.e. as [+HUM]) than when the antecedent can be categorized as a Thing (i.e. as [-HUM]), his Figur 1, Juel Jensen (2009b)[7]). Figure 2.6 sums up parts of his results and it is notable for the current topic that the *Ting* (Thing) column contains only 5 % examples with bound pronominals compared to the 23 % bound pronominals in the Person column. Accordingly, this data makes it seem more correct to claim that sentences with *animate* antecedents show a greater tendency to bind non-reflexives. Juel Jensen's *Ting* category does not only contain inanimates, however, as e.g. his example

- Aarhus universitet/PhD/Afhandling/Figures/REFLEXIVE-DESCRIPTION-juel-jensen-2009-figur-1.png - Aarhus universitet/PhD/Afhandling/Figures/REFLEXIVE-DESCRIPTION-juel-jensen-2009-figur-1.png



Figur 1: Refleksivt anvendte pronominer, samlet

Figure 2.6: Figur 1 from Juel Jensen (2009b)[7]. The chart shows the proportion of reflexive to bound pronominal use in the LANCHART data.

(8) has the animate subject *hunden* (Eng. 'the dog') as an exemplar of the sentences contained in the *Ting* category. His *Ting* category is also much smaller (n=55) than his *Person* category (n=838).

I find precisely one example of an inanimate antecedent of a non-reflexive in the LANCHART data, shown in (111).

(111) **The single example of an inanimate antecedent of a locally bound non-reflexive in LANCHART**

- a. jeg plejer at sammenligne det med en med en men
 I use to compare it with an with an with
en øh olieplatform₁ der simpelthen får revet en
 an uh oil-drilling-platform that basically gets
 af **dens₁** grund søjler væk og står i stormvejr
 ripped one of its support-pillars away and stands in
 stormweather

'I usually compare it to an to an to an uh oil drilling platform that basically gets one of its supporting pilers.'

In the other direction, I find several examples of inanimate antecedents of *sin* in the same LANCHART data and a selection of them are reproduced below.

(112) **Examples of inanimate antecedents of *sin* in LANCHART.**

- a. og vores organisation₁ har sine₁ blade
and our organisation has REFL's magazines
'and our organisation has our magazines'
- b. det er fantastisk at Citroen kan tillade sig at gå væk fra sit
it is amazing that Citroen can allow REFL to go away from REFL's
affjedringssystem
suspension-system
'it is amazing that Citroen can allow themselves to walk away from their
suspension system'

All in all, I find nothing in the Danish spoken corpus to substantiate that inanimates are more likely to occur as antecedents of locally bound non-reflexives. Jul Nielsen (1986) describes reflexive usage in various older Jutlandic dialects and describes a situation where non-reflexives occur locally bound with animate (and especially gendered) antecedents. This is precisely the reverse of the claim that inanimates tend to occur as antecedents of non-reflexives. It also supports the notion that gender may be a relevant factor for (at least Jutlandic) Danish speakers in terms of reflexive choice, which then could carry over into the use of plural antecedent *sin*.

2.5.4 Agreement variation also regularly present in the standard language

There is, irrespective of the specific mechanism, variation in the use of *sin* in modern Danish in terms of whether or not it allows plural (or not-gender-specified) antecedents. This variation is very reminiscent of a previous development in the use of *sig* which, in an earlier stage of the language, also mainly occurred with singular antecedents (a development which is discussed in greater detail in chapter 3). This variation, which could be a sign of syntactic change in progress, is what the remainder of this thesis is about. However, in order to discuss the actual cases of variation from the norm, we need a firmer hold on what the norm is and what factors that could be contributing to the variation found. This is the topic of the remaining parts of this chapter.

Ahn (2019) reviews several cases where an English reflexive may mismatch the antecedent. One such case is that of *imposters* where a third-person nominal can be interpreted as the (first person) speaker.

- (113) a. **Mommy and Daddy**₁ need some time to **themselves**₁.
 b. **Mommy and Daddy**₁ need some time to **ourselves**₁.
 (Ahn (2019, example 6))

Mommy and Daddy can either be the antecedent of the third person plural *themselves*, (113a), in which case the antecedent and the reflexive are directly matched in terms of ϕ features. *Mommy and Daddy* can also be the antecedent of first person plural *ourselves*, (113b). The antecedent and reflexive are overtly mismatched in terms of person but the sentence is still acceptable.

Another case of mismatch is with collective nouns. These are nouns which are morphologically singular but describe a group of individuals. The collective *U.N.* may either be the antecedent of singular *itself* as in (114a) or of plural *themselves* as in (114b). The first example is a feature match in terms of number (singular *U.N.* – singular *itself*) and the second example is a feature mismatch (singular *U.N.* – plural *themselves*).

- (114) a. The **U.N.**₁ finds **itself**₁ in a difficult position.
 b. The **U.N.**₁ finds **themselves**₁ in a difficult position.
 (Ahn (2019, example 12))

Ahn further notes that not all English antecedent types may take part in mismatches. Specifically, reflexives that are specified as 3rd person singular are not allowed to mismatch with the features of their antecedent. He speculates that this is due to the gender specification on the 3rd person singular reflexives (Ahn (2019, section 3)), which is not present on e.g. the 3rd plural English reflexives.

Similar kinds of agreement variation are also present in Danish with regard to the use of *sig selv* (~themselves) and *os selv* (~ourselves). There is perhaps an interpretive difference in that (115a) can be read either distributively (where *mom and dad* each get some time alone but not necessarily together as a couple) or collectively (where *mom and dad* get some time together as a couple). The distributive reading is not readily available with *os selv*.

- (115) a. **Mor og far₁** har brug for noget tid til **sig selv₁**.
 Mom and dad have need for some time to REFL self
 'Mom and dad need some time to themselves.'
- b. **Mor og far₁** har brug for noget tid til **os selv₁**.
 Mom and dad have need for some time to us self
 'Mom and dad need some time to ourselves.'

A similar set of 'imposter' examples can be made with the possessives *deres* (En. *their*) and *vores* (En. *our*) but not with *sin*. This is not too surprising, as the agreement mismatch that we find here is not dependent on number but rather on person.

- (116) a. ***Mor og far₁** har brug for noget tid til **sin₁** hobby.
 Mom and dad have need for some time to REFL's hobby
 'Mom and dad need some time for their hobby.'
- b. **Mor og far₁** har brug for noget tid til **deres₁** hobby.
 Mom and dad have need for some time to their hobby
 'Mom and dad need some time for their hobby.'
- c. **Mor og far₁** har brug for noget tid til **vores₁** hobby.
 Mom and dad have need for some time to our hobby
 'Mom and dad need some time for our hobby.'

The mismatch with collectives is illuminating in a different way with regards to *sin* and *sig* since the mismatch here specifically is due to number. I have changed the wording of Ahn's specific example to make it more idiomatic in Danish. *Sig* and *sig selv* are number-neutral in Danish and this has the consequence that there is no possibility of a number mismatch in the examples in (117). *Sig selv* in (117a) presumably covers both options (singular and plural) and none of the alternative forms in (117b) are possible.

- (117) a. **FN₁** spænder ben for **sig selv₁**.
 U.N. trip leg for REFL
 'The U.N. trips itself/themselves up.'
- b. ***FN₁** spænder ben for **dem selv₁** / **den selv₁** / **det selv₁**.
 U.N. trip leg for them self it.C self it.N self
 'The U.N. trips itself/themselves up.'

The possibility of a number mismatch is visible with *sin* and *deres* in (118). The two examples are attested examples from a quick Google search. *Deres* is available as

an alternative, in contrast with e.g. *dem selv* in (117b) presumably precisely because *sin* does not generally allow plural antecedents.

- (118) a. **FN**₁ har **sin**₁ egen hjemmeside. (UNRIC.org)
 U.N. has REFL's own website
 'The U.N. has its own website.'
- b. **FN**₁ har **deres**₁ egne frimærker. (NewYorkCity.dk)
 U.N. has their own stamps
 'The U.N. has their own stamps.'

There are certainly cases even in the language of those Danes who do not otherwise accept plural antecedent *sin* (such as myself) where *sin* and *deres* are both acceptable. This is the case for (some, perhaps all) collectives and some quantifiers. According to Hansen and Heltoft (2011, 590) *deres* may only refer to a singular subject when this subject is a semantically plural nominal. Their example of this is provided in (119a). The intuition is that SF, a Danish political party, could either be interpreted as a (morphosyntactically) singular entity, or as a collection of individual politicians. With the semantically plural interpretation, *deres* is possible and grammatically unmarked. The sentence is just as acceptable with *sin* rather than *deres*, as in the constructed counterpart (119b).

- (119) a. **SF**₁ vil i den kommende folketingssamling søge at koncentrere
 SF will in the upcoming parliament.session seek to concentrate
deres₁ bestræbelser om skattelovgivningen
 their efforts about tax.legislation
 '(The political party) SF will in the upcoming session in parliament attempt to concentrate their efforts on tax legislation'
- b. **SF**₁ vil i den kommende folketingssamling søge at koncentrere
 SF will in the upcoming parliament.session seek to concentrate
sine₁ bestræbelser om skattelovgivningen
 REFL's efforts about tax.legislation
 'SF will in the upcoming session in parliament attempt to concentrate their efforts on tax legislation'

Attested examples with *sin* instead of *deres* are readily available in KorpusDK, e.g. (120). These are examples where the antecedent is a political party that is the binder of either a *sin* or a *deres*. The results of a cursory corpus check indicate that *sin* is the most commonly used pronoun in this kind of context at a ratio of 2-4 to 1 compared to *deres* in KorpusDK.

- (120) a. ... når **SF**₁ afgør **sin**₁ stilling til den næste folkeafstemning.
 when SF determine REFL's position to the next referendum
 '... when SF determine their position about the next referendum.'

(KorpusDK)

Corbett (Corbett 2023 and various earlier works) has suggested the term "semantic agreement" as (part of) a formalization of similar findings of agreement variation in a range of languages. He illustrates the agreement variation contrast for English with the minimal pair in (121).

- (121) a. This family have lost everything
 b. This family has lost everything

"The plural target *have* in [(121a)] shows what many would label semantic agreement, while singular *has* is also found [(121b)]" (Corbett 2023, 1). In this terminology, the *deres* in (119a) would be a case of semantic agreement because it agrees with the plurality of individuals of the group. The *sin* in (119a) would be labelled syntactic agreement because it agrees with the morphologically singular number of the group.

Corbett proposes an Agreement Hierarchy (e.g. Corbett 2023) to capture the fact that not all positions in the clause are as likely to be able to perform semantic agreement.

- (122) THE AGREEMENT HIERARCHY
 attributive > predicate > relative pronoun > personal pronoun

The claim of the Agreement Hierarchy is that elements to the right on the scale are more likely to allow semantic agreement than elements to the left on the scale. Sturt (Sturt 2022, 3, his Table 1) provides examples in English of the elements of the hierarchy and I give his examples in (123). Sturt does not provide examples with relative pronouns "because English relative pronouns are not marked for number" (Sturt 2022, 3). This may be a misunderstanding of where the agreement is supposed to be found. Corbett himself does provide English examples of the relative pronoun category, where it is not the relative pronoun itself that agrees overtly but the verb inside the relative clause, and I consequently borrowed one of his examples and found another similar one in The Corpus of Contemporary American English (COCA).

- (123) a. ATTRIBUTIVE (could be DETERMINER - NOUN)
 i. Syntactic agreement: this family

- ii. Semantic agreement: *these family
- b. PREDICATE (could be SUBJECT - VERB)
 - i. Syntactic agreement: the group agrees
 - ii. Semantic agreement: the group agree
- c. RELATIVE PRONOUN
 - i. Syntactic agreement: ... you are defending an illegal group who has usurped the rights of the Eritrean people. (COCA)
 - ii. Semantic agreement: There was (?were) a witch and a wizard who were (*was) living in the moat. (Corbett 1979, example (27), 207)
- d. PERSONAL PRONOUN (could be ANTECEDENT - ANAPHOR)
 - i. Syntactic agreement: the group filmed itself
 - ii. Semantic agreement: the group filmed themselves

The table in (124) is a Danish parallel to the English examples.

- (124)
- a. ATTRIBUTIVE (could be DETERMINER - NOUN)
 - i. Syntactic agreement: denne familie
 - ii. Semantic agreement: *disse familie
 - b. PREDICATE (could be SUBJECT - ADJECTIVE as Danish has no verbal agreement)
 - i. Syntactic agreement: gruppen er enig
 - ii. Semantic agreement: gruppen er enige
 - c. RELATIVE PRONOUN
 - i. Syntactic agreement: Udspillet møder ligeledes opbakning fra SF, der er enig-Ø i at fastsætte et klimamål for sundhedsvæsenet.⁷ (information.dk)
 - ii. Semantic agreement: ... at det er en samlet ledelse hos de radikale, der er enig-e i Hegaards beslutning om at trække sig.⁸ (nyheder.tv2.dk)
 - d. PERSONAL PRONOUN (could be ANTECEDENT - ANAPHOR)
 - i. Syntactic agreement: gruppen₁ solgte sin₁ film⁹

⁷The proposal also receives support from (the political party) SF who agree-SG.C with setting a climate goal for the health care system.

⁸that it is a united party leadership within the Danish Social Liberal Party who agree-PL with Hegaard's decision to resign.

⁹the group sold REFL's movie

ii. Semantic agreement: *gruppen*₁ *solgte deres*₁ *film*¹⁰

Agreement variation, or a difference between syntactic and semantic agreement, can be found in both English and Danish in the three lower (or right-most) categories in Corbett's Agreement Hierarchy. We would also expect, if Corbett is right in calling his hierarchy a hierarchy, that there is at least as much variation in the use of e.g. *sin* and *deres* (which belong to the PERSONAL PRONOUN group) as there is in predicative adjectival inflection (which belongs to the hierarchically higher PREDICATE group).

It seems that *sin* and *deres* (for some speakers, if not necessarily for all) are interchangeable with (syntactically singular) collective nouns (e.g. (119)). These nouns also typically allow either singular or plural inflection on their predicative adjectives. One such example pair is shown in (125) where the same news media (*Sjællandske Nyheder*) has posted both types of inflection with the same adjective (*glad-Ø, glad-e*) and the same political party (*SF*).

- (125) a. SF er glad-e for Rompuys udkast til ny pagt. (www.sn.dk)
 SF are happy-PL for Rompuy's draft to new treaty
 'SF are happy about Rompuy's draft of new treaty.'
- b. SF er glad for Allerøds gode økonomi. (www.sn.dk)
 SF are happy-SG for Allerød's good economy
 'SF is happy about Allerød's good economy.'

I will compare the variable agreement patterns for *sin* and *deres* with Danish adjectival agreement which exhibits many of the same variation options. Interestingly, there is at least one case (specifically, the impersonal pronoun *man*) where the options for variable adjectival inflection and the option of choosing between *sin* or *deres* do not match.

There is a tendency in modern Danish to allow the impersonal pronoun *man* to take either singular or plural inflection on an associated predicative adjective. However, even in the context of a plural-inflected adjective, *man* cannot be the antecedent of a bound *deres*. Hansen and Heltoft (2011, 589) write the following:

Pronomenet *man* forbindes ofte med pluralisprædikat, men mht. refleksivpronomen er *man* entydigt singularis:

Spørgsmålet er om man er for dårlig-e til at holde styr på sin økonomi DR 17.3.2002 TVA

¹⁰the group sold their movie

The pronoun *man* is often seen with a plural predicate but it is unambiguously singular with respect to reflexive pronouns:

The question is whether MAN is too bad-PL at controlling REFL's finances

(126a) is completely impossible with coreference between *man* and *deres*, even with a plural-marked predicative adjective in between. The same sentence with *sin* instead of *deres* in (126b) is perfectly fine, and so is (126c) with a singular-inflected adjective and *sin*.

- (126) a. *Spørgsmålet er om man₁ er for dårlig-e til at holde styr på deres₁
 Question.DEF is if MAN is too bad.PL to to keep control on their
 økonomi.
 finances
 'The question is whether people are too bad at controlling their finances.'
- b. Spørgsmålet er om man₁ er for dårlig-e til at holde styr på sin₁
 Question.DEF is if MAN is too bad.PL to to keep control on REFL's
 økonomi.
 finances
 'The question is whether people are too bad at controlling their finances.'
- c. Spørgsmålet er om man₁ er for dårlig-Ø til at holde styr på sin₁
 Question.DEF is if MAN is too bad.C to to keep control on REFL's
 økonomi.
 finances
 'The question is whether someone is too bad at controlling their finances.'

Impersonal *man* allows either singular or plural inflection on its predicative adjectives as seen in (126b) and (126c). Where it differs from the collectives is that it is always ungrammatical as the antecedent of *deres* (e.g. (126a)), i.e. the bound possessive form with *man* as the antecedent must always be *sin* (or *ens* in those cases where *sin* is ruled out for structural reasons). *Man*, in other words, may bind *sin* and it may agree with predicative adjectives either marked with singular or plural agreement. It may not, however, bind *deres* which is overtly plural. In this sense the impersonal *man* is very much a mirror of the Danish *sin* which, in the standard language, seems to allow various antecedents that are not strictly plural, and not just those that are strictly coded as singular.

Table 2.3: The Danish quantifier inventory

Universal quantifiers	English counterpart
alle (al/alt)	all, everything, everyone
begge	both
enhver	any(one)
hver	every, everyone, each
hvem som helst	anyone
Indefinite quantifiers	
nogen/nogle	someone, some
en	one
Negative quantifiers	
ingen/intet	no one, nothing

Danish quantifiers and agreement variation

The Danish quantifiers also show variation in the reflexive and adjectival forms that they occur with. The Danish quantifier inventory (according to Hansen and Heltoft (2011, 550)) is given in (2.3).

Alle and *begge* primarily occur with plural-marked predicative adjectives in KorpusDK, or with uninflected adjectives such as *klar* or *gift* that are typically not inflected in predicative positions in Danish. Some examples can be found in the corpus where *alle* (but not, as far as I can see, *begge*) occurs with a singular-marked predicative adjective. A selection of these are reproduced in (127). These examples are all grammatically unacceptable (or at least very degraded) in my own native speaker grammar and they are very infrequent in the corpus compared to plural-marked predicative adjectives in similar utterances. *Al* and *alt* are, somewhat simplified, the singular counterparts of *alle* and they only take singular predicative adjectives in the corpus. Note that (127a) and (127b) are examples of what Christensen and Nyvad (2019), Christensen and Nyvad (2021), and Nyvad and Christensen (2023) describe as *alternativ kongruens* or *atypisk kongruens* (En. *alternative or atypical concord*) where they find that a nominal inside a prepositional object may interfere with the standard subject concord. Their group of papers on the topic explain the variation found mainly as a result of performance errors.

- (127) a. Og det tror jeg, vi alle er opmærksom-Ø på, at der er én god
 And that think I we all are attentive-SG on that there is one good

grund til... (KorpusDK, political speech)
reason to

'And I think that we are all attentive to the fact that there is one good reason for...'

- b. Det store nummer i påsken var Put and Take-søerne, hvor alle
The big number in easter.DEF was Put and Take-lakes.DEF where all
er sik-ker på gevinst. (KorpusDK, newspaper)
are certain-SG on prize.

'The big hit during Easter was the Put and Take lakes where everyone is guaranteed a catch.'

- c. Alle er velkom-men og der er gratis entré. (KorpusDK,
Everyone is welcome-SG and there is free admission.
local news)

'Everyone is welcome and admission is free.'

The plurality of *alle* and *begge* suggest that they should occur with *deres* as the preferred bound form and this is indeed the case. There are, however, a number of examples of both *begge* and *alle* in KorpusDK as the antecedents of *sin*.

Enhver primarily occurs with singular predicative adjectives in KorpusDK and primarily with *sin* as the bound reflexive form. I find a few examples of bound *deres* with *enhver* as the antecedent, reproduced in (128), but *sin* is by far the most frequent reflexive form with *enhver* as the antecedent.

- (128) a. ... hvor enhver havde deres egen private... (KorpusDK)
where anyone had their own private
'... where anyone had their own personal...'
- b. Enhver medbragte deres såkaldte "Luftschutzgepäck". (KorpusDK)
Anyone brought their so-called "Luftschutzgepäck"
'Everyone brought their so-called "Luftschutzgepäck".'

Hver is described as a distributive quantifier. Hansen and Heltoft (2011, 572) write that "[hver har] ENTYDIG DISTRIBUTIV FUNKTION", En. *hver* functions unambiguously distributively). In a simple corpus search for *hver* followed by a verb and either *sin* or *deres*¹¹ *hver* occurs most frequently as the antecedent of *sin*. The three examples in (129) are the only examples in the given search where *hver* is the antecedent of a

¹¹Search strings: [word="hver"] [pos="V"] [word="sin|sit|sine"] or [word="hver"] [pos="V"] [word="deres"]

bound *deres*. I find 28 examples of *hver* as the antecedent of *sin* in the same search environment.

- (129) a. Hver har deres fortrin og ulemper, muligheder og vilkår.
 each has their pros and cons options and conditions
 (KorpusDK)
 'Each have their pros and cons, options and conditions.'
- b. ... og hver har deres unikke kampenheder. (KorpusDK)
 and each has their unique fight-squads
 '... and each have their unique fighting squads.'
- c. Der er to døtre i familien, og hver har deres lille hund på
 There are two daughters in family.DEF and each has their little dog on
 hovedpuden om aftenen. (KorpusDK)
 pillow.DEF at night.DEF
 'There are two daughters in the family, and each has their little dog on
 the pillow at night.'

Hver is often used with a nominal complement (*hver NOM*, e.g. *hver forælder* (En. *each parent*)). *Sin* is still the most frequent bound form in this context but examples with *deres* also do occur in KorpusDK. A few examples with *hver N* antecedents that bind *sin* or *deres* are reproduced in (130)¹².

- (130) a. Hver aldersgruppe har deres egne til at sælge junk-en. (KorpusDK)
 Each age-group has their own to to sell junk.DEF
 'Each age group has their own that sell the junk.'
- b. Hver bruger får deres egen virtuelle personlighed og kan deltage
 Each user gets their own virtual personality and can participate
 i et socialt liv på EON. (KorpusDK)
 in a social life on EON
 'Each user gets their own virtual personality and can take part in a social
 life on EON.'
- c. Hver klasse hænger sine kritikpunkter op et sted på skole-n...
 Each class hangs REFL's criticism-points up a place on school-DEF
 (KorpusDK)

¹²The search environments that produce these examples are [word="hver"] [pos="N"] [pos="V"] [word="deres"] which turns up 5 relevant examples (and 10 that are not) and [word="hver"] [pos="N"] [pos="V"] [word="sin|sit|sine"] which turns up 191 examples all in all, out of which the majority are relevant, i.e. *hver N* followed by a bound *sin/sit/sine*

'Each class hangs up their points of criticism somewhere in the school.'

- d. ... at hver flok bruger sin egen "dialekt" af "spækhuggersprog-et".
that each pod uses REFL's own dialect of orca-language-DEF
(KorpusDK)

'... that each pod uses its own dialect of the orca language.'

A few different searches in KorpusDK only give examples of *hver N* with singular-inflected predicative adjectives. I do not, however, judge that plural-inflected predicative adjectives are impossible with *hver N*, as the acceptability of the constructed parallel example in (131) shows. Note that I have chosen to insert a *deres* in the constructed example since the utterance with the plural-inflected predicative adjective followed by *sin* feels much less acceptable. This is in line with the results found for English mismatched verbal inflection and reflexives in Sturt (2022).

- (131) a. ... at hver faggruppe er ansvarlig-Ø for sit / sine lokaler.
that each course-group is responsible.C for REFL's / REFL's rooms
(KorpusDK)

'... that each group is responsible for their own room or rooms.'

- b. ... at hver faggruppe er ansvarlig-e for deres lokaler.
that each course-group is responsible.PL for their rooms
(constructed)

'... that each group is responsible for their own room or rooms.'

A search for *hvem som helst* as the antecedent of either *sin* or *deres* provides very few results and out of the few that do, only one example is relevant, reproduced in (132)¹³. I gloss *hvem som helst* as *who so ever* in order to capture the meaning as closely as possible while still having three words in the gloss, even if the standard spelling of *whosoever* is as one orthographic unit.

- (132) a. Men klausuler gør, at ikke hvem som helst kan få opfyldt sine
But provisos do that not who so ever can get granted REFL's
ambitioner. (KorpusDK)
ambitions

¹³The specific searches are [word="hvem"][word="som"][word="helst"] []{0, 15}[word="sin|sit|sine"] and [word="hvem"][word="som"][word="helst"] []{0, 15}[word="deres"]

'But provisos have the effect that not just anyone can have their ambitions come true.'

Some very specific Google searches¹⁴ turn up a few relevant examples where both *sin* and *deres* may be bound by *hvem som helst*.

- (133) a. Hvem som helst kan sidde hjemme i deres stue og bidrage
Who so ever can sit home in their living-room and contribute
til Wikipedia. (ruc.dk, university website)
to Wikipedia
'Anyone can sit at home in their living room and contribute to Wikipedia.'
- b. ... at hvem som helst både kunne dele deres poesi og komme med
that who so ever both could share their poetry and come with
kommentarer til andres digte (information.dk, newspaper)
comments to others' poems
'... that anyone was able to both share their poetry and comment on
other people's poems.'
- c. ... at hvem som helst kunne købe en guitar og sige sin mening.
that who so ever could buy a guitar and say REFL's opinion
(jyllands-posten.dk, newspaper)
'... that anyone could buy a guitar and voice their opinion.'
- d. ... hvem som helst friktionsfrit kunne flytte sin kapital til hvor
who so ever frictionlessly could move REFL's capital to where
som helst... (information.dk, newspaper)
so ever
'... anyone could move their capital to anywhere without trouble...'

The indefinite quantifiers *nogen*, *noget*, and *nogle* can be used as the antecedent of both *sin* and *deres*. The form *nogle* is usually only found in written Danish. In spoken modern Danish both *nogen* and *nogle* are typically indistinguishable, according to Hansen and Heltoft (2011, 584). The examples in (134) are from KorpusDK and show both available options.

- (134) a. Nogle giver deres børn pligter (KorpusDK)
Some give their kids chores
'Some people give their kids chores'

¹⁴Such as site:dk "hvem som helst kunne * deres" and similar searches with various modals and either *sin*, *sit*, *sine*, or *deres*

- b. Nogle fik deres diagnose stillet for flere år siden (KorpusDK)
Some got their diagnosis made for several years ago
'Some people had their diagnosis made several years ago'
- c. ... at nogen siger deres mening (KorpusDK)
that someone says their opinion
'... that someone voices their opinion'
- d. Det er nu første gang, jeg har set nogen ta' deres bøger med,
It is now first time I have seen someone bring their books with
når de vil drukne sig. (KorpusDK)
when they will drown REFL
'It really is the first time that I have seen someone bring their books when
they want to drown themselves'
- e. ... når nogen bruger sin position over for andre. (KorpusDK)
when someone uses REFL's position over for others
'... when someone uses their position against other people.'
- f. Måske har nogen læst sine opgaver over på dig (KorpusDK)
Perhaps has someone loaded REFL's tasks over on you
'Perhaps someone has loaded their tasks onto you'
- g. ? Eller man ser hvorledes nogle ikke kan rumme egne grådige og
Or one sees how some not can hold own greedy and
destruktive følelser i sit indre (KorpusDK)
destructive emotions in REFL's inside
'Or you see how some people cannot cope with their own greedy and
destructive internal emotions'

Nogle, as the English translation and gloss make explicit, encodes plurality (translated as *some people*) while *nogen* (*someone*) rather encodes an unclear number of people (one or more). This could very well contribute to the observation that *nogle* does not participate as much in the *sin/deres* variation.

The indefinite quantifier *en* (*one*) is unambiguously singular and cannot be used as the antecedent of a bound *deres* but only *sin* (*sit, sine*), such as the example in (135).

- (135) a. Én viser sin rotte frem. (KorpusDK)
one shows REFL's rat off
'One shows off his/her rat.'

The negative quantifiers *ingen* and *intet* may occur as pronominals where they mean *no one* and *nothing*, respectively. They may also occur as negative determiners:

Ingen grund, intet hus, ingen sange (En. *no reason, no house, no songs*). The form *ingen* co-occurs with common gender singulars and all plural nouns and *intet* co-occurs with neuter gender singular nouns. I focus only on the pronominal form here because the number ambiguity of the pronominal is cancelled by any nominal complement of *ingen*.

Intet is fairly rare in spoken Danish where it is typically replaced with the form *ikke noget* (En. *not something*). *Intet* occurs very infrequently as the antecedent of either *sin/sit/sine* or *deres*. I found only one instance of *intet* as the binder of *sin* in KorpusDK with the wide search [word="intet"][pos="V"] []0, 20[word="sin|sit|sine"] within s;. I found no instances of *intet* as the binder of *deres* in KorpusDK and constructed examples are grammatically unacceptable with coreference between *intet* and *deres*. *Intet* perhaps encodes singular number more clearly than *ingen* does and does not allow the same variation in bound forms or associated predicates.

Ingen occurs in KorpusDK both as the antecedent of *sin* and of *deres*. More examples of *ingen* as the antecedent of *sin* (*sit/sine*) than *ingen* as the antecedent of *deres* turn up in the corpus in my simple searches¹⁵. This could be a substantial difference or just an effect of the fact that *sin* (*sit, sine*) occurs more frequently than *deres* in the corpus overall.

- (136) a. Ingen fik deres biler, men havde tabt mange tusinde kroner.
no-one got their cars but had lost many thousands kroner
(KorpusDK)
'No one got their cars but had lost several thousand kroner.'
- b. Ingen gav deres liv forgæves. (KorpusDK)
no-one gave their life in-vain
'No one gave their life in vain.'
- c. Ingen havde i deres vildeste fantasi forestillet sig dette
no-one had in their wildest imagination imagined REFL this
tilsyneladende massive angreb (KorpusDK)
apparently massive attack
'No one had imagined this apparently massive attack even in their wildest dreams.'
- (137) a. ...ingen tog sine smykker på. (KorpusDK)
no-one took REFL's jewelry.PL on

¹⁵[word="ingen"][pos="V"] []0, 3[word="sin|sit|sine"] within s; and
[word="ingen"][pos="V"] []0, 3[word="sin|sit|sine"] within s;

'No one put on their jewelry.'

- b. I de næste 12 uger må ingen ringe hjem til sin familie eller
in the next 12 weeks can no-one phone home to REFL's family or
kæreste. (KorpusDK)
partner
'In the next 12 weeks no one is allowed to phone their family or partner.'
- c. Ingen ville ofre sit fine lille sandslot for ideen om
no-one would sacrifice REFL's nice little sand-castle for idea.DEF about
proletariatets diktatur. (KorpusDK)
proletariat.DEF's dictatorship
'No one would sacrifice their nice little sand castle for the idea of the
dictatorship of the proletariat.'

Ingen can be found with singular as well as plural inflected predicative adjectives in KorpusDK and the singulars are slightly more frequent.

Hansen and Heltoft (2011, 580) write that the contrast between singular and plural is neutralized with *ingen* and *intet* and that this explains the many possibilities of variation in the forms that co-occur with *ingen* and *intet*. This does not prevent a language like English to encode *no one* as singular with respect to verbal agreement (*No one is happy that...*, however, so this apparent semantic number-neutrality is not necessarily something that also has number-neutral morphosyntactic consequences.

Table 2.4 sums up the options for agreement variation (semantic agreement in the terms of Corbett (2023) and previous work) within the categories discussed in the two previous sections. The categories that straightforwardly exhibit variation between singular and plural adjective agreement or the bound pronominal form are marked with Yes in the relevant cell. The categories that only allow one form are marked with No. The categories that primarily occur with one form but can also be found with the other are marked with (No). The preferred form (singular or plural adjective, bound *sin* or bound *deres*) is indicated in the right-most column in the table. *Nogen* could either be regarded as one form with a great deal of variation allowed (indicated in the row NOGEN) or as one morphological form with two different meanings. In order to show this option, I also split *nogen* into two separate categories: *nogen* meaning *someone* and *nogen* meaning *some people*. This makes explicit that *nogen* on the outside does permit a lot of variation but primarily because *nogen* in essence performs two functions, one (*someone*) which is primarily associated with singular adjective agreement and bound form *sin*, and another (*some people*) which is primarily associated

Table 2.4: Options for agreement variation with collectives, impersonal *man*, and certain quantifiers

Variation in:	Predicative adj. singular/plural agr.	Bound pronominal form sin/deres	Preferred form
Collectives	Yes	Yes	
Impersonal <i>man</i>	Yes	No	<i>sin</i>
QUANTIFIERS			
Alle	No	(No)	pl., <i>deres</i>
Begge	No	(No)	pl., <i>deres</i>
Enhver	(No)	(No)	sg., <i>sin</i>
Hver	(No)	(No)	sg., <i>sin</i>
Hvem som helst	?	?	sg., <i>sin</i>
NOGEN	Yes	Yes	
Nogen (= someone)	(No)	(No)	sg., <i>sin</i>
Nogen (= some people)	(No)	(No)	pl., <i>deres</i>
Nogle	No	(No)	pl., <i>deres</i>
Én	No	No	sg., <i>sin</i>
Ingen	Yes	Yes	sg., <i>sin</i>

with plural adjective agreement and bound form *deres*.

The point of the preceding discussion is to show that we find a great deal of variation in number agreement with predicative adjectives and bound possessive forms (*sin/deres*) with Danish collectives and quantifiers. The scope of investigation in this thesis is the number variation with *sin* that is outside of these contexts.

2.5.5 Distributivity

According to the Oxford Dictionary of English Grammar, something *distributive* is "(A word or phrase) that relates to individual members of a class separately, not jointly... **Distributive plural concord** is common in expressions such as *The children all had such eager faces* (where clearly each child had only one face), but a **distributive singular** is often possible, e.g. *They all had such an eager expression*" (Aarts 2014).

At least as early as Diderichsen (1939) we find the idea that the acceptability of *sin* with plural antecedents could be improved if the clause that contains *sin* can be read as distributive. This is the point in the quote from Diderichsen below; that *sin* with plural antecedents is the *least offensive* in a distributive context. Diderichsen's data does not support this observation.

Uafhængigt af disse Kilder har Dr. Aage Hansen meddelt mig den sikkert rigtige Iagttagelse, at Brugen af *sin* virker mindst stødende, naar man tænker paa hver enkelt af en Flerhed (naar Ordet *hver(t)* kunde indsættes foran *sine*), hvilket vel oftest er Tilfældet naar *sin* staar ved et Flertalsord.

Man vil dog formentlig ved Gennemlæsning af det anførte Citatmateriale, som jeg derfor ikke har ment at burde erstatte med Henvisninger, konstatere, at dette ikke særlig hyppigt er Tilfældet. (Diderichsen 1939, 68)

English: Independently of these sources, Dr. Aage Hansen has informed me of the likely correct observation that the use of sin seems the least offensive when one considers each part of a plurality (when the word hver(t) could be inserted in front of sine), which presumably is most often the case when sin is by a plural word.

After reading the given quotes, which I consequently have not replaced with references, one must concede that this is not very often the case.

Diderichsen's examples notwithstanding, distributivity could still in principle play a part in the occurrence of *sin* with plural antecedents. A reason that distributivity could be a factor in improving the acceptability of plural antecedent *sin* is that the distributive reading might introduce a one-to-one or one-to-many reading to a sentence with a plural subject, i.e. where the plural subject gets broken down into atomic entities that each individually have a relationship to the reflexive. This could in principle result in a reading that is conceptually closer to *sin* with a singular antecedent than a plural antecedent, improving the overall acceptability of the sentence even for speakers who do not normally accept plural antecedent *sin*. (138) is an example of precisely this from KorpusDK, where the plural subject *krokodiller* (En. *crocodiles*) most likely should be read as a group of individual crocodiles that each have their own jaw and their own prey.

- (138) a. ... hos **krokodiller**₁ der netop har smækket kæberne omkring **sit**₁
 in crocodiles who just have closed jaws.DEF around REFL's

forsvarsløse bytte...

defenseless prey

'... in crocodiles who have just closed their jaws around their defenseless prey...'

Another relevant term here is *dependent plurality*, which is a term first introduced by Mey (1981) as plural nominals used "in what would appear to be a singular meaning". The concept is discussed at least as early as in Chomsky (1975) but without the specific term. Chomsky notes that plural nominals in predicates may sometimes be more like their corresponding singular expressions, such as in the examples in (139).

- (139) a. Unicycles have wheels.
 b. The boys have living parents.

(139a) does not convey the obviously false statement that unicycles have more than one wheel. The sense of (139a) must be the tautological statement that all unicycles have precisely one wheel, in which case the plural *wheels* indeed seems to correspond more to a singular. The plural *living parents* in (139b) must correspond to a one-or-more sense, as the statement has to include the case where at least one boy has one living parent and at least one boy has more than one.

The project ScanDiaSyn (the data from which is available in the Nordic Syntax Database, Lindstad et al. 2009a) attempted to test the hypothesis that a distributive reading would improve the acceptability of a sentence with plural antecedent *sin* in Danish. The three sets of sentences in (140)-(142) were used to investigate this hypothesis and the acceptability of plural antecedent *sin* in general. The summary, examples, and analysis of this part of the ScanDiaSyn investigation are from Lundquist (2014). I have modified Lundquist's example sentences slightly to make them correct Danish (which I assume they would also have been by the Danish researchers who collected the data, or that the original sentences were accidentally slightly modified by Lundquist for the paper only).

- (140) a. **Forældre**₁ var kede af ikke at have tid nok til **deres**₁ barn.
 parents.DEF were sad of not to have time enough to their child
 'The parents were unhappy with not having time enough for their child.'
- b. **Forældre**₁ var kede af ikke at have tid nok til **sit**₁ barn.
 parents.DEF were sad of not to have time enough to REFL's child
 'The parents were unhappy with not having time enough for their child.'

- (141) a. **Forældre**₁ var kede af ikke at have tid nok til **deres**₁ børn.
 parents.DEF were sad of not to have time enough to their children
 'The parents were unhappy with not having time enough for their children.'
- b. **Forældre**₁ var kede af ikke at have tid nok til **sine**₁ børn.
 parents.DEF were sad of not to have time enough to REFL's children
 'The parents were unhappy with not having time enough for their children.'
- (142) a. **Spillerne**₁ snakkede om **deres**₁ præstationer.
 players.DEF talked about their performances
 'The players talked about their performances.'
- b. **Spillerne**₁ snakkede om **sine**₁ præstationer.
 players.DEF talked about REFL's performances
 'The players talked about their performances.'

(140) and (141) are intended to test the acceptability of plural antecedent *sin* compared to reflexive *deres*, and further to test whether it makes a difference that *sin* is contained within a plural nominal compared to a singular nominal. (142) is intended to test whether the possibility of a distributive reading (each player talks about their own individual performance) improves acceptability of plural antecedent *sin*.

The maps in figure 2.7 to figure 2.9 plot the ratings of the three plural antecedent *sin* sentences. All three maps show fairly low acceptability scores for the sentences overall: The black dots correspond to a low score on average (1-2 out of 5) and the grey dots correspond to a medium score on average (3 out of 5). There are no areas with high scores on average, but it is interesting to see that all three plural antecedent *sin* sentences actually do receive high scores (4 or 5) from some informants. The results of the study do indeed show a higher acceptability of (141b) and (142b) which both have *sin* contained in a plural nominal than (140b) where *sin* is contained in a singular nominal. This lines up well with the corpus results discussed in section 4.8.2 of this thesis (starting on page 198) which show that plural antecedent *sin* is more often contained within plural nominals than *sin* in general is.

Lundquist (2014, 530) further suggests that distributivity is a relevant factor in improving the acceptability of plural antecedent *sin* but it is not obvious from this data that this is the case. (141b) and (142b) seem to receive much the same ratings even though (142b) is potentially more obviously distributive than (141b). At the same time, I would argue that (140b) can in principle be read as distributive, too,

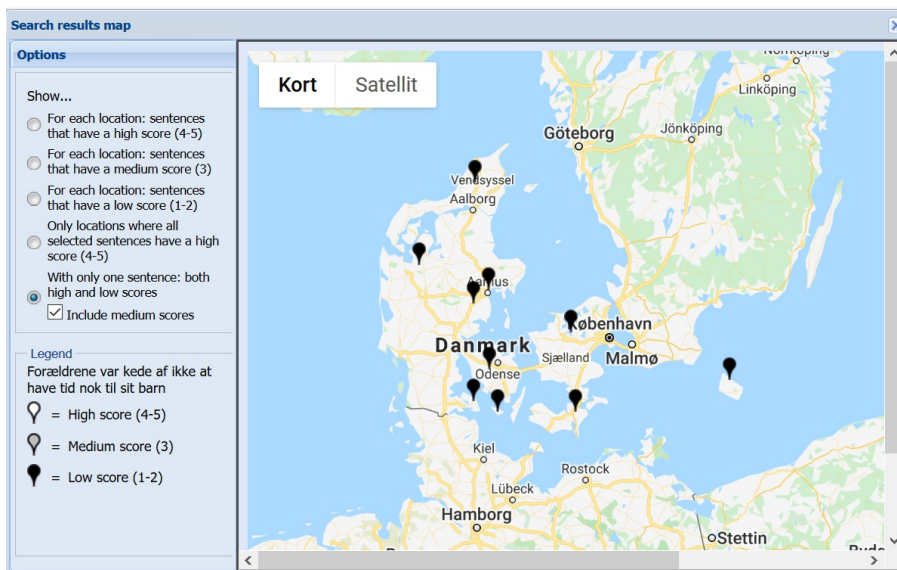


Figure 2.7: Forældrene var kede af ikke at have tid nok til sit barn

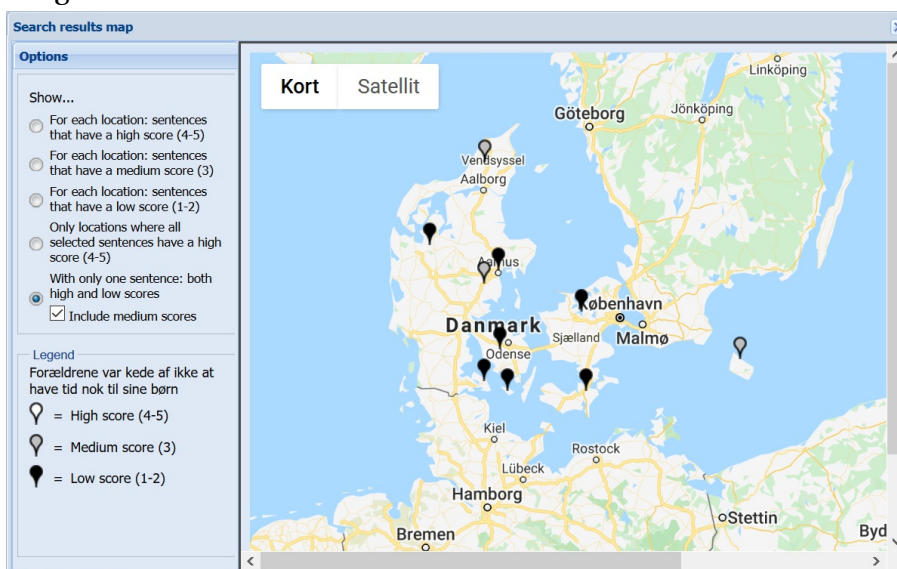


Figure 2.8: Forældrene var kede af ikke at have tid nok til sine børn

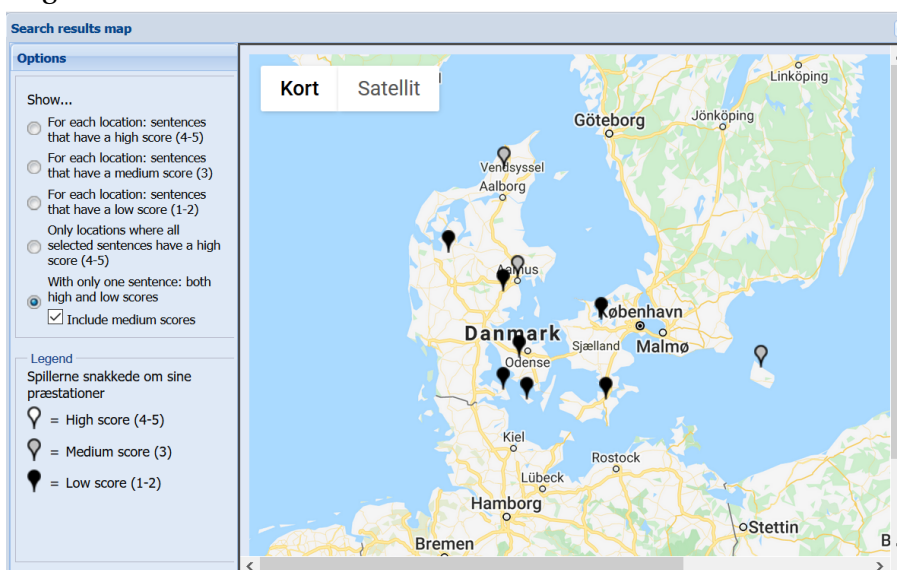


Figure 2.9: Spillerne snakkede om sine præstationer

against Lundquist's claim that a sentence such as (140b) with a singular possessed noun phrase cannot be read as distributive.

- (143) a. **Forældre**₁ var kede af ikke at have tid nok til **sit**₁ barn.
 parents.DEF were sad of not to have time enough to REFL's child
 'The parents were unhappy with not having time enough for their child.'

NON-DISTRIBUTIVE READING: One set of parents do not, collectively, have enough time for their one child.

DISTRIBUTIVE READING: Several parents, each with one child, do not individually have enough time for their one child.

In this case the only factor that seems to explain the difference in acceptability between (141b) and (142b) on the one hand and (140b) on the other hand, is that the two former have *sin* contained within a plural nominal while *sin* is in a singular nominal in the latter. It is in no way unthinkable that distributivity *is* a relevant factor in improving plural antecedent *sin* but the data from ScanDiaSyn unfortunately does not illuminate the issue very much. I further investigate whether distributivity can be said to be a decisive factor for the occurrence of *sin* with plural antecedents in chapter 4.

Chapter 3

Reflexive pronouns with plural antecedents from runes to LANCHART

3.1 Introduction

An earlier, Danish version of this chapter has been published in the journal *Danske Studier* as Ehlers (2020).

The examples in (144) and (145) illustrate a difference between *sig* and *sin* that is found in modern Danish but less so in previous stages of the language. *Sig selv* and *sig* allow both singular and plural antecedents, as shown in the pairs in (144)a,b and (144)c,d.

- (144) a. **Anne**₁ fotograferer **sig selv**₁.
Anne photographs REFL self
'Anne photographs herself.'
- b. **Anne**₁ skammer **sig**₁.
Anne shames REFL
'Anne is ashamed.'
- c. **Vennerne**₁ fotograferer **sig selv**₁.
Friends.DEF photographs REFL self
'The friends photograph themselves.'
- d. **Vennerne**₁ skammer **sig**₁.
Friends.DEF shame REFL
'The friends are ashamed.'

As discussed at length in chapter 2, *sin* may be coreferent with singular antecedents, (145a), but rarely occurs with plural antecedents, as indicated by the ungrammatical-

ity of (145b). The non-reflexive plural possessive *deres* is used instead of *sin* when the antecedent is plural, (145c).

- (145) a. **Anne**₁ fotograferer **sin**₁ morgenmad.
Anne photographs REFL's breakfast
'Anne photographs her breakfast.'
- b. ***Vennerne**₁ fotograferer **sin**₁ morgenmad.
Friends.DEF photograph REFL's breakfast
'The friends photograph their breakfast.'
- c. **Vennerne**₁ fotograferer **deres**₁ morgenmad.
Friends.DEF photograph their breakfast
'The friends photograph their breakfast.'

In modern standard Danish, non-reflexive *dem* cannot have a local antecedent, as opposed to the possessive *deres*. This is the case both when *dem selv* would substitute a *sig selv* as in (146a) and when *dem* would substitute a non-theta local *sig* as in (146b).

- (146) a. ***Vennerne**₁ fotograferer **dem selv**₁.
Friends.DEF photograph them self
Intended: 'The friends photograph themselves.'
- b. ***Vennerne**₁ køber **dem**₁ en bold.
Friends.DEF buy them a ball
Intended: 'The friends buy a ball.'

Both sentences are perfectly grammatical without the intended coreference, in which case the *them* who is being photographed or bought a ball is a third party, not the friends themselves.

This difference between number-neutral *sig (selv)* and number-sensitive *sin* in standard Danish is not found in the closely related Swedish and Norwegian where *sig (selv)* as well as *sin* may be coreferent with antecedents of any number¹. (Swedish examples in (147) from Hinchliffe and Holmes (2018, 54) and Norwegian examples in (148) from Holmes and Enger (2018, 163).) Note that even though both examples with *sin* with plural antecedents have *sin* in a plural nominal, this is not a requirement for having a plural antecedent. (147)b and (148)b would have been just as grammatical with *sin* in a singular nominal.

¹This is not to say that the equivalents of locally bound *dem* or *deres* never occurs in the two languages. Jan Terje Faarlund (p.c.) has noted that the Norwegian equivalent of *deres* as a locally bound reflexive can sometimes be found in Norwegian, too.

(147) Swedish

- a. **Han**₁ lämnar **sin**₁ fru.
He leaves REFL's wife
'He leaves his wife.'
- b. **Deras barn**₁ saknar **sina**₁ föräldra.
Their children miss REFL's parents
'Their children miss their parents.'

(148) Norwegian

- a. **Kari**₁ liker sønnen **sin**₁.
Kari likes son.DEF REFL's
'Kari likes her son.'
- b. **Gry og Ola**₁ klemte sønnene **sine**₁.
Gry and Ola hugged sons.DEF REFL's
'Gry and Ola hugged their sons.'

This difference between *sig* and *sin* in terms of possible antecedents is a fairly modern difference. In earlier stages of Danish, both *sig* and *sin* posed number-restrictions on their antecedents to a greater extent. *Dem* and *deres* were used as locally bound pronouns when the antecedents were plural and *sig* and *sin* were used when the local antecedents were singular. In e.g. the first complete Danish Bible translation, Christian III's Bible from 1550, there are multiple examples of a locally bound *dem*, such as the example from Genesis 3:8 in (149). In the modern Danish translation of this passage, (150), Adam and Eve hide *sig* (*REFL*) rather than hide *dem* (En. *them*).

(149) Chr. III's Bible (1550): Genesis 3:8

- a. Da skiulte **Adam oc hans hustru**₁ **dem**₁ for Guds HERRENS ansigt
Then hid Adam and his wife them for God's LORD.DEF's face

'Then Adam and his wife hid themselves from the Lord God's face.'

(150) Newest Danish Bible Translation (1992): Genesis 3:8

- a. Da gemte **Adam og hans kvinde**₁ **sig**₁ for Gud Herren mellem
Then hid Adam and his woman REFL for God Lord.DEF between
havens træer.
garden.DEF's trees
'Then Adam and his wife hid themselves from the Lord God's between
the trees of the garden.'

In this chapter, I describe and discuss the development of *sig*, *sin*, *dem*, and *deres* in Danish from the language of the runic stones to present day Danish. The first part of the chapter, section 3.2, concerns the use of reflexives on runic stones and in the early Danish area laws. The data show that *dem* and *deres* have been occurring locally bound with plural antecedents as early as the 13th century. In section 3.3 I survey the use of reflexives in a selection of texts from 1500 to 1900. In particular, I investigate the variation in reflexive usage in Christian III's Bible from 1550, which turns out to be different from what has been claimed in the literature. In section 3.4 I use speech data from LANCHART, a corpus of spoken modern-day Danish, to investigate whether locally bound *dem* can be said to still be in use.

In the new major work on the history of the Danish language (*Dansk sproghistorie* (2016)), the main time periods are classified as in the following, and I adopt this classification. The period from 800 to 1100 is called *Runic Danish* (Da. 'olddansk' or 'runedansk') in accordance with the mode of the written language of the time. Around 1100, Danish manuscripts written in Latin script start appearing. This is marked as a change to the Middle Danish period (Da. 'gammeldansk' or 'middeldansk') from 1100 to 1500. The Middle Danish period is divided into older Middle Danish, 1100-1350, and younger Middle Danish, 1350-1500. After 1500, the handwritten Danish of the previous period starts being supplemented by printed texts and this development marks the transition into older (1500-1750) and younger (1750-present day) modern Danish.

3.2 The reflexives in Runic Danish and Middle Danish

Counterparts of *sig* and *sin* both existed as reflexive pronouns without number restrictions in Proto-Germanic and Old Norse (see e.g. Mikkelsen (1911, 266)). That is, both *sig* and *sin* could be bound by singular as well as plural antecedents. In *Grammatik over det danske sprog* (henceforth GDS, Hansen and Heltoft (2011, 590)), the authors give examples of *sin* with plural antecedents on a runic stone from the 9th century, (151a), in the oldest Danish Bible translation from the 15th century, (152a), and in Malling's *Store og gode Handlinger af Danske, Norske og Holstenere* from 1777, (153a). The authors also note that *deres* has been used reflexively as early as the Scanian area laws from the 13th century.

(151) 9th century

- a. **Alla** **syniR**₁ gærþu kumbl þøsi æft faþur **sin**₁
 Alle.GEN sons made runes these after father REFL's
 'Alle's sons made the runes after their father'

(152) 15th century

- a. **the**₁ skulle plawes for **syn**₁ fathers, og **synæ**₁ syndhe
 they should be-plagued for REFL's father's and REFL's sins
 'They were to be tormented because of their own and their father's sins.'

(153) 1777

- a. [han] klappede paa Kaarden: og spurgte: om **Officererne**₁ havde
 he patted on rapier.DEF and asked if officers.DEF had
 baaret **sin**₁ Væрге saa kiekt i Fyen...
 carried REFL's weapon as boldly in Funen
 'He patted the rapier and asked if the officers had carried their weapons
 as boldly in Funen...'

It is likely not the case that *sin* and *deres* (and *sig* and *dem*, which are not mentioned by GDS in this regard) have been more or less interchangeable through the history of Danish. In some of the earliest post-runic Danish manuscripts, both *sig* and *sin* have been restricted in their usage in terms of possible antecedents. In most of the Danish area laws from the 13th century, *sig* and *sin* are used primarily with singular antecedents and *dem* and *deres* with plural antecedents. In the following sections, I chart the development in use of *sig* and *sin* from runic Danish (section 3.2.1) to Middle Danish (section 3.2.2).

3.2.1 Runic Danish (800-1100)

The University of Copenhagen and the National Museum of Denmark have published a searchable database of runic inscriptions² from all the inscribed items that have been found in the area that comprised Denmark in the middle ages, including Schleswig and Scania. The database contains approximately 900 runic inscriptions, ranging from the Iron Age to the middle ages. This means that the youngest inscriptions are from the same period as the oldest written Danish manuscripts from the 13th century. However, all of the inscriptions that contain a plural antecedent *sin* are from before 1125 and out of these, all of the late inscriptions (from between 1025 and 1125)

²accessed at <http://runer.ku.dk/Search.aspx> in 2019, as *Danske Runeindskrifter - runedatabase* in 2023. Also available as part of a larger runic database at <https://app.raa.se/open/runor/search> in 2023.

are from Bornholm. The older inscriptions are from between 900 and 1020 and come from various sites in the Danish areas, including Scania.

I have searched for *deres*, *sin*, *dem*, and *sig* in the translated modern Danish text in the database and manually examined all of the results. A possible issue with this search strategy is that details from the original inscriptions may be lost in the transition from runic original to modern Danish, perhaps as a consequence of faulty translation. This seems to be the case with the translation of the text on the Kuregård stone in (154) where *sin* in the transliteration³ becomes *sinn* in the transcription and again *sin* in the modern Danish translation. In the translation, *sin* should have been translated into *deres* in accordance with the modern norm.

(154) Kuregård-stenen, Bornholm, c1025-1075

a. **suin : auk : ketil₁** : reistu : sten : eftiR : (i)u(l)k|ir : faþur : **sin₁**

Svénn ok Kætill réstu stén æftiR Í[g]ulgéR, faþur sinn

Sven og Ketil rejste stenen efter Igulger, sin fader.

'Sveinn and Ketill raised the stone in memory of Ígulgeirr, their father.'

Similar translation errors should not pose a problem, however, as my search includes both *sin* and *deres*, in which case the faulty translation shows up as a *sin* and is then still included. To make completely certain, I have also searched for *sin* in the transliterated text and *sinn* in the transcription and I have only found the one translation error in (154).

There are 30 items in the database that contain the word *deres* in the modern Danish translation. Some of the items contain more than one *deres* so that the database contains 32 *deres* in all. Out of these 32, 27 *deres* represent translations of possessive reflexives (*sin*, *sit*, *sine*) bound by plural antecedents and 5 are non-reflexive *deres*. Altogether, I find 28 examples of possessive reflexives with plural antecedents: The 27 found in the *deres* search and the one example in (154). I provide examples of reflexively used *deres* in (155) and non-reflexively used *deres* in (156). Both examples and the translation into modern Danish are taken directly from the database.

(155) Aarhus, 970-1020. Reflexive *sin* with plural antecedent.

³All the examples from the runic stones are structured as shown below:

Line 1: TRANSLITERATION: Text written with runes interpreted and converted into Latin letters

Line 2: TRANSCRIPTION: The inscription transcribed with normalized spelling, intended to represent sounds of the actual language of the time

Line 3: MODERN DANISH TRANSLATION

- a. Side A : kunulfr : auk : augutr : auk : aslakR : auk : rulfr : risθu Side B : stin : θansi : eftiR : ful : fela(k)a : sin : | : iaR : uarθ (:) ..y- : tuθr : Side C θa : kunukaR : | barθusk :

Gunnulfr ok Øgotr/Øðgotr ok ÁslākR ok Rōlfr₁ rēsθu stēn θannsi æftiR **Fūl, fēlaga sinn₁**, eR varð ... døðr, θā kunungaR barðusk.

Gunulv og Øgot og Aslak og Rolf rejste denne sten efter deres fælle Ful. Han fandt døden ... da konger kæmpede.

Gunnulfr and Eygautr/Auðgautr and Áslakr and Hrólfr raised this stone in memory of Fūl, their partner, who died when kings fought.

(156) Bornholm, c1075-1125. Non-reflexive *deres*.

- a. Krist hælgi hialpi siolu **θeRa** brøθra bæggia
Den hellige Krist hjælpe deres, begge brødres, sjæle.
May holy Christ help the souls of both these brothers.

In (155) *deres* (in the translation) is used as a reflexive: It is bound by the subject of the minimal clause that contains *deres*, *Gunulv og Øgot og Aslak og Rolf*. In the original inscription, the pronoun is reflexive possessive *sin*. In (156) *deres* is used non-reflexively: It is not bound within the minimal clause. *Deres* corefers with the nominal *brøθra bæggia* (En. *both brothers*), which functions as an apposition to *deres* in the transliterated text, but *deres* is not bound by *brøθra bæggia*. The translated *deres* is **θeRa** in the transliteration, the non-reflexive possessive pronoun. This is the case for all of the 32 translated *deres* in the database. When *deres* in the translation is used reflexively, it is a translation of *sin*. When *deres* in the translation is used non-reflexively, it is a translation of the non-reflexive possessive pronoun **θeRa**. The runic inscriptions consistently represent a grammar where *sin* is used as a reflexive pronoun with singular and plural antecedents and *deres* is only used in non-reflexive contexts.

It is not possible to make the same observation from the available data with regard to the use of *sig* and *dem*. However, this is probably a question of lack of data. *Sig* shows up in the translated text of 34 items. Out of these 34, 30 have been excluded for various reasons. Either the text on the item has not been interpreted (and the *sig* comes from the name of the item, e.g. the *Sigerslev-brakteat*), or the *sig* appears as part of another word on the item on those that have been interpreted (such as in the name *Sigtryg* or the verb *velsigne* (En. *bless*)). On the four items where *sig* does appear as a pronoun, *sig* is bound by singular antecedents. *Dem* appears on the translation of three items in the database. Two of these are originally in Latin and consequently

not relevant. On the third item, *dem* is used non-reflexively. Due to scarcity of data, it is not possible to conclude anything about whether *sig* allowed plural antecedents and whether *dem* could be bound locally. According to Brøndum-Nielsen (1965, 25), the reflexive pronouns allowed antecedents of any number and person prior to the Common Germanic stage but were limited to third person antecedents, still of any number, in Germanic (as well as e.g. Greek and Latin). It is possible that this use of *sig* continued into the grammars of the time of the runic inscriptions, even if the available runic data does not provide any evidence for (or against) it. In the language of the Danish provincial laws from the 13th century, which is the topic of section 3.2.2, *sig* is used with plural antecedents in the Scanian manuscripts but not (or very little) in the West Danish (the parts of Denmark that are west of Scania) manuscripts. Presumably, then, *sig* would have retained the Common Germanic option of being bound by antecedents of all numbers at least in the Scanian area through the time of the runic inscriptions.

3.2.2 Middle Danish (1100-1500)

The following sections build primarily on data from Diderichsen (1939). It is a comprehensive study of the use of reflexives in Danish over time but it also not very accessible. This section is an attempt at making some of Diderichsen's data more easily available.

Diderichsen (1939) contains, among many other topics, an investigation of the use of pronouns with plural antecedents in some of the 13th century provincial laws: *Skånske Lov* (En. *Law of Scania*), *Valdemars Sjællandske Lov* (En. *Valdemar's Zealandic Law*), *Eriks Sjællandske Lov* (En. *Erik's Zealandic Law*) and *Jyske Lov* (En. *Code of Jutland* or *The Provincial Law of Jutland*). The investigation is a response to the claims from Falk and Torp (1900) that are cited below:

I slutningen af 15. aarh. blev «dem» det fremherskende, maaske tildels under indflydelse af det adjektiviske reflexiv [sin], som efter mønster af det tyske «ihr» lød «deres». [...]

Navnlig blev fra reformationstiden af ved indflydelse af tysk («ihr») pronomenet «deres» almindeligt, hvor der henvises til et flertalsord; saaledes stadig hos Chr. Pedersen. Denne brug er trængt helt igjennem i dansk. (Falk and Torp (1900, 131-4))

(En. translation: In the end of the 15. century *dem* became the predominantly used form, perhaps partly due to influence from the adjectival reflexive [sin], which had acquired the form *deres* in accordance with the German *ihr*. [...])

From the time of the Reformation in particular, the pronoun *deres* became the common form due to influence from German (*ihr*) when the pronoun refers to a word in the plural; this is still the case in Chr. Pedersen's texts. This usage is currently the only option in Danish.)

Falk and Torp, in short, claim that *dem* and *deres* became the forms most commonly used for reflexive binding by plural antecedents from the end of the 15th century and onwards. This means that *dem* and *deres* would have supplanted *sig* and *sin* with plural antecedents but not with singular antecedents. I illustrate this allegedly new usage in the examples in (157) from Chr. III's Bible from 1550 (Genesis 6:3 and 8:5).

(157) Chr. III's Bible (1550): Genesis 6:3

- a. **Menniskene**₁ ville icke mere lade **dem**₁ straffe aff min Aand
humans.DEF would not more let them punish by my spirit
'The people would no longer let themselves be punished by my spirit.'

Chr. III's Bible (1550): Genesis 8:5

- b. Paa den første dag i den tiende Maaned lode **toppene** aff
on the first day in the tenth month let tops.DEF of
Biergene₁ **dem**₁ til siune.
mountains.DEF them to sight
'On the first day of the tenth month the mountain tops became visible.'

In the Runic Danish language described in section 3.2.1, dated prior to 1100, *sig* and *sin* were used with singular and plural antecedents and *dem* and *deres* were used as non-reflexive pronouns. According to Falk and Torp, a shift in usage happened around the end of the 15th century which restricted the set of antecedents allowed by *sig* and *sin* to singulars while *dem* and *deres* were drafted into use as reflexives with plural antecedents, perhaps due to influence from German.

As Diderichsen (1939, 71) points out, the claims about dating and causality from Falk and Torp merit scrutiny. I will lead with a discussion of the claims about German influence from page 132 and continue with a discussion of the dating of the change from page 133.

German influence?

Falk and Torp suggest that *dem* came to be used reflexively because speakers generalised from the reflexive use of *deres*. Due to German influence, speakers started using *deres* (rather than *sin*) with plural antecedents. This usage would then have pushed *dem* in the same direction, leading to a situation where *sig* was restricted to singular antecedents and speakers used *dem* as a reflexive pronoun with local plural antecedents. The logic behind the claim about German influence must be that the German *sein*, cognate of the Danish *sin*, is a singular pronoun. The German *sein* cannot refer to plural antecedents or feminine singular antecedents where the pronoun *ihr* is used instead for both forms. *Sein* is not a reflexive pronoun but rather a possessive personal pronoun of masculine and neuter gender. It can be used reflexively as in (158) and non-reflexively as in (159).

(158) German, bound *sein*-

- a. **Er**₁ **liebt** **seine**₁ Mutter.
he loves his mother
'He loves his mother.'

(159) German, unbound *sein*-

- a. Seine Mutter ist sehr nett.
his/its mother is very nice
'His/its mother is very nice.'
- b. Seine Mutter ist auch ein Hund.
his/its mother is also a dog
'His/its mother is also a dog.'

If *deres* came to be used reflexively due to German influence, we might expect a system not just with number differences but also gender differences. Specifically, a reflexive system where *sin* is limited to masculine and neuter antecedents and speakers use *deres* (En. *their*) with reference to plural antecedents and *hendes* (En. *her*) with reference to feminine antecedents which is how the German system works. This is actually what seems to be the case in the those areas of Jutland where the dialects were heavily influenced by German, as they were spoken around 18-1900 in some parts of Schleswig (Jul Nielsen (1986, 51, 75–79)), including Fjolde (Bjerrum and Bjerrum (1974, 23–24)). Noesgaard studied dialectal variation in the written language of school children in the middle of the 20th century and he describes that some of the students use *sin* as though it were the German *sein*:

Nogle elever, især med tysk skolegang, anvender *sin* hvor tyskerne bruger *sein* [...]. Den tyske sætning: 'Er nahm *seinen* Hut' bliver altså til den danske: 'Han tog *sin* hat', medens det tyske: 'Sie nahm *ihren* Hut', bliver til: 'Hun tog hendes hat'.

(Noesgaard 1951, 74)

(En. translation: Some students, particularly those with German schooling, use *sin* where the Germans use *sein*... The German sentence: 'Er nahm *seinen* Hut' consequently becomes the Danish: 'Han tog *sin* hat' (En. gloss 'He took REFL's hat') while the German: 'Sie nahm *ihren* Hut' becomes: 'Hun tog hendes hat' (En. gloss 'She took her hat').)

It would be interesting to see whether this kind of German influence can be found in older texts, e.g. from around the time of the Reformation, but that investigation is outside the scope of this chapter.

The story of German influence on *sin* and *deres* is not syntactically impossible, i.e. it is possible that influence from the use of the German *sein* could lead Danish speakers to adopt a reflexive system with a number-sensitive *sin* and a locally bound *deres* (although we then may also expect *hendes* to be common as a locally bound reflexive, which is not the case). However, the same path of German influence is less likely for *sig* and *dem* where German *sich*, cognate of Danish *sig*, allows both singular and plural antecedents. The expectation here must be that strong German influence would support the preservation of a *sig* that allows both singular and plural antecedents. This prediction is not immediately borne out in the language of the dialects mentioned above as heavily influenced by German: In the language of Fjølde, *dem* was used as a reflexive with plural antecedents and *sig* as a reflexive with singular antecedents (Bjerrum and Bjerrum (1974, 24)) and Jul Nielsen (Jul Nielsen (1986, 44)) just notes that *dem* is often used as a reflexive in the Jutlandic dialects but also in other dialects and in the older stages of the language.

Dating the change: *dem/deres* was already predominant by 1200

In this section, I address the dating of the change that led to a period where speakers primarily used *sig* and *sin* with singular antecedents. Falk and Torp (Falk and Torp (1900, 131-4)) suggest that the change happened around 1500 but the data from Diderichsen (1939) indicate that this dating is likely too late. Example (160) is from Law of Scania from around 1200⁴. It contains a *dem* and a *deres* that are both used as

⁴The example is from Law of Scania, chapter 48 in the version which is available at tekstnet.dk. I use the translation from Kroman and Juul (1945) as inspiration for translating the original text for the

reflexives.

- (160) a. at þe₁ muhu æi fōþæ þem utæn þe₁ sæli þeræ₁ iorþ
that they can not feed them without they sell their land
'... that they cannot support themselves without selling their land...'

Falk and Torp (1900, 131) note that both *dem* and *sig* were used with plural antecedents in the old provincial laws from the 13th century so they do recognize that *dem* could be used as a reflexive before 1500. Their claim is specifically that *dem* became the most frequent form from around 1500. Diderichsen (1939, 72–4) tests this claim empirically and in the illustration in fig. 3.1 I reproduce his data graphically. The data for the graph comes from Diderichsen's counts of the four relevant pronouns, *dem*, *sig*, *sin*, and *deres*, in four of the old provincial laws. Diderichsen communicates his results partly by providing the counts and partly by providing a list of where the pronouns appear in the relevant manuscripts. I have counted the items in the lists where Diderichsen does not himself provide counts. There are cases where the pronoun choice varies between manuscripts, which Diderichsen notes in his counts. The frequencies in fig. 3.1 represent the most frequent forms used across the variant manuscripts.

The graph in fig. 3.1 is split into four in accordance with the four law texts, *Law of Scania*, *Valdemar's Law of Zealand*, *Erik's Law of Zealand*, and the *Law of Jutland*. For each law, the columns represent the occurrences of respectively *dem*, *sig*, *deres*, and *sin* used as reflexive pronouns bound by plural antecedents.

In Law of Scania the data support Falk and Torp's description of the use of *dem* and *sig* at the time. Both *dem* and *sig* are used as reflexives with plural antecedents, but *sig* (18 instances) is much more frequent than *dem* (2 instances). The use of *deres* and *sin* is rather more varied: Both forms are used and the frequency of occurrence is approximately the same (12 *deres*, 11 *sin*).

In the laws from Zealand (Erik's and Valdemar's) and Jutland the usage pattern is quite different both from Law of Scania and from Falk and Torp's descriptions. The non-reflexive *dem* and *deres* are almost the only used forms in reflexive contexts with plural antecedents. There are almost no occurrences of *sig* and *sin* with plural antecedents. There are seven occurrences of both *dem* and *deres* in Valdemar's Law but none of *sig* and only one of *sin*. In Erik's Law Diderichsen finds five occurrences of *dem* and 21 occurrences of *deres*. There are no occurrences of *sig* and only one of *sin*. In Law of Jutland he finds eight *dem* and 21 *deres* but zero instances of *sig* and *sin*.

gloss. A newer translation is Tamm and Vogt (2016).

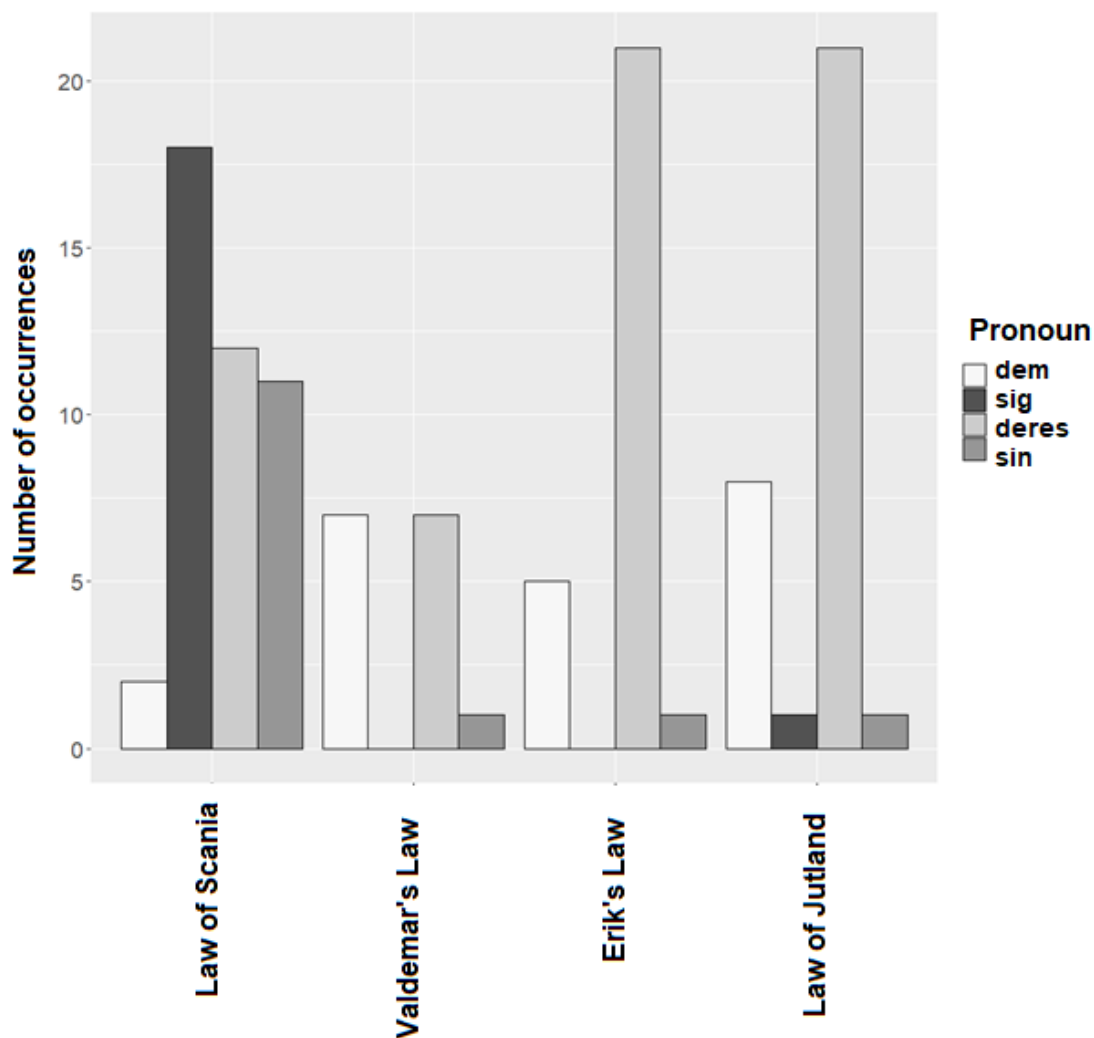


Figure 3.1: Occurrences of reflexive uses of *dem*, *sig*, *deres*, and *sin* in four of the medieval Danish provincial laws. The data for the plot is extracted from the counts in Diderichsen (1939). Law of Scania (Skånske Lov) is from the beginning of the 13th century. Valdemar's Law (Valdemars Sjællandske Lov) is from the first half of the 13th century. Erik's Law (Eriks Sjællandske Lov) is from the middle of the 13th century. Law of Jutland is from 1241 according to the book's own prologue.

Brøndum-Nielsen (1965, 34) nuances Diderichsen's counts as he finds that *Valdemar's Law* does contain some instances of *sig* (*sik*, *sich*) as verbal complements. One example is (161) from chapter 16 of *Valdemar's Law*. Diderichsen evidently did not find in his excerpt, which necessarily is a risk when working with only parts of a text. This also further underscores the fact that the data presented in fig. 3.1 is only an approximation of the actual usage patterns in the manuscripts.

- (161) a. tha ær e **børnæbørn**₁ nærmær at witæ **sich**₁ til mæth witnæ...
 then are the grandchildren closer to prove REFL to with witnesses
 'then the grandchildren are closer to proving (to their own benefit) with
 witnesses...'

Brøndum-Nielsen further notes that *dem*, and never *sig*, occurs as the reflexive form as the complement of prepositions in the text. This also seems to be a truth with modifications, however, as I find at least one instance of *sig* as a prepositional complement in chapter 16 of *Valdemar's Law* (*oc hauær fathærn fleræ børn hemæ æftær sich*, En. *and has the father more children at home after REFL*). I also, however, find this distributional difference between verbal and prepositional reflexives in Chr. III's Bible, as further discussed in section 3.3.1. In keeping with the counts in Diderichsen (1939), *Erik's Law* and *Law of Jutland* both have *dem* as the plural antecedent reflexive form both as verbal and prepositional complement (Brøndum-Nielsen (1965, 35)).

To sum up, only the language of *Law of Scania* fits Falk and Torp's description of the reflexive use of pronouns with plural antecedents when they write that both *dem* and *sig* were used in the old provincial laws (Falk and Torp (1900, 131)). In the three other laws the pronoun usage strongly suggests that the number restriction on *sig* and *sin* was in place in the language of the time, meaning that *dem* and *deres* were already the most frequent forms with reference to plural antecedents by the 13th century and onwards. It does not seem to be the case, as Falk and Torp claim, that there was a major shift in reflexive usage around 1500: That shift had already taken place at least three centuries earlier.

The use of *sin* with singular antecedents in the provincial laws

Following on Diderichsen's study on the use of pronouns with plural antecedents in the provincial laws, it is worth including a study on the use of pronouns with singular antecedents in the same texts. In the language on the runic stones, at least *sin* is used exclusively over *deres* as the reflexive form with both singular and plural antecedents.

In the language of the West Danish (i.e. excluding Scania) provincial laws, *sig* and *sin* almost do not occur with plural antecedents. This pattern could in principle be caused by two different strands of development: Either *sig* and *sin* have become restricted to singular antecedents, or else *sig* and *sin* have gone out of use more or less completely. In the first case, *sig* and *sin* should still occur with singular antecedents even if *dem* and *deres* are used in their stead when the antecedent is plural. In the second case, *sig* and *sin* should be rare both with singular and plural antecedents. The data that will be presented in this section supports the first hypothesis, i.e. that *sin* (and presumably *sig*, which I have not found nor made a similar study of) has been restricted to singular antecedents.

The illustration in fig. 3.2 presents data from Wellejus (1972, 130-7). The primary purpose of Wellejus' study is to investigate which language features that can be used to distinguish between Middle Danish manuscripts from Scania, Zealand, and Jutland. The use of possessive non-reflexive pronouns in place of *sin* is often presented as a Jutlandic language feature. For this reason, Wellejus includes reflexive use of *hans* (En. *his*), *hendes* (En. *her*), and *sin* with singular antecedents in the study. It turns out, perhaps surprisingly, that manuscripts from Jutland as well as Zealand contain reflexive use of non-reflexive *hans* and *hendes* but that *sin* nevertheless is the most frequent form in reflexive contexts.

I show Wellejus' data in a simplified graphical format, just like I did with Diderichsen's data. This necessarily gives a decrease in the number of details that can be presented: Wellejus' study contains much more information about the occurrence of the pronouns in the various different manuscripts of the same text than what I am able to present in a graph. In addition, Wellejus surveys many more texts than I have chosen to include in the graph, but the tendency is the same across all the texts in her study: *Sin* is by far the most frequent form and *hans/hendes* occurs from time to time as reflexives in manuscripts from both Jutland and Zealand. Wellejus includes several Zealandic laws and the one that I have chosen to include in the graph in fig. 3.2 is the one with the highest proportion of *hans/hendes* relative to *sin*.

Figure 3.2 shows the occurrences of *sin* compared with reflexively used *hans/hendes* with singular antecedents in three provincial laws, *Law of Scania*, *Erik's Law*, and *Law of Jutland*. The three columns show the proportion of *sin* relative to *hans/hendes* in the three laws. When Law of Scania has a proportion of 100 % *sin*, it means that *sin* is the only possessive form used reflexively with singular antecedents in Law of Scania.

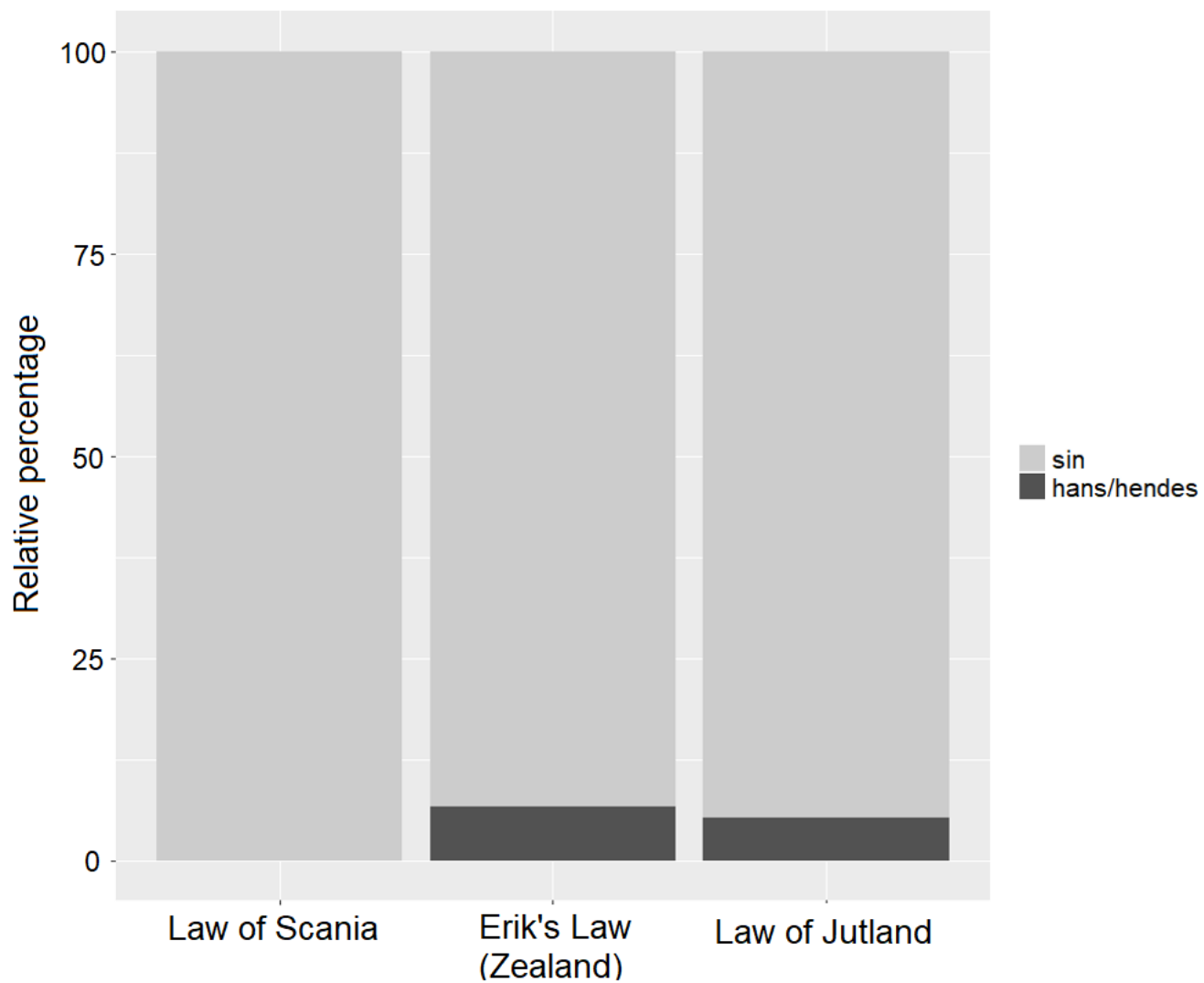


Figure 3.2: Reflexively used pronouns (*sin*, *hendes*, *hans*) with singular antecedents in three medieval provincial laws: Law of Scania, Erik's Law, Law of Jutland. Data from Wellejus (1972, 130-7).

The usage pattern in Law of Scania clearly differs from those of *Erik's Law* and *Law of Jutland*. The Scanian text contains no examples of reflexively used *hans* or *hendes*. Wellejus' investigation into the variant manuscripts finds just a single instance of a reflexively used *hans/hendes* across all her surveyed manuscripts. These findings fit in well with Diderichsen's data where the pronoun use in Law of Scania also appears considerably different from the West Danish laws.

In Erik's Law and Law of Jutland *sin* is still by far the most frequent form in reflexive contexts with singular antecedents. Erik's Law contains 283 instances of *sin* with singular antecedents and 20 instances of reflexively used *hans/hendes*. Law of Jutland contains 231 instances of *sin* with singular antecedents and 13 instances of reflexively used *hans/hendes*. The variant manuscripts of both laws contain some instances of *hans/hendes* where the primary manuscript has *sin*. I conclude, with Wellejus, that there does not seem to be much of a difference between the Jutlandic and the Zealandic manuscripts with respect to the use of *sin* and *hans/hendes*. Mikkelsen (1911, 266) remarks that the use of *hans/hendes* as a reflexive did figure in both the Jutlandic and Zealandic provincial laws but that the use was primarily retained in the later Jutlandic dialects.

In the specific context of this chapter, I conclude that it is primarily in regards to reference to plural antecedents that the reflexive system changed between Runic Danish and Middle Danish. With reference to singular antecedents, *sin* (and presumably also *sig*, even though I do not have data to substantiate this assumption) is the primarily used reflexive form throughout both time periods. Additionally, Wellejus' data shows that variation between *sin* and *hans/hendes* in reflexive contexts cannot be said to be a new feature of Danish, just like variation between *sin* and *deres* (and *sig* and *dem*) also seems to have been present in the language for at least 800 years.

3.3 The reflexives in Older Modern Danish

In section 3.2.2 I used data from Diderichsen (1939) and Wellejus (1972) to substantiate the claim that *dem* and *deres* were used as locally bound reflexives with plural antecedents as far back as the 13th century, and that *sin* and *sig* have been used as reflexives with singular antecedents throughout the same period. The relevant difference between Runic Danish and the later stages of Danish in this respect is that *sin* is used much less with plural antecedents than in Runic Danish. Speakers of modern stan-

dard Danish still almost exclusively use *sin* for binding by singular antecedents and *deres* for binding by plural antecedents, preserving the pattern found in the provincial laws of Middle Danish. The use of *sig* and *dem*, on the other hand, has changed quite a bit. *Sig* is used as a reflexive with both singular and plural antecedents and *dem* is only used as a reflexive in a very limited number of contexts where *sig* is also possible, e.g. in non-local binding contexts. The examples in (162) show that *sin* and *deres* have a clear division of labour in modern Danish where *sin* is used with singular antecedents and *deres* with plural antecedents.

- (162) a. Ja, og **hun**₁ spiser **sin**₁ mad, indskød pigen.
yes and she eats REFL's food interjected girl.DEF
'Yes, and she eats her food, the girl interjected.' (KorpusDK)
- b. * Det var, mens **de**₁ sad og spiste **sin**₁ mad i middagsstunden...
it was while they sat and ate REFL's food in middaytime.DEF
'It was while they were eating their food around noon...' (constructed)
- c. Det var, mens **de**₁ sad og spiste **deres**₁ mad i middagsstunden...
it was while they sat and ate their food in middaytime.DEF
'It was while they were eating their food around noon...' (KorpusDK)

(163a) and (163b) show that *sig* is used as a reflexive with singular and plural antecedents, and (163c) shows that *sig* not only can but must be used in local binding contexts with a plural antecedent. (163d) is an example of *dem* as a non-local reflexive in one of the few contexts where *dem* and *sig* may be used interchangeably, as in the parallel example in (163e).

- (163) a. **Han**₁ skammer **sig**₁ ikke over at vise sin svaghed.
He shames REFL not over to show REFL's weakness
'He is not ashamed of showing weakness.' (KorpusDK)
- b. Adam og hans kvinde var nøgne, men **de**₁ skammede **sig**₁ ikke.
Adam and his woman were naked but they shamed REFL not
'Adam and his woman were naked but they were not ashamed.' (KorpusDK)
- c. * Adam og hans kvinde var nøgne, men **de**₁ skammede **dem**₁ ikke.
Adam and his woman were naked but they shamed them not
'Adam and his woman were naked but they were not ashamed.' (constructed)

- d. **De**₁ var klædt som forretningsmænd og bad mig køre **dem**₁ til
 they were dressed as businessmen and asked me drive them to
 South Bronx.
 South Bronx
 'They were dressed as business men and asked me to drive them to South
 Bronx.' (KorpusDK)
- e. **De**₁ lagde sig i sandet og lod solen tørre **sig**₁.
 they laid REFL in sand.DEF and let sun.DEF dry REFL
 'They lay down in the sand and let the sun dry them.' (KorpusDK)

The question is when (and why) *sig* regained the option of being bound by plural antecedents and *dem* correspondingly had its usage limited. Another question is whether a similar development has taken place (or is taking place) with *sin* and *deres*. In an attempt to answer these questions I have surveyed a number of texts from the older modern Danish period (1500-1750). I only use sources that are available in digital, searchable formats which necessarily limits the amount of available materials and authors. The advantage of this, on the other hand, is that it allows me to go through large amounts of complete texts, rather than looking at smaller excerpts (the conventional method of *excerpering*).

3.3.1 *sig/dem* in Chr. III's Bible (1550)

The diachronic development of *sig* and *dem* has most recently been treated in Pedersen (2017). On the basis of Diderichsen (1939) and her own text surveys Pedersen describes a fairly abrupt shift in the use of *sig* and *dem* in the written language after the Reformation. Before the Reformation, as e.g. Diderichsen's data from the provincial laws in fig. 3.1 on page 135 show (discounting Scania), *dem* was by far the most frequent reflexive form in binding contexts with plural antecedents. After the Reformation, *sig* seems to have gained ground quickly at the expense of reflexive *dem*. Pedersen (2017, 5-6) suggests (in line with Diderichsen (1939, 83) and with what I discussed on page 132) that this change happened due to German influence, as the German *sig*-cognate *sich* allows both singular and plural antecedents. That this shift happens only after the Reformation – given that German language influence would have been ubiquitous before the Reformation, too – Pedersen explains through an interaction between an increased contact with German in academic contexts and the introduction of the printing press.

As a supplement to Pedersen (2017), I have carried out a small investigation into the use of *sig* and *dem* in Chr. III's Bible from 1550. According to both Diderichsen (Diderichsen (1939)) and Skautrup (Skautrup (1947)), this text diverges quite a lot from the new use of *sig* and *dem* that became the written standard after the Reformation. Diderichsen describes that Chr. III's Bibel "[står] paa det middelalderlige Standpunkt, idet der i al Fald i flere af dens Bøger[...] er tydelig Overvægt af *dem* henvisende til Flertalssubjekt" (Diderichsen (1939, 80)) ('takes the Medieval approach, as there at least in some of its books is a clear majority of *dem* with reference to plural antecedents'). Skautrup repeats this assessment and claims that the practice of using *dem* as a reflexive form with plural antecedents is also used "overvejende i 1550" ('predominantly in 1550', i.e. in the 1550 Bible) (Skautrup (1947, 201)). Pedersen (2017, 6) wonders why this use in Chr. III's Bible did not become the norm in the written language and gives as a possible explanation that there is variation between *sig* and *dem* in the Bible texts text, which has then not been able to give a unified usage impression that could have an impact on the standard language.

The data in this section nuance the descriptions from Diderichsen, Skautrup, and Pedersen, since it does not actually seem to be the case that reflexive *dem* has a clear majority in the 1550 Bible. From the Old Testament, I have read through Genesis, Exodus, First Kings, and the Book of Job. From the New Testament, I have read through Matthew, Mark, and the Book of Revelation. I use the digital, searchable edition of Chr. III's Bible that was available on tekstnet.dk, *Tekster fra Danmarks middelalder og renæssance 1100-1550* in 2019-20 (and unfortunately is not available at the time of writing in 2023). I have searched for *sig* and *dem* (which are the orthographic variants used in the specific text) and manually read through all of the examples in order to separate the reflexive examples from the non-reflexives. The results of the survey are presented in table 3.1.

Genesis and Exodus do indeed exhibit a clear majority of *dem*, and it might not seem too unreasonable to conclude from this that the rest of the text follows the same pattern. This hypothesis must, however, be rejected on the basis of the rest of the data. The usage pattern is rather different in the New Testament and in the two other surveyed works from the Old Testament: In Matthew, First Kings, and Book of Job *sig* and *dem* appear with approximately the same frequency; In Mark and Book of Revelation *sig* is the most frequent form.

There is an interesting difference in the way that *sig* and *dem* are used in the Old

Table 3.1: *Sig* and *dem* in Chr. III's Bible (1550)

	<i>sig</i> with plural antecedent	<i>dem</i> with plural antecedent
The Old Testament		
Genesis	4	24
Exodus	5	27
First Kings	9	13
Book of Job	21	14
The New Testament		
Matthew	14	22
Mark	41	2
Book of Revelation	8	2

Testament compared to the New Testament, also apart from the obvious difference in frequency. In just a single case in Exodus *sig* is used as the complement of a preposition, (164a). All the other *sig* examples from the Old Testament are complements of verbs, such as (164b) and (164c). In all the four surveyed books from the Old Testament *dem* is used both as the complement of a preposition, (165a), and as the complement of a verb, (165b). The most frequent context for *dem* is as the complement of a verb.

- (164) a. Oc **de**₁ ana~~m~~mede til **sig**₁ aff Mose al den Løfftning...
and they took to REFL of Moses all the offerings
'And they took to themselves from Moses all the offerings...' (Exodus, 36:3)
- b. Oc **fulene**₁ formere **sig**₁ paa iorden
and birds.DEF multiply REFL on earth.DEF
'And the birds multiply on the earth.' (Genesis, 1:22)
- c. MEn der **Menniskene**₁ begynte at formeris paa iorden / och
but there humans.DEF began to multiply on earth.DEF and
afflede **sig**₁ døtter
bred REFL daughters
'But people began to multiply there on the Earth and bred themselves daughters.' (Genesis, 6:1)
- (165) a. Jeg haffuer giffuit eder alle Vrter / ath **de**₁ skulle giffue sæd aff **dem**₁
I have given you all herbs / that they should give seed off them
offuer all iorden /
over all earth.DEF

'I have given you every herb so that they should give seeds from them all over the earth.' (Genesis, 1:29)

- b. Mig tœcte at **Solen oc Maanen oc elleffue Stierner**₁
me appeared that sun.DEF and moon.DEF and eleven stars
bøyde **dem**₁ ned for mig.
bowed them down for me
'It seemed to me that the sun and the moon and eleven stars bowed down for me.' (Genesis, 37:9)

The distribution is different in the books from the New Testament that I have looked through. In all three books, *sig* functions as the complement of a preposition in 20-30 % of all occurrences of *sig* and otherwise as the complement of a verb. (166) provides examples of both complement types.

- (166) a. De daarlige toge deris Lamper / Oc **de**₁ toge icke Olie met **sig**₁.
The bad took their lamps / and they took not oil with REFL
'The bad ones brought their lamps and they did not bring oil with them.'
(Matthew, 25:3)
- b. Men Maria Magdalena vaar der / oc den anden Maria / **de**₁ sette
but Mary Magdalene was there / and the other Mary / de sat
sig₁ tuert offuer fra Graffuen
REFL across over from grave.DEF
'But Mary Magdalene was there as well as the other Mary, they sat themselves opposite the grave.' (Matthew, 27:61)
- c. At **de**₁ skulde inted bære met **sig**₁ paa veyen /
that they should nothing carry with REFL on road.DEF
'That they should carry nothing with them on the road.' (Mark, 6:8)
- d. Thi der skulle **mange falske Christi**₁ reyse **sig**₁ op /
for there should many false Christs rise REFL up
'For many false Christs should rise there.' (Mark, 13:22)
- e. Oc **de**₁ haffde en Konge offuer **sig**₁ / en Engel aff affgrunden /
and they had a king over REFL / an engel of abyss.DEF
'And they had a king over them, an angel of the abyss.' (Revelation, 9:11)
- f. Oc ieg saa Stole / oc **de**₁ sette **sig**₁ der paa / oc dem bleff
and I saw chairs / and they sat REFL there on / and them was
Domen giffuen /
judgment.DEF given
'And I saw chairs and they sat upon them and they were given their judgment.' (Revelation, 20:4)

Dem is still used both as the complement of verbs and the complement of prepositions in the New Testament, as the examples in (167) show. It is worth noting that the use of *dem* in (167e) and (167f) (from the Old Testament, The Book of Job) would also be acceptable in Modern Danish. In both examples the reflexive *dem* in the embedded non-finite embedded clause is non-locally bound by a subject antecedent in a matrix clause. This is precisely one of the contexts where *dem* and *sig* are interchangeable in modern Danish (see section 2.3 on page 69). The newest modern Danish authorized Bible translation (*Bibelen : autoriseret af Hendes Majestæt Dronning Margrethe II* (1992)) has *sig* instead of *dem* in (167f) but still *dem* in (167e).

- (167) a. At **de**₁ skulle icke træde dem met deris Føder / Oc vende **dem**₁
 that they should not tread them with their feet / and turn them
 om / oc riffue eder sønder.
 around / and rip you asunder
 'That they should not step on them with their feet and turn around and
 rip you asunder.' (Matthew, 7:6)
- b. Da tenckte **de**₁ ved **dem selff**₁ / oc sagde /
 then thought they by them self / and said
 'Then they thought to themselves and said' (Matthew, 16:7)
- c. At **de**₁ skulde inted bære met sig paa veyen / Vden alene
 that they should nothing carry with REFL on road.DEF / without only
 en Staff / ey Taske / ey Brød / ey Pendinge i beltet / Men skulde
 a staff / not bag / not bread / not money in belt.DEF / but should
 haffue sko paa / oc icke føre **dem**₁ i tho Kiortle.
 have shoes on / and not dress them in two kirtles
 'That they should carry nothing with them on the road, apart only from
 a staff, not bag, not bread, not money in the belt, but they should wear
 shoes and not dress themselves in two kirtles.' (Mark, 6:8)
- d. Oc **de**₁ forundrede **dem**₁ paa hans lerdøm /
 and they wondered them on his learning
 'And they were amazed at his teaching.' (Mark, 1:22)
- e. Oc **de**₁ hørde en stor røst aff Himmelen sige til **dem**₁ / Stiger hid
 and they heard a large voice of heaven say to them / rise here
 op.
 up
 'And they heard a large voice from heaven say to them: Come up here'
 (Revelation, 11:12)

Table 3.2: *Sig* and *dem* with plural antecedents as PP and VP complements in Chr. III's Bible (1550)

	<i>sig</i>		<i>dem</i>	
	V comp	P comp	V comp	P comp
The Old Testament				
Genesis	4	0	16	8
Exodus	4	1	25	2
First Kings	9	0	11	2
Book of Job	21	0	13	1
The New Testament				
Matthew	9	5	18	4
Mark	31	10	2	0
Book of Revelation	6	2	1	1

- f. oc **de**₁ sende hen / oc bade deris tre søster til at æde oc
and they sent there / and asked their three sisters to to eat and
dricke met **dem**₁.
drink with them
'and they sent [an invitation] and asked their three sisters to eat and
drink with them.' (Job, 1:4)

I summarize the frequencies of *sig* and *dem* as complements of verbs (V comp) and complements of prepositions (P comp) in table 3.2. The counts are given for each of the four surveyed books in the Old Testament and each of the three surveyed books in the New Testament.

The most notable difference is found in the usage domain of *sig* in the two Testaments. In the Old Testament *sig* appears almost exclusively as the complement of verbs, both in Genesis and Exodus where *sig* is much less frequent than *dem* as the reflexive form with plural antecedents and in First Kings and Book of Job where *sig* and *dem* appear more or less equally often. The only example of *sig* as a prepositional complement is found in Exodus where *sig* appears as the complement of a verb of German origin (annamme, *Ordbog over det Danske Sprog*), where a German original text might have prompted the use of *sig* (cognate of German *sich*, as discussed on page 132). *Dem* is used most frequently as a verbal complement but also does occur a number of times as a prepositional complement. In the three surveyed books from the New Testament *sig* is by far the most frequent plural antecedent reflexive form, at the expense of *dem*. The use of *sig* is also qualitatively different from the Old Testament, as *sig* in the New Testament appears both as verbal and prepositional

complement. *Dem* continues to be used as both verbal and preposition complement, although overall at a rather lower frequency than in the books in the Old Testament.

It is difficult to provide a definite explanation of the qualitative difference in reflexive use between the two Testaments, but my suggestion is that the Old Testament reflects an older stage of the language or a transition stage, which is both evident in a lower frequency of *sig* and in the fact that *sig* is limited to the domain of verbal complements. The New Testament correspondingly reflects a later stage of the development of the language where *sig* has also moved into the domain of prepositional complements and where *dem* is being pushed out of the (written) language. It may be the case that the editors of the individual books from the 1550 Bible had adopted the new reflexive use of *sig* with both singular and plural antecedents to various degrees, and that they then introduced this usage (or preserved the old one) to the extent that they themselves had adopted plural antecedent *sig*. Or alternatively, to the extent that they themselves were influenced by a new written standard, whether or not they would have had plural antecedent *sig* in their own spoken language. Brøndum-Nielsen suggests that, in dialects and sources where both *sig* and *dem* are used as plural antecedent reflexives, *sig* is used in the more fixed prepositional expressions with reflexives, such as "*tage paa sig* el. *ved sig* 'trives godt'" (Brøndum-Nielsen (1965, 36)). Conversely, *dem* is used in the less fixed expressions, as the complement of verbs and prepositions that also take non-reflexive complements. If this is the case, as it seems to be at least in the Old Testament in the 1550 Bible, *dem* must be considered the standard, the most productive form, while *sig* is lexically constrained.

Plural antecedent, locally bound reflexive *dem* entirely loses the option of appearing as a verbal complement in the later standard Danish language, cf. the unacceptability of a sentence like *Børnene₁ hygger dem₁* (En. *The kids are having a good time*). In the domain of prepositional complements I believe that there still is a degree of optionality in the modern standard language between *sig* and *dem*, as seems to be the case in the parallel examples in (168). As mentioned previously in this section, optionality is also found in non-local binding contexts. *Dem* with plural antecedents is not completely gone from the standard language, but it has been severely restrained in its usage domains.

- (168) a. Så svømmede de ind mod land, mens **de₁** trak jollen
 then swam they in towards land while they dragged dinghy.DEF
 efter **dem₁**.
 after them

'Then they swam towards land while they dragged the dinghy after them.'

(KorpusDK)

- b. Scrooge kom i tanke om at have hørt, at **spøgelses**₁ i
Scrooge came in thought about to have heard that ghosts in
hjem søgte huse ofte trak tunge lænker efter **sig**₁.
haunted houses often dragged heavy chains after REFL
'Scrooge recalled hearing that ghosts in haunted houses often dragged
heavy chains after them.' (KorpusDK)

3.3.2 *Sin* and *deres* from Chr. Pedersen to H.C. Andersen (1500-1900)

This section explores the use of *sin* with plural antecedents in various texts throughout the older and younger modern Danish period. Diderichsen has a large collection of examples of plural antecedent *sin* from a range of authors and texts, and he writes that plural antecedent *sin* for the most part only appears "lejlighedsvis (som Regel sikkert kun ved Lapsus...)" (En. *occasionally (usually only by mistake)*) (Diderichsen 1939, 68-69). His collection of examples shows that *sin* has indeed been used with reference to plural antecedents throughout his surveyed time period, but it does not say anything about the frequency of use. I have carried out a small study on the use of *sin* and *deres* in a selection of sources from older modern Danish and younger modern Danish in order to complement Diderichsen's collection of examples and to attempt to quantify the extent to which *sin* (and by extension, locally bound *deres*) has been used with plural antecedents in the period from 1500 to 1900.

I have looked through Chr. Pedersen's *En nøttelig Legebog* from 1533 and found two examples of plural antecedent *sin* and around 40 examples of plural antecedent *deres*. The two examples with *sin* are reproduced in (169) and in both cases they seem more like errors than evidence of plural antecedent *sin* in the language of Chr. Pedersen.

- (169) a. Ma skall oc giffue de **siwge**₁ Saffra i **sin**₁ mad thi det gør at
One shall also give the sick saffron in REFL's food for it makes to
soffue
sleep
'One should give the sick saffron in their food as it helps them sleep'
- b. De som haffue denne brøst, **de**₁ skulle gerne æde Agerhøne
those who have this illness they should preferably eat partridge

æggh, Eller oc støde dem raa, och smørge sig met dem paa **sin**₁
 eggs, or also break them raw and smear REFL with them on REFL's
 lønlige ting
 secret thing
 'Those who have this illness, they should preferably eat partridge eggs
 or crack them raw and smear themselves with them on their secret thing
 (i.e genitals)'

The *sin* in (169a) does not refer to a subject but to the object *de siwge* (*the sick*), which in itself could be considered non-standard usage apart from the fact that the antecedent is plural (but cf. e.g. Lødrup (2008) who argues that the Scandinavian reflexives are not necessarily subject-oriented). The *sin* in (169b) is so far away from its subject that it is possible that the writer lost his sense of the plural of the subject in the process of writing. (169b) also contains a *sig* that is bound by the same antecedent as *sin*. It could be the case that this preceding reflexive form inspired the use of a reflexive form in the possessive nominal.

In Poul Helgesen's translation of Erasmus of Rotterdam (translated in the 1520s and printed in 1534) I have read through the first third of the text and found a single example of *sin* bound by a plural antecedent, (170). There are many more examples of plural antecedent reflexive *deres* in the same excerpt.

(170) a. **bønder**₁ bruge icke thenom wden till **sitt**₁ eget gaffn
 farmers use not them without to REFL's own benefit
 'Farmers do not use them apart from for their own benefit.' (Helgesen's translation of Erasmus)

In Hans Tausen's *Postil*, which I have read excerpts from (Vinterdelen, and The gospel for Christmas from 1539), there are no examples of plural antecedent *sin* and around seven examples of plural antecedent reflexive *deres*. In *Kirkeordinansen* (1539), probably primarily translated by Peder Palladius but with contributions from other editors and authors, there is a great deal of variation in the reflexive use. I find at least 16 examples of plural antecedent *sin* (including the inflectional varieties *sit* and *sine* and the orthographic variants *syn*, *syt(t)* and *syne*) and approximately 40 *deris*. A few of the plural antecedent *sin* examples are given in (171). Note that the *sin* in 171c is not c-commanded by its antecedent and further that the same example is included in the article for *tromol* in *Ordbog over det danske Sprog* (Tro-maal, *Ordbog over det Danske Sprog*) but with a *deris* (*deres*) in the place of *sin*.

- (171) a. hoss **dennom**₁ som ere dømde til døde oc skulle affliffues for
with those who are sentenced to death and should kill.PASS for
sine₁ misgierninger/
REFL's misdeeds
'with those who are sentences to death and will be put to death for their
misdeeds' (*Kirkeordinansen*)
- b. at **de**₁ med rett alworlighed bekiende **sin**₁ tro oc begere saa
that they with proper seriousness confess REFL's faith and desire the
sacramentet
sacrament
'that they confess their faith and desire the sacrament with proper seri-
ousness' (*Kirkeordinansen*)
- c. Des ligeste skulle oc **scholemesterne**₁ (...) giffue forældrene i
as well should also schoolmasters.DEF give parents.DEF in
rett tromol tilkiende/ huilcke aff **syn**₁ studeren intet siunes
proper confidence to-know which of REFL's students nothing seems
at bliffue forbedrede/
to become improved
'similarly, the school masters should inform the parents about which of
their students that do not appear to have any improvements (*Kirkeordi-
nansen*)

Diderichsen reports that he read through the first 50 pages of Palladius' *Sct. Peders Skib* where he finds no examples of plural antecedent *sin* and around 21 *deres* (Diderichsen 1939). Given the use of *sin* and *deres* in *Kirkeordinansen*, I would expect to see more examples of plural antecedent *sin* in Palladius' other texts. That this does not seem to be the case, at least in the one sermon that Diderichsen has read, suggests that Palladius did not in fact have plural antecedent *sin* in his grammar. The plural antecedent *sin* that do appear in *Kirkeordinansen* could be the result of influence from the grammars of other editors or translators.

I find almost no variation in the use of *sin* and *deres* with plural antecedents in Genesis and Matthew in Chr. III's Bible, contrary to the great deal of variation in the use of *sig* and *dem* described in section 3.3.1. I only investigated Genesis and Matthew since the usage patterns for *sin* and *deres* seem much less varied across the texts than the usage patterns do for *sig* and *dem*. There is a single sentence with two plural antecedent *sin* in Genesis and 44 instances of *deres* in the same chapter. I find a single example of plural antecedent *sin* in Matthew and 37 *deres*. Both examples are given

in (172).

- (172) a. Och Gud skabte diur paa Jorden huert effter sin art
 and God created animals on Earth.textscdef each after REFL's kind
 oc **fæ**₁ effter **sin**₁ art och **alle haande orme paa**
 and livestock after REFL's kind and all sorts worms on
Jorden₂ effter **sin**₂ art.
 Earth.textscdef after REFL's kind
 'And God made animals on the Earth, each according to their own kind,
 and livestock according to their own kind and all sorts of worms on the
 Earth after their own kind.' (Genesis 1:25)
- b. Der hand saa nu **Phariseer oc Saduceer**₁ komme til
 there he saw now many Pharisees and Saducees come
sin₁ Daab sagde hand til dem
 to REFL's baptism said he to them
 'When he saw many Pharisees and Saducees coming to their baptism, he
 said to them' (Matthew 3:7)

The referent of *sin* in (172b) is ambiguous and it is not completely clear from either context or from the sentence itself which reading that should be preferred. In one reading, the *Pharisees and Saducees* are the antecedent and they come to be baptized. In this reading, *sin* has a plural antecedent. In another reading, *hand (he, John the Baptist)* is the antecedent, the binding is non-local and it is John's ceremony of baptism that the Pharisees and Saducees are coming to attend. In this reading, *sin* has a singular antecedent and the example is then not relevant here. Interestingly, different translations have chosen different readings: The authorized Danish translation seems to lean towards the plural antecedent *sin* reading ("Men da han så, at mange af farisæerne og saddukæerne kom for at blive døbt af ham...", En. *But when he saw that many of the Pharisees and Sadducees came to be baptized by him...*) while the New International Version (an English translation) seems to learn towards the singular antecedent reading (But when he saw many of the Pharisees and Sadducees coming to where he was baptizing,...)⁵.

In Anders Sørensen Vedel's *Predicken over Kansler Friis* (1571) I find no examples of plural antecedent *sin* in a search of the first 48 pages and *deres* is the only used plural antecedent reflexive form. In Niels Hemmingsen's *Om Ecteskab* (1572) I have searched

⁵I found the English line on biblegateway.com where there are a number of English Bible translations available. Both interpretations are found across several of the other English translations.

for *sin* and *sit* in the first half of the book and found no examples of plural antecedent *sin*. I have read short excerpts to ascertain that *deres* is indeed used instead.

Leonora Christina (1621-1698) exclusively uses *deres* or *derres* – and not *sin* – with reference to plural antecedents in the excerpt from *Jammers Minde* that I have searched through (page 94-184 of the version of the text available at adl.dk).

I have read through every example of *sin* and *deres* in the texts by hymn writer Thomas Kingo (1634-1703) available on renæssancesprog.dk.⁶ Kingo's texts contain a remarkably large number of plural antecedent *sin*. In almost all of the texts, *sin* appears at least as frequently as *deres* in plural antecedent reflexive contexts. It seems reasonable to imagine that rhyme and rhythm may have played a part in the choice of reflexive form and I provide a selection of representative examples in (173).

- (173) a. **Alle**₁ hâr **sit**₁, Stort eller Lit!
all have REFL's, large or little
'Everyone has their (thing), large or small.' (Aandelig Siunge-Koor (1681))
- b. Deyligste Roser hâr stindeste Toorne, **Skjønneste Blomster**₁ **sin**₁
loveliest roses have pointiest thorns, loveliest flowes REFL's
tærende Gift
corroding poison
'The loveliest roses have the pointiest thorns, the most beautiful flowers have their corroding poison' (Aandelig Siunge-Koor (1681))
- c. Huor **deris**₁ Hænder **de**₁ til nogen Gierning sætter,
where their hands they to some action set
'They they set their hands to some action' (Kroneborgs Korte Beskrivelse)
- d. Gud trøste **dem**₁ der ere nu i **deris**₁ Døds minut
God soothe those who are now in their death's minute
'May God soothe those who are in the minute of their death' (Siunge-Koor (1674))
- e. Og **de**₁ med **deres**₁ Blood **sin**₁ troskab vidne maa.
and they with their blood REFL's loyalty witness may
'And they must give witness to their loyalty with their blood' (Hosianna)

In (173a), *sit* presumably rhymes with *Lit* and a *deres* instead of a *sin* would break the rhythm in (173b). I find no examples of *sine*, so it seems that *deres* is used anywhere

⁶Specifically Aandelig Siunge-Koor (1681), Christian Vs første Ledings-Tog (1676), De Fattige udj Odensee Hospital (1682), Hosianna (1671), Kroneborgs Korte Beskrivelse (1672), Mølle-borups Velkom (1675), Paa-Skrifter udi Niels Juels Epitaphio i Holmens kirke, Samtale med Rygtet (1699), Siunge-Koor (1674), Til... Brigitte Baltzlow (1689), Vinterparten af Danmarks og Norges forordnede Psalmebog (1689)

that the possessee is plural, such as in (173c). However, as (173d) shows, *deres* also appears with singular possessives. In (173e) *sin* and *deres* appear in the same sentence and with the same plural antecedent.

Kingo may have had a grammar in which *sin* and *deres* were both available as reflexives with plural antecedents. I find at least 20 examples of plural antecedent *sin* (and a small handful where it is not completely clear what the antecedent is) and the same amount of plural antecedent *deres*. There are too many examples with plural antecedent *sin* to make it likely that the use is purely an expression of grammatical errors. Kingo evidently uses the different options given by the two forms, *sin* and *deres*, actively in his texts, so that *sin* appears when he needs a single syllable word and *deres* when he needs a word with two syllables. If plural antecedent *sin* were not part of his grammar, he could in principle just have written his verses differently in a way that only required a *deres*. Kingo is the only writer in the present study with texts that are strictly poetic, and it must be an open question whether other poets may also have used plural antecedent *sin* as a rhythmic tool.

Holberg (1684-1754) was born in Bergen in Norway and must necessarily have heard more Norwegian language in his early years than many other Danish authors. For this reason, it could be the case that Holberg had a grammar that followed the Norwegian use of plural antecedent *sin*, i.e. where plural antecedent *sin* is the standard form. As Diderichsen (1939, 68, footnote 2) also notes, this is, perhaps surprisingly, not the case. Skautrup writes as a summary of Holberg's reflexive use that Holberg uses "sine for deres: (de) tugtede sine Fiender... mange forlode sine Booliger" (Skautrup (1953, 28)) (En. Holberg uses *sine* instead of *deres*: (they) hurt REFL's enemies... many left REFL's homes). Somewhat in opposition to this, Skautrup writes a few pages later that Holberg actually "undgår [...] norvagismene *sin* og *sine*, om end exemplar kan findes" (En. *Holberg actually avoids the Norwegianisms sin and sine, although examples can be found*) (Skautrup (1953, 37)). This last description also appears in the dedicated Holberg dictionary which gives *deres* as the most frequent reflexive form with plural antecedents. This claim is supported by the actual frequency of use in three Holberg texts that I have read through. In *Naturens og Folkerettens Kundskab* (adl.dk, p. 51-141) I find a single example of plural antecedent *sin* (reproduced in (174a)) and 43 *deres*. In *Mascarade* I find four *sin*, (174b)-(174d), and eight (possibly nine) reflexive *deres*. In *Jean de France* I find one plural antecedent *sin*, (174e), and 13 *deres*.

- (174) a. thi her contribuere **begge**₁ noget i Gierningen til **sin**₁
for here contribute both something in deed_{DEF} to REFL's
Næstes Vanære
neighbour's dishonour
'For here both of them contribute something in their deeds to their neighbour's dishonour.' (*Naturens og Folke-rettens Kundskab* p. 105)
- b. **De**₁ demaskerer sig begge, tales ved, og gir hinanden
they unmask REFL both, speak.PASS at, and give each other
sine₁ Ringe
REFL's rings
'They both unmask, talk, and give each other their rings.' (*Mascarade* p. 191, stage direction)
- c. **Unge Folk**₁ har **sine**₁ Tidsfordrive, og **gamle Folk**₂ **sine**₂
young folk have REFL's pastimes and old folk REFL's
'Young folk have their pastimes and old folk theirs.' (*Mascarade* p. 198, spoken line)
- d. Men Hr. Jeronimus, I kand lære af denne Historie, at
but Mr Jeronimus, you can learn of this story that
Mascarader₁ har ogsaa **sin**₁ Nytte
mascarades have also REFL's usefulness
'But Mr Jeronimus, you can learn from this story that mascarades can also be useful.' (*Mascarade* p. 238, spoken line)
- e. ...at **Cavalliers**₁ bære saadant om **sin**₁ Hals for at lade see
that cavalliers carry such around REFL's neck for to let see
Estime for de Damer,
reverence for the ladies
'... that cavaliers carry such things around their neck to show reverence for the ladies' (*Jean de France* p. 189)

Holberg does have quite a few examples of plural antecedent *sin*, but *deres* is indeed the most frequent form.

Johannes Ewald (1743-1781) only uses *deres*, never *sin*, with reference to plural antecedents in *De Fremmede*, in which I have read through the first 50 pages in the edition on adl.dk. The same is the case in *Fiskerne* (the 1969 edition on adl.dk).

I have read through the first 60 pages of Jens Baggesen's (1764-1826) *Labyrinten* (the 1971 edition on adl.dk) and find no examples of plural antecedent *sin*. In Grundtvig's *Nordens Mytologi* (1808) I find a single example of plural antecedent *sin*, reproduced in (175).

- (175) a. **Han og hans Følge**₁ tillagde sig Odins og de øvrige Gunders he and his company adopted REFL Odin's and the other gods' Navne, eller gave maaske Guderne **sine**₁, og vilde holdes for names, or gave perhaps gods.DEF REFL's and would hold.PASS for deres Repræsentantere paa Jorden. their representatives on Earth.DEF
'He and his company adopted the names of Odin and the rest of the Gods, or perhaps gave the Gods their own, and wished to be considered their representatives on the Earth.' (*Nordens Mytologi* (1808))

I have read through four texts by Hans Christian Andersen and I find the two examples of plural antecedent *sin* that I report in (176). *Deres* is by far the most frequent reflexive form with reference to plural antecedents.

- (176) a. Der gives **Mennesker**₁, der ret kunne udtale for Andre **sine**₁ there give..PASS people that right could express for others REFL's Lidelser og Elendigheder afflictions and miseries
'There are people who truly could express to others their afflictions and miseries.' (*Fodreise* p. 27)
- b. Luften var saa tør, at **Mund og Hals**₁ tabte **sin**₁ naturlige air.DEF was so dry that mouth and neck lost REFL's natural Fugtighed moisture
'The air was so dry that mouth and neck lost their natural moisture.' (*Spanien* p. 47)

Through the entire older and younger modern Danish period and all the way back to the oldest manuscripts *deres* is the most frequent reflexive form with plural antecedents. Plural antecedent *sin* can be found in texts by a range of authors and various kinds of texts through the entire period. Some speakers, Kingo being the most likely example in my survey, may have had *sin* and *deres* as optional variants in their grammars. I find no signs of a change in the use of *sin* and *deres* akin to the changes that can be found for *sig* and *dem* from the time of the Reformation and onwards, which have lead up to the language of the present day where *dem* is no longer a local reflexive form in the standard language. This is, however, just as expected if the change in *sig* and *dem* was spurred on by language contact with German: German *ihr* is used with reference to plural antecedents, and it corresponds to the Danish *deres*

in that both *ihr* and *deres* may be used both reflexively and non-reflexive. The German cognate of *sin*, *sein*, is singular, masculine/neuter. Consequently, there is no German form that would directly encourage a greater frequency of plural antecedent *sin*, as opposed to the case of *sig* where the German *sich* could inspire to more frequent use of plural antecedent *sig*. German language contact in the domain of possessives might rather support the retention of *deres* as a reflexive form and discourage plural antecedent *sin*.

3.3.3 *sig/dem* in the dialects in younger modern Danish (1750-now)

Diderichsen writes about the language of his own time that "i Pluralis har Dialekterne og det ældre (københavnske?) eller lavere Talesprog i al fald ofte *dem* for *sig* og normalt *deres* for *sin*, mens Skriftsprogsnormen og Grammatikerne kræver henh. *sig* og *deres*" (Diderichsen (1939, 67)) (En. *In the plural the dialects and the older (Copenhagen?) or lower spoken language at least often have dem instead of sig and normally deres instead of sin, while the written norm and the grammarians require sig and deres, respectively*). Diderichsen's observations are supported by the various available descriptions of dialects that are more or less contemporary with Diderichsen. In the small grammar that belongs to *Ordbog over Fjoldemålet* (Bjerrum and Bjerrum (1974, 24)) the authors describe that *sig* is used as the reflexive when the antecedent is singular and that *dem* is used as the reflexive when the antecedent is plural. Fjorde was a Danish dialect area in Schleswig. In *Jysk Ordbog* (a dictionary on the Jutlandic dialect) it is also noted that *dem* is the plural form that corresponds to the singular form *sig*, i.e. that *dem* is used reflexively in the area included in *Jysk Ordbog*. The material in the dictionary covers the time period from 1700 to 1930, in principle, but the great majority of the data describes the dialects in Jutland as they were spoken between 1850 and 1900. The dictionary provides the two examples in (177). The antecedent in (177a) is *fløde* (En. 'cream', transcribed phonetically as *flø* in the example), which is a singular or non-count word in standard Danish. In the Jutlandic dialects, however, *fløde* falls into the category *stof-pluralis* ('matter plural', Arboe (2016, 106)) where certain, typically liquid, collectives (*grød*, *suppe*, *kål* (En. *porridge*, *soup*, *cabbage*), among others) function as plurals.

- (177) a. om 'somə kuŋ ə 'flø₁ blyw 'su· åw dæm 'sjæ₁
in summer could the cream become sour of them self
'In summer the cream could go sour by itself.'

- b. a wo jo et' o **de Slaw**₁, dæ ku fjedt' **dæm**₁ igjæm'mel
 I was yeah not of the kind that could dawdle them through
 'I was not one of those who could dawdle their way through.'

Ømålsordbogen (I specifically use volume 3 from 1996, page 30 here) aims to describe the everyday language of rural speakers (fishermen, farmers, and their wives) on Zealand, Funen, Lolland-Falster, and surrounding islands. It here appears that *dem* is a normal reflexive form and the dictionary provides a set of examples that includes the ones in (178). Neither of the editorial boards of *Jysk Ordbog* and *Ømålsordbogen* have reached *sig* in the work on the dictionaries, which limits the available search space for data on reflexives to the entries on *dem* and *deres*. I assume that it is safe to say that *dem* in both Jutlandic and the *ømål* dialects seems to have been a normal reflexive form, in line with what Diderichsen describes.

- (178) a. for så forsluger **de**₁ **dem**₁ jo
 for then overeat they them yeah
 'for then they overeat, you know.' (said about horses that do not work)
 (Northern Zealand)
- b. **de**₁ tog godt til **dem**₁
 they took well to them
 'they really dug in' (meaning that they eat a lot) (Northern Zealand)
- c. **de**₁ sidder og keder **dem**₁ og har ikke noget at tage **dem**₁ til
 they sit and bore them and have not anything to take them to
 'they sit around and are bored and have nothing to do' (Falster)
- d. **de**₁ drak **dem**₁ fulde
 they drank them drunk
 'they got themselves drunk' (Tåsinge)

Jul Nielsen not does deal much with *sig* but he does note "at brugen af *dem* i stedet for rigsmålets *sig* ved henvisning til pluralissubjekt (hvilket er hyppigt forekommende, men ikke enerådende) er fælles for jysk og en række andre dialekter og talesprogsvarieteter (incl. lidt ældre sprog)" (En. *the use of dem instead of the sig of the standard language with reference to a plural subject (which is frequent but not the only form) is common to Jutlandic and a range of other dialects and spoken language varieties (including the slightly older language)*) (Jul Nielsen 1986, 44). The informants that provide the data for Jul Nielsen's investigation are for the most part born between 1880 and 1920, and the rest of the study draws on grammars and dialectal descriptions of

speakers that are even older. In *Ordbog over det danske Sprog*, which covers Danish between 1700 and 1950, it appears that *dem* is used as a reflexive in colloquial speech, particularly in Jutlandic, when the antecedent is plural.

Pedersen (2017) studies the development of *dem* and *sig* in the dialects and she provides the detailed, focussed data that the dictionaries cannot aspire to. I report her conclusions briefly here. In the spoken language of lower class Copenhagen, *dem* has been evident as a reflexive at least as late as the 1950s. *Sig* would have been the only possible reflexive in the spoken language of upper class Copenhagen at the same time. Pedersen concludes on the basis of a small text corpus that a similar development is evident in three provincial towns on Funen: Reflexive *dem* disappears first in Odense, the large provincial city, and later on in the smaller towns, just as it disappears first among the upper class Copenhagen speakers with longer educations. Eastern Jutlandic has most likely had the reflexive *dem* up until the beginning of the 20th century where *sig* began to gain ground according to Pedersen's investigation. This is also what happens to some extent in Western Jutlandic. Southern Jutlandic may have retained the reflexive *dem* a bit longer into the 20th century than the other Jutlandic dialects. The Funen dialects probably adopted *sig* slightly earlier than the Jutlandic dialects, and Pedersen (2017, 24) suggests that *dem* has been the most used reflexive form with plural antecedents up until the 1850s on Funen. The Zealandic dialects seem to reflect a similar developmental timeline, as Pedersen's earliest examples of Zealandic plural antecedent *sig* are from the second half of the 19th century. On the basis of a small amount of sound recordings and a manuscript of a dictionary, Pedersen finds that reflexive *dem* was the only used form on the eastern Danish island Bornholm at least until the 1930s.

3.4 LANCHART: Present-day spoken Danish

The Centre for Language Change in Real Time at the University of Copenhagen hosts a large database of Danish spoken language from the 1970s to the present day, the LANCHART corpus (see e.g. Gregersen (2009) for more details on the project). I have examined the use of reflexive *dem* in the LANCHART corpus in order to see whether it can be said to still exist in present day spoken Danish.

For the present study I read through approximately 15,000 individual sentences in LANCHART with the structure [V *dem*], i.e. all the sentences in the corpus that con-

tain a verb directly followed by *dem*. I chose this subgroup of all *dem* in LANCHART in order to reduce the amount of irrelevant results as much as possible. Irrelevant results are sentences with non-reflexive *dem*, which still make up the great majority of the data, even in the reduced subgroup. My specific search criterion unfortunately excludes possible examples of reflexive *dem* as prepositional complements. This is especially unfortunate given my previous assertions that this is a context where reflexive *dem* is actually useable also in the standard language. This specific context, in other words, would be very relevant to include in an extra search. The results show that locally bound reflexive *dem* is extremely infrequent: Out of 15,153 [V *dem*] sentences I found 32 examples of reflexive *dem*. This corresponds to a vanishingly small share of 0.2 %. I would expect a greater proportion if prepositional complement *dem* were included. I provide a few of the relevant examples of reflexive *dem* in (179). Most of the examples are non-theta, non-argument, instances of reflexive use (see section 2.3.5). This is not surprising given that the non-theta uses of *sig* greatly outnumber the theta uses of *sig* in general.

- (179) a. de sad og havde deres øh frokost og de havde deres øh øl og
 they sat and had their uh lunch and they had their uh beer and
 hvad **de**₁ nu hyggede **dem**₁ med ikke
 what they now cozied them with right
 'they sat and had their uh lunch and they had their uh beer and whatever
 they had a good time with, right' (LANCHART, woman, working class,
 born 1928, BYSOC0 1987)
- b. nogle lærlinge de kan få firmabil og andre lærlinge eller
 some apprentices they can get companycar and other apprentices or
 voksenlærlinge **de**₁ må klare **dem selv**₁
 adult-apprentices they must manage them self
 'some apprentices can have a company car and other apprentices or grown-
 up apprentices must manage on their own' (LANCHART, man, working
 class, born 1989, Familie1 2006)
- c. ja der er **svende**₁ der bare går derhjemme og passer **dem**
 yes there are journeymen who just go at-home and watch them
selv₁ sådan ja **de**₁ går og hygger **sig**₁ på værkstedet men
 self like yes they go and cozy REFL on workshop.DEF but
 alligevel...
 still
 'sure, there are journeymen who stay home and mind their own busi-

ness, like, yeah, they have a good time in the workshop but still.' (same speaker as (179b))

- d. jamen jeg tror da de festede simpelthen til langt ud på
well I think surely they partied absolutely to long out on
natten og jeg tror **de**₁ hyggede **dem**₁
night.DEF and I think they cozied them
'well I do think they absolutely partied long into the night and I think
they had a good time' (LANCHART, woman, working class, born 1965,
Familie1 2006 - (179b)'s mother?)
- e. i et hjem der hører der børn til for det var derfor folk
in a home there hears there children to for it was therefore people
de₁ giftede **dem**₁
they married them
'a home needs children for that was the reason that people got mar-
ried' (LANCHART, man, born 1939, Familie1 2007 - related to (179b) and
(179d))
- f. hvor man havde indkaldt forskellige uddannelsesinstitutioner
where one had summoned various educational-institutions
hvor **de**₁ skulle komme og præsentere **dem selv**₁
where they should come and present them self
'where various educational institutions had been summoned where they
were to come and present themselves' (LANCHART, woman, born 1968,
Odder2 2008)
- g. nej men det må **de**₁ jo bukke **dem**₁ for jo
no but that must they yeah bow them for yeah
'no but they have to bow down for that after all' (LANCHART, man, work-
ing class, born 1970, Tinglev2 2010)
- h. og det ville **de**₁ egentlig godt øh frasige **dem**₁ på daværende
and that would they actually good uh renounce them on that
tidspunkt
time
'and they actually did want to renounce that at that time' (LANCHART,
woman, working class, født 1964, Vinderup2 2006)
- i. og det synes jeg også der er rigtig fedt at **de**₁ ligesom tør at
and that thought I also that is really cool that they like dare to
være **dem selv**₁ og sige jeg er da ligeglad med alle jer andre
be them selv and say I am surely indifferent with all you others

'and I also think that that is really cool that they like dare to be themselves and say I totally don't care about the rest of you' (LANCHART, woman, working class, born 1991, Vinderup3 2006)

Six of the 32 examples are produced by the same speaker, a man born in 1939, and it seems reasonable to suppose that this speaker has *dem* as a local reflexive as an active part of his grammar. The speaker does use *sig* as a reflexive but only with singular antecedents ('jeg tror ikke det₁ har ændret sig₁ ret meget', En. 'I don't think it has changed that much'). With the other speakers, it seems that the reflexive *dem* is a more marginal variant. The speaker in (179b), a man born in 1989, uses *sig* with plural antecedents ('så hygger de₁ sig₁ med det', En. 'then they have a good time with that') and singular antecedents ('men han₁ har altid interesseret sig₁ for det', En. 'but he has always been interested in that') in addition to his two sentences with reflexive *dem*.

It is worth noting that almost half of the examples, 14 out of 32, are from the (North) American part (one from Argentina) of the *Danske Stemmer* (*Danish Voices in the Americas*) research project (see e.g. Kühl (2014)). The data for *Danske Stemmer* consists of recordings of Danish speakers who for various reasons emigrated to the Americas in the late 19th and early 20th century (or the descendants of these Danes). Examples specifically from *Danske Stemmer* are given in (180).

- (180) a. men der var også **mange danskere**₁ der skammed **dem**₁ over jo
 but there were also many Danes who shamed them over yeah
 at være danske
 to be Danish
 'but there were also many Danes who were ashamed of being Danish'
 (LANCHART, woman, no information about birth year, AmDa-tk)
- b. åh jeg rejste nærmest til Amerika ligesom **alle andre unge folk**₁
 oh I travelled basically to America like all other young people
 der gerne vil gerne ud at se **dem**₁ om
 who happily will happily out to see them about
 'oh I basically travelled to America just like all other young people who
 want to have a look around' (LANCHART, man, born 1906, AmDa-kbl)
- c. han foretog sig jo at stille an med stor diletant hvert år
 he managed REFL yeah to set up with large diletante every year
 hvert forår **de**₁ øvede **dem**₁ på det om vinteren
 every spring they practised them on it during winter.DEF
 'he did manage to set up a large diletante every year every spring they
 practised for it in the winter' (this speaker also has some plural antecedent

sig) (same speaker as b.)

Some of the examples definitely sound as though the reflexive form has been influenced by the English reflexive construction. The *dem selv* in (181a) is completely equivalent to the English intensifier *themselves*, as should also be clear from the gloss and translation. In standard Danish, the intensifier would have been *selv*, not *dem selv*. In the example in (181b) it is less clear that English language contact should be responsible for the reflexive *dem*, and there is no English expression that directly translates *klare sig* (En. 'manage') which could be thought to influence the reflexive.

- (181) a. and de de fælder kun træer **de**₁ kan bruge **dem selv**₁
 and they they cut only trees they can use them self
 and they they only cut down trees they can use themselves' (LANCHART,
 man, born 1899, AmDa-kbl)
- b. jeg kunne ikke noget engelsk men mine to brødre kunne en lille
 I could not any English but my two brothers could a little
 smule så **de**₁ klarede **dem**₁ lidt
 bit so they managed them somewhat
 'I did not know any English but my two brothers knew a little so they
 managed somewhat' (LANCHART, woman, born 1906, CanDa 1982)

If the emigrants brought along a dialect with reflexive *dem*, this could very well be what is still seen in their language in the examples given here, in addition to some impact from the English reflexives. Reflexive *dem* was quite standard in the time before and around their period of emigration in the late 19th and early 20th century, cf. the results from Pedersen (2017) discussed in section 3.3.3. The Danish speakers in the Americas will not have felt as much pressure from the Danish standard language or from the Danish written norm (as the speakers in the study typically do not write Danish at all), which both point towards using *sig* as the plural antecedent reflexive. The speakers in *Danske Stemmer* are also quite a bit older than most other LANCHART speakers, most of them born around 1900. This places them, age-wise, right at the time where *sig* is gaining ground in most of the Danish dialect areas, according to Pedersen (2017). The biographical information in the database does not list when the speaker emigrated (or whether they were born in the Americas), but birth years around 1900 presumably either indicate that they are descendants, or that they emigrated at a time where plural antecedent *sig* was still not completely adopted in most dialects. The youngest speaker in my selection of *Danske Stemmer* examples was

born in 1934, which makes 99 % of the other LANCHART speakers younger than her. Out of 15,153 sentences with *dem* 1286 are from *Danske Stemmer*, which corresponds to 8 %. Out of 32 sentences with reflexive *dem* 14 are from *Danske Stemmer*, which corresponds to 44 %. Older age, language contact with English and Spanish, and lack of contact with standard Danish and written Danish together make up one way of explaining the disproportionately large share of Danish speakers in the Americas in the 32 examples of reflexive *dem*.

I find examples of locally reflexive *dem* here and there in sentences spoken by both older and younger speakers in the LANCHART database, but it is definitely more frequent in the speech of older speakers in the database, particularly the speakers who emigrated to the Americas in the late 19th and early 20th century who are also by far the oldest speakers in the corpus. The results from Pedersen (2017) and the descriptions from various dialect studies all point towards reflexive *dem* as a feature of spoken language which largely started disappearing from the beginning of the 20th century and onwards. Reflexive *dem* has been maintained in the language in non-local contexts (discussed in section 2.3), apparently to some extent in various override *dem selv* constructions (discussed in Sørensen, Ehlers, and Vikner (2020)), and with binding into certain prepositional phrases where there is quite an amount of optionality between *sig* and the personal, non-reflexive, pronouns, including *dem* (see e.g. Rooryck and Wyngaerd (2011) which has an entire chapter on these so-called *snake-sentences*). However, locally bound reflexive *dem* in the modern language, perhaps apart from some few older dialectal speakers, must be considered a very marginal variant at most.

3.5 Conclusion

The Danish reflexives *sig* and *sin* make different demands on their antecedents. *Sin* normally does not allow plural antecedents, whereas *sig* allows both singular and plural antecedents. This is why Danish speakers say *forældrene₁ elsker deres₁ børn* (not *sine børn*) (En. 'the parents love their children', not 'REFL's children'), but *forældrene glæder sig til weekenden* (not *glæder dem*) (En. 'the parents look-forward REFL to the weekend', not 'look-forward *dem*'). In earlier stages of the language, both *sin* and *sig* were number sensitive and the non-reflexive *deres* and *dem* were used as local reflexives with plural antecedents. This difference between *sin* and *sig* is a reasonably

modern difference which was gradually standardized in the written language after the Reformation, i.e. after around 1500, and in the spoken language likely a few hundred years later, after 1850.

In the Danish found in runic inscriptions from around 1000 AD, *sin* is used with both plural and singular antecedents. This corresponds to the use of *sin* in Common Germanic and Old Norse. The same would probably be the case for *sig*, but the available language data is fairly narrow in scope and shows neither positive nor negative evidence for the case of *sig*. However, both Common Germanic and Old Norse had a reflexive *sik* which was used with both singular and plural antecedents and this makes it likely that Runic Danish would have had this feature, too, and that it carried over into the language which can be read in the Law in Scania.

A few centuries later, in the provincial laws from the 13th century, the reflexive use has changed. *Sin* and *sig* are primarily used with singular antecedents and *deres* and *dem* are used as local reflexives with plural antecedents. This is the generally the case in the West Danish laws (discounting Scania), while Law of Scania still contains an approximately equal proportion of plural antecedent *sin* to plural antecedent *deres* and a large majority of plural antecedent *sig*. The reflexive use of *sin* and *sig* with singular antecedents and *dem* and *deres* with plural antecedents most likely stayed in the spoken language until the 20th century, even though *sig* became the most frequent form with both singular and plural antecedents in the written language after the Reformation. *Deres* is by far the most frequent plural possessive reflexive form through the entire period between 1200 and the present day, and *sin* occurs sporadically with plural antecedents in the texts of various writers in the older and younger *nydanske* period, from 1500 to the present day. The hymn writer Thomas Kingo (1634-1703) is the sole exception in the data set as his poems contain approximately equal shares of plural antecedent *sin* and plural antecedent *deres*. Kingo writes poetry and it is very possible that the constraints of the genre play a part in his usage. Perhaps a rigid formal structure generally allows or encourages a looser use of marginal items in the grammar, such as more frequent reflexive *sin*.

Data from LANCHART shows that locally bound reflexive *dem* is an extremely marginal form in modern Danish. *Sig* is the standard form with both singular and plural antecedents. Local reflexive *dem* can still be found sporadically but it is rather more frequent with older than with younger speakers.

Chapter 4

Sin with plural antecedents in KorpusDK

This chapter is the first of two empirical investigations into the use of *sin* with plural antecedents in modern Danish. This first study is a corpus study of the occurrence of *sin* with plural antecedents in the large (56 million words) Danish text corpus KorpusDK.

Corpus data can be used in various ways and for many different purposes and the use of corpus data within the field of experimental syntax seems to be gaining ground (Francom 2021). One strong advantage of corpus data is that they are widely, and usually freely, available also for smaller languages such as Danish. They can be used reasonably easily to investigate naturally occurring language and are thus a good supplement to e.g. the various introspection approaches that are widely used within the generative grammar tradition.

I use the data from my corpus study to compare the behaviour of plural antecedent *sin* with the behaviour of the standard forms in Danish, i.e. singular antecedent *sin* and plural antecedent *deres*. The results from the corpus investigation were further used to inform the design of the acceptability judgment study of plural antecedent *sin* that I discuss in chapter 5. Juel Jensen (2009b) did a similar study on plural antecedent *sin* in a subset of the spoken LANCHART corpus and he found a much higher, if still relatively low, frequency of plural antecedent *sin* (7 %) than what I find in the written corpus (which is 0.6 % of all occurrences of *sin*).

I was initially interested in knowing whether plural antecedent *sin* could be found in KorpusDK at all, or whether the form was too marginal (or perhaps rather too much

of a feature of spoken language). *Sin* does indeed occur with plural antecedents in KorpusDK but it is very infrequent. The original, perhaps rather optimistic, intention of the study was to find and analyze all the instances of *sin* with plural antecedents that can be found in KorpusDK. *Sin* itself is a highly frequent form in Danish and it occurs 188,404 times in KorpusDK across its three inflectional forms (*sin*, *sit*, *sine*). As a scale of comparison, the first person singular pronoun *jeg* (En. *I*) occurs about twice as frequently as *sin* (405,298 times in KorpusDK) and the second person singular pronoun *du* (En. singular *you*) occurs about half as frequently as *sin* (106,604 times in KorpusDK). KorpusDK is not dependency parsed, which means that there is no way to search automatically and directly for the relationship between *sin* and its antecedent. Consequently, my search for all instances of plural antecedent *sin* was very much a search for the proverbial needle in the haystack. Given enough time and various experimental approaches, however, I did eventually succeed in finding a great deal (if not necessarily every single instance) of plural antecedent *sin* in KorpusDK. The standard approach within corpus linguistics is to take a smaller subset or a random smaller selection of a corpus in order to approximate the picture in the entire corpus. With a more limited amount of time (or tolerance for repetition), that approach is certainly much more reasonable. The approach that I chose is not something that I would generally recommend but it has resulted in a data set with a very high degree of detail, fidelity, and representativeness for the corpus as a whole.

In the remaining parts of this chapter, I go through the process of data collection in KorpusDK with various tools. I initially started out with using the online version of KorpusDK which is unfortunately not ideally suited for a corpus investigation of this magnitude. I eventually ended up using the offline version of the corpus which further necessitated the creation of several *R* (R Core Team 2021) and *Python* (Python Software Foundation 2016) scripts in order to even be able to perform searches in the corpus. I describe and discuss the steps in this process, and the differences between the online and offline versions of the corpus. Finally, I discuss the actual data and how plural antecedent *sin* patterns similarly to and differently from the standard reflexive forms, non-plural antecedent *sin* and plural antecedent *deres*.

4.1 Description of the online version of KorpusDK

The largest Danish-language annotated corpus is KorpusDK. The description of the corpus here in section 4.1 is largely borrowed from my Master's thesis, Ehlers (2017). KorpusDK is owned and created by *Det Danske Sprog- og Litteraturselskab*. The corpus consists of two approximately equal halves, Korpus90 and Korpus2000, which together make up the full KorpusDK corpus. Korpus90 consists of texts that cover the time span from 1983 to 1992 and contains 28 million words. The corpus is mixed, as opposed to specialized, and contains text from newspapers both local and national, magazines, Danish literature, various specialized journals, the two, at the time, major Danish news stations (Danmarks Radio and TV2), passages from a Danish translation of the Bible, and finally texts written by high school students.

Korpus2000 is the same size as Korpus90, 28 million words, and covers the years 1998 to 2002. Just like Korpus90, Korpus2000 contains texts from newspapers, journals, and literature. Additionally, Korpus2000 contains material from several more schools than Korpus90, from a sports association, and a student exchange association, from companies, from websites, and from individuals. The two corpora together make up the 56 million word KorpusDK.

The corpus has been automatically annotated using the tagger *DanPars*, developed by Eckhard Bick. The tagger is a Constraint Grammar parser, meaning that it annotates the text based on multiple context dependent rules, usually defined by the linguist working on tagging the text. The words in KorpusDK are tagged for part of speech (PoS) and morphology but not for any kind of dependence between the words. Because the annotation is automatic, there will always be some mistakes in the corpus. For instance, a verb could be tagged as a noun if the verb for some reason is in a context or has a form that makes it more likely to be a noun, as defined by the tagging criteria. This will be true of all corpora and should be kept in mind when using corpus data.

The online corpus search engine has three levels of complexity. Each level adds more freedom to specify the search strings but also demands a higher level of technical proficiency with the search language. All three ways of searching result in a *concordance*, which is the collection of text strings found by the search in the corpus. The concordance allows for the strings to be sorted in different ways and the user can check a few boxes to reveal the tagging that is behind the results. Additionally, it is possible to click on the individual strings to reveal more of the textual context and the

biographical information encoded about the text.

In the basic search (*Standardsøgning*), words or sentences are entered in the text box and the resulting concordance of strings comes up. This comes at a price, namely that the underlying search string always is more general than the search input. As a standard, the basic search is programmed to search for the words entered in the text box at a distance of 0 to 3 words apart. Additionally, both the entered word (the *word form*) and the non-inflected form (the *lemma form*) will be included in the underlying search string. This means that a basic search usually provides a larger and less specific concordance than what might be desirable.

The next level of search complexity (*Udvidet Søgning*) allows the user somewhat more freedom to search for specific morphological tags, while providing an interface that helps the less proficient user perform searches.

The most complex level (*Formel søgning*) allows the user full freedom to manipulate and refine search strings but it also requires a working understanding of the corpus processing language CQP that the search language in KorpusDK is based on. The website provides a good deal of assistance and a short manual in using the query language, as well as handy overviews of the available PoS and morphological tags.

4.2 Description of the offline version of KorpusDK

The online version of KorpusDK has a number of limitations and occasional bugs. This makes it challenging to work with big data sets and large-scale searches. For this reason, and after discovering that an offline version exists at all, I chose to work with an offline version of the corpus.¹ This, first of all, allows much greater flexibility in performing large-scale searches heavy in computer processing. Similar searches in the online version of KorpusDK will otherwise often lead to the search engine breaking down repeatedly. More importantly, the results from the online KorpusDK search engine are limited to a maximum cap of 5000 results. In the cases where a search returns more than 5000 results, the additional results are unavailable. KorpusDK reports that it contains 188404 instances of *sin*. Since the intention of my corpus search was to obtain information about all instances of plural antecedent *sin* in the corpus, or as

¹The offline version of KorpusDK is available to researchers and other interested parties and is password protected. At the time of writing (2019, and still in 2023) the corpus can be downloaded from the website <https://korpus.dsl.dk/resources.html> and the password is available by request from Jørg Asmussen at *Det Danske Sprog- og Litteraturselskab*.

close to all as realistically possible, the cap of 5000 coupled with frequent technical errors from the website itself in larger searches made this goal practically infeasible.

There are three corpora available for download: Korpus90, Korpus2000, and Korpus2010. Korpus90 and Korpus2000, as described in more detail in section 4.1, make up the KorpusDK that is available online. Korpus2010, when I began the corpus investigation in 2016, was only available in this offline version or as part of the still-under-development search engine CoREST (Asmussen 2016). From 2018 and onwards, CoREST is available for download at <https://korpus.dsl.dk/>. In addition to Korpus90, Korpus2000, and Korpus2010, the publically available² CoREST as of 2023 contains the Wikipedia 2017 corpus and the annually updated corpora TiDK 2018, 2019, 2020, and 2021, as well as Korpus90, Korpus2000, and Korpus2010.

The offline corpora come packaged as three ZIP files of approximately 800 MB in all. Each corpus has its own folder. For Korpus90 and Korpus2000, each folder contains precisely 10,000 .txt files of text and annotation. Together, Korpus90 and Korpus2000 take up 2.3 GB of space and the larger Korpus2010 in itself takes up 1.65 GB. The texts are scrambled, meaning that the files only contain individual sentences and not full texts.

The individual files are structured through lines, columns, and sentence boundaries. The beginning of each sentence is marked with an identifying string and ends with a generic closing string which is always `</s>`. The opening string has the general structure in (182a) and (182b) is a specific example. I point this structure out because it allows me to incorporate sentence boundaries into the software that I designed for searching in the corpus.

- (182) a. `<s id="year-file-sentence">`
 b. `<s id="90-21-83">`

The organization of the full corpus texts into scrambled sentences was done automatically when the offline text files were created. This automatic division and scrambling has resulted in some amount of errors in the files. For instance, the system seems to deal fairly badly with sentence-internal punctuation, leading to frequent instances of erroneous sentence boundaries prompted by abbreviations or integers written with full stops. For instance, the abbreviation *bl.a.* (En. *for instance*) would prompt the system to create a sentence boundary after *bl* because of the full stop, and so would

²The team behind *Den Danske Ordbog* maintain an even larger version of the corpus, BAKSPEJLET, to use for updating the dictionary. This resource is not publically available.

Table 4.1: Word structure in offline KorpusDK files. The word "var" (En. *was*) is used as an example to demonstrate the six columns.

1	2	3	4	5	6
var	var	_	være	V	VF:—:ta:—

¹ Simplified orthography (no capital letters)

² Orthography as in the text

³ Punctuation or empty space that follows the word

⁴ Lemma form (uninflected)

⁵ Part of speech

⁶ Morphological tagging

e.g. the number *10.000*.

Each word has its own line with six columns (pers. comm., Jørg Asmussen), structured as summarized in table 4.1.

Column 1 is a simplified orthographic version of the word. Column 2 is the orthography of the word as it is in the text. Column 3 gives the punctuation or empty space (indicated by an underscore, `_`) that follows the word. Column 4 is the lemma (uninflected) form of the word. Column 5 gives the word's part of speech (PoS).

Column 6 gives more detailed morphological information with tagging according to the tagging system outlined in Asmussen (2015) and presented in a condensed form in table 4.2. The tagging system of column 6 is positional and depends on the specific word class. What this means is that each tag always occurs in the same position out of the 12 slots and that the available tags for e.g. a noun and a verb will be different. Table 4.2 is adapted from Asmussen (2015, 9) and presents the full system for PoS tagging in the corpus. An asterisk, `*`, indicates that this position is in use for the word class in question. A hashtag, `#`, indicates that this value is "underspecified" (Asmussen 2015, 6) and that any value could in principle be used here. Some letters (e.g. *C* as in *common*, *cardinal* and *reciprocal*) are used multiple times across the table. This has the consequence that the individual tags only have specific meaning when they are coupled with a particular word class. The last category in the table, Residual, represents the cases where the automatic tagger has been unable to assign a tagset to the word or words. There are whole sentences in the corpus just tagged XY which equates to no tagging at all.

Table 4.3 shows the tags that are applied to the words beyond the basic word class information from table 4.2. Table 4.4 spells out the abbreviations used in table 4.3.

Table 4.2: ePOS word class tag set (Asmussen 2015, 9)

POS		Sub.		Paradigm
V	Verb	I	infinitive	VI:---:~*~---
		F	finite	VF:---:~*~---
		M	imperative	VM:---:~*~---
		G	gerund	VG:****:~*~---
		P	participle	VP:****:~*~---
		T	past part.	VT:siu#:~*~---
		D	adv. part.	VD:---:~*~---
A	Adjective	C	common	AC:****:~*~---
		D	adverbial	AD:---:~*~---
L	Numeral	C	cardinal	LC:~*~:~*~---
		O	ordinal	LO:~*~:~*~---
N	Noun	C	common	NC:****:~*~---
		P	proper	NP:****:~*~---
P	Pronoun	C	reciprocal	PC:~*~*~*~:~*~---
		M	demonstrative	PM:~*~*~*~:~*~---
		I	indefinite	PI:~*~*~*~:~*~---
		O	possessive	PO:~*~*~*~:~*~---
		P	personal	PP:~*~*~*~:~*~---
R	relative	PR:~*~*~*~:~*~---		
D	Adverb	-		D:---:~*~---
I	Interjection	-		I:---:~*~---
T	Preposition	-		T:---:~*~---
C	Conjunction	C	coordinating	CC:---:~*~---
		S	subordinating	CS:---:~*~---
U	Unique	I	inf. marker	UI:---:~*~---
		S	<i>som/der</i>	US:---:~*~---
E	Lexical element	W	word formation	EW:---:~*~---
M	Inflectional ending	N	attached to a noun	MN:****:~*~---
		V	attached to a verb	MV:---:~*~---
		A	attached to a adj.	MA:****:~*~---
X	Residual	S	symbol	XS:---:~*~---
		F	foreign	XF:---:~*~---
		Y	tagging error	XY:---:~*~---

Table 4.3: ePOS additional tags (Asmussen 2015, 7)

CLASS	nominal			verbal			additional			
	NUM	DEF	CAS	GEN	TMP	VOC	DEG	PER	RFL	POS
	s	i	u	c	s	a	p	1	y	s
	p	d	g	n	t	p	c	2	n	p
			f				s	3		
			n				a			

Table 4.4: ePOS abbreviations

Class	SUBCLASS		value	
Nominal	NUM	number	s	singular
			p	plural
	DEF	definiteness	i	indefinite
			d	definite
	CAS	case	u	unmarked
			g	genitive
			f	fossilized
			n	nominative
			c	common
	GEN	gender	n	neuter
Verbal	TMP	tense	s	present
			t	past
	VOC	voice	a	active
			p	passive
Additional	DEG	degree	p	positive
			c	comparative
			s	superlative
			a	absolute superlative
	PER	person	1	first person
			2	second person
			3	third person
	RFL	reflexiveness	y	yes
			n	no
	POS	possessor	s	singular
			p	plural

4.3 Scripts and technical considerations

I have written and later revised several pieces of software (*scripts*) that allow me to search through the text files that constitute the offline KorpusDK. Professor Søren Feodor Nielsen (CBS) kindly provided me with a first draft for a search script based on the make-up of the data files. The code for this initial script, `Katrine.indlæs.data.R`, as well as the revised versions with my own additions can be found in appendix B. The initial script is written for the open-source programming tool R. The script uses a *for* loop that runs through a `.txt` corpus file and writes the words of every sentence that contains *sin*, *sit*, or *sine* to a new file. The final output of the code are two files, one called `sentences.txt` and one called `output.txt`. `sentences.txt` contains a list of sentences separated by newline commands (`\n`). The words in the sentences are all lower case and there is no punctuation. `output.txt` gives the full picture for the results, in that it is a list of all the meta information (the information presented in table 4.1 and table 4.2) that belongs to the words in `sentences.txt`. `output.txt` has the same content as the raw data files but only for the relevant sentences, i.e. the ones that contain *sin*, *sit*, or *sine*. For the same reason, the file `sentences.txt` is suited to be read by a human and can be used to manually sort the data. The `output.txt` files can be used to perform more searches as desired on a smaller data set.

In order to make the code useful for my purposes, I changed the script to accommodate the large amount of different `.txt` files that make up the corpus. Feodor's script only works for one file at a time and it would be far too time-consuming to load each of the 20,000 files individually. The updated script (`Katrine.indlæs.data-UPDATED.R`) loads every file in a given folder that fits the name pattern `"*.txt"`. `*` is a regular expression in R that matches every string of characters that occurs before the file extension `.txt`. The corpus files are labelled `"abcd.txt"` where `abcd` are all the integers in the span of 0000 to 9999, making up precisely 10,000 individual data files. The updated script also accommodates Danish special letters `æøå` through a specification of UTF-8 encoding. Finally, it contains the specification `quote=""` which ensures that all rows are processed as though of uniform length, i.e. as having the six columns indicated in table 4.1. This was a necessary addition to let the script process the initial and final line of each sentence which have a slightly different column structure. It also lets the script process lines that have a different structure due to errors in the automatic annotation.

Having made these changes, the code works reasonably well for my specific pur-

poses. It is, however, significantly hampered by the fact that R is not ideally suited to handle these kinds of *for* loops and that the script consequently runs fairly slowly. Combined with the very large amounts of data, I opted for running the scripts mainly overnight and splitting the processing into smaller chunks of 1,000 to 5,000 text files rather than the full 10,000 files. Another script design or another programming language would likely have cut down processing time significantly but I did not have the programming experience to make those changes at the time. The programming language Python would be my first choice if I were to redo this process.

I encountered a further technical issue related to the interplay between my script and the errors in the raw data files of the corpora. The errors that I discovered were all minor ones probably to do with faulty mark-up of special characters. Specifically, there were issues with the characters ", {} and _ which are all characters that have a special function in most programming languages. Most likely, the code for creating the corpus has had problems handling these which leads to issues in the raw data set. These errors make the running script crash and I spent a significant amount of time on finding and correcting the errors in the text files. The Korpus2010 data set appears to contain many more errors than the Korpus90 and Korpus2000 data sets. I found the problematic files in the Korpus2010 data set and modified them by removing faulty (empty, i.e. no text was removed) lines. I have not processed Korpus2010 more than the initial automatic search for sentences with *sin*, *sin* and *sine* and, separately, *deres*. Due to the higher number of errors in the data set, the output from Korpus2010 is also more prone to errors than Korpus90 and Korpus2000. This is most easily visible as a higher number of duplicate sentences in the output for Korpus2010. I chose to not include Korpus2010 in the final data set because of the amount of errors which, in combination with the sheer size of the corpus, made an already very time-consuming data collection process far too time-consuming.

4.4 Search methodology for plural antecedent *sin*

At an initial stage in the project, before having access to the offline version of the corpora, I performed a thorough search through the online version of KorpusDK in order to find as many examples of plural antecedent *sin* as possible. In the interest of limiting the amount of irrelevant data, and to make the searches feasible given the limit of 5000 results per search, I chose to limit my search environment to a maximum of

Table 4.5: KorpusDK - online search environments

noun, PL ^a				verb				sin
noun, PL	w ^b	(w) ^c	(w)	verb	w	(w)	(w)	sin
noun, PL				verb	w	(w)	(w)	sin
noun, PL	w	(w)	(w)	verb				sin
noun, PL								sin
noun, PL	!verb ^d	(!verb)	(!verb)	(!verb)	(!verb)	(!verb)	(!verb)	sin

^a noun tagged as plural

^b word, any part of speech

^c word, any part of speech, optional

^d word, any part of speech but verb

7 words between the plural nominal (the potential plural antecedent) and *sin* (short-hand for all three inflectional variants: *sin*, *sit*, *sine*). I chose this (somewhat arbitrary) distance limit on the basis of experimentation with the sizes of the concordance returned by smaller and larger distance limits. The specific search environments are illustrated in table 4.5. They are intended to be complementary, i.e. they cover the full span of possible results without too much overlap.

I performed the same searches with all other plural-marked non-nominal antecedents but ended the process about halfway through the full data collection. I cut this online search short after finding out about the offline corpus. The results from the online search provide every sentence in the corpus where *sin* is not topicalised, but only in a limited environment: that where *sin* and its potential plural-marked antecedent are a maximum of seven words apart and where the antecedent is a nominal.

In the offline search, there is no such arbitrary limit on how large a concordance can be and consequently no reason to put a limit on the distance between antecedent and *sin*. This also means that the offline searches returned very large results files, e.g. around 32,000 sentences with *sin* and a preceding plural from Korpus90 and similar numbers for Korpus2000. The plural antecedent *sin* examples are so rare in the corpus that a manual search is exceedingly time consuming due to the amount of false positive results. False positive results are sentences that contain a *sin* and a preceding plural word where the plural word is not the antecedent of *sin*. A later random sample and extrapolation (discussed in 4.5.3) suggested that I could expect to find around 500 examples of plural antecedent *sin* in Korpus90, meaning that around 31,500 of the 32,000 results sentences would be false positives. I provide such an example in (183) where *sin* has a singular antecedent, *Linda*, but is also preceded by two plural marked

words, *nogle* (En. *some*) and *gange* (En. *times*).

- (183) a. **Linda**₁ havde *nogle gange* beskrevet **sin**₁ far...
 Linda had some times described REFL's father
 'Linda had sometimes described her father...' (Korpus90)

The corpus files are not dependency parsed so the search engine has no way of determining which word in the sentence that is the antecedent of *sin*. Consequently, I had to rely on linear order as the sorting factor, which gives rise to the large amount of false positives just mentioned. In the initial stages of the search process, I experimented with how much irrelevant data that could be removed automatically. The various iterations of this experimental process are listed in (184).

- (184) a. All instances of *sin* in the corpus (14,204 KB – 89,809 sentences in Korpus90)
 b. All of instances of *sin* preceded by a plural word (5,554 KB – 32,779 sentences in Korpus90)
 c. The subset of (184b) where there is a verb in between the plural word and *sin* (4,900 KB – 27,963 sentences in Korpus90)
 d. All instances of *sin* preceded by the first verb in the sentence which itself is preceded by the first plural word in the sentence (2,251 KB – 14,294 sentences in Korpus90)

In the final data file, it is evident that the great majority of the plural *sin* examples contain a plural word followed by a verb and then *sin*. This would match a prototypical finite V2 main clause where the plural subject antecedent is moved to CP-spec and followed by the finite verb in *C^o*. Consequently, (184d) seemed like a good search strategy that would both drastically cut down the number of false positives and return the majority of the relevant examples. However, the logic of the argument and the practicality of the search structure do not quite match up in the way that I wrote the code: The demand that the verb must be the first verb rules out a number of relevant examples such as the one in (185). The problem here is that *gennemtænkte* is coded in the corpus as a verb, giving the effect that the first verb-coded element in the sentence precedes the first plural word, *sjove*, and the sentence is consequently excluded.

- (185) a. Mening med galskaben og **gennemtænkte sjove film**₁, der
 meaning with craziness.DEF and through-thought funny movies that
 henter **sin**₁ humor i nuet.
 get REFL's humour in now.DEF

'Method to the madness and thought out funny movies that get their humour from the now.' (Korpus90)

I eventually chose to use the search result from (184b), since (184d) excluded too many relevant results. Additionally, the difference between (184b) and (184c) seemed too small to justify the possible exclusion of e.g. interrogatives and even regular V2 constructions with non-subject constituent movement to CP-spec. 25 relevant examples from Korpus90 and 32 relevant examples from Korpus2000 were included in the final data sample by choosing search strategy (184b) instead of (184c). (186) gives a few examples of sentences that are excluded by search strategy (184c) but not (184b). (186a) and (186b) are main clauses with V2 constituent movement of a non-subject: The complex adverbial *i dette perspektiv* in (186a) and the one-word adverbial *så* in (186b). The finite main verbs *fik* and *følger* are moved to C^0 and therefore linearly precede the plural subject. This means that they are not included in the relevant results with search strategy (184c) but they are included in the relevant results with search strategy (184b). In (186c) the reflexive *sit* sits in a constituent that modifies the CP-spec of the relative clause and precedes the verb *må*, which would again exclude the sentence if I had chosen search strategy (184c).

- (186) a. I dette perspektiv fik **fagene**₁ **sin**₁ ofte uudtalte
 in this perspective got subjects.DEF REFL's often unspoken
 begrundelse.
 reason
 'In this perspective the subjects got their often unspoken reason.' (Korpus90)
- b. ... så følger **fårene** **med kors på ryggen**₁ **sin**₁ hyrdes
 then follow sheep.PL-DEF with crosses on back.DEF REFL's shepherd's
 røst.
 voice
 '... then the sheep with crosses on their back follow the voice of their shepherd.' (Korpus2000)
- c. ... men om **alle dem**₁ der enten i **sit**₁ job eller af psykisk
 but about all them who either in REFL's job or of mental
 nødvendighed må afsted.
 necessity must away
 '... but about all of those people who either in their job or out of mental necessity must leave.' (Korpus90)

The choice of search strategy (184b) still excludes instances of topicalised *sin*, although some of these do appear in the results files by happy accidents. An example of this is e.g. where the topicalised *sin* is nested within a longer clause that contains a preceding plural-marked word, as illustrated in the example in (187). The plural-marked *marginaler* linearly precedes the topicalised *sin*. This places the sentence in the results file even though *sin*'s antecedent *fynboerne* is further down in the clause.

- (187) a. ... og med Sepp Pionteks små *marginaler* på **sin**₁ side, scorede
and with Sepp Piontek's small marginals on REFL's side scored
fynboerne₁ to gange inden for tre minutter...
Funen-players.DEF two times within for three minutes
'... and with Sepp Piontek's small marginals on their side, the players
from Funen scored twice in the span of three minutes...' (Korpus2000)

The great advantage of the offline search is that it sets no restrictions on how far apart the antecedent and *sin* can be. This means that the general coverage of my offline search is much greater than the online search. Having made the online search, however, allows me to compare the results from the online search with the relevant results from the offline search. By comparing these two, I can to some extent estimate how much data my offline search is erroneously missing and either attempt to remedy this lack or just bear it in mind when analyzing the data. The online data provided the argument that search strategy (184d) was excluding too much relevant data. In addition, the online data provided 33 relevant examples from Korpus90 and 33 relevant examples from Korpus2000 which were missed by the initial automatic offline sorting. In most cases, the reason that the offline automatic search missed these two times 33 sentences is that the words in the sentences had been tagged erroneously. Two examples of wrong tagging are shown in (188) with online and offline tags given for each word. It is evident here that the online and offline versions use different tagging standards (DanPARS in the case of online KorpusDK, see page 167, and the ePOS tagger in the case of offline KorpusDK, see page 170). The two taggers lead to rather different results in cases such as these. In (188a), the plural antecedent *alle* is tagged as plural (the *_P_* in the tag) and in (188b), the same word is tagged as singular (the tag *NP:siu* means *Noun Proper: singular indefinite unmarked*). The difference is perhaps even more striking in the examples in (189). In (189a), the plural antecedent *kunderne* is tagged as plural, while everything, including the plural antecedent, in (189b) is given the error tag *XY:—:—:—*.

- (188) a. Alle, der skrev sit navn...
nG_P_NOM nG_nN_NOM IMPF_AKT NEU_S NEU_S_IDF_NOM
'All who wrote their name...' (Korpus90, online tagging)
- b. Alle, der skrev sit navn...
NP:siu#:--:-- US:--:--:-- VF:--:ta:-- PO:s-n:--:--3ys NC:siun:--:--
'All who wrote their name...' (Korpus90, offline tagging)
- (189) a. ... at kunderne så kan koncentrere
... USPEC UTR_P_DEF_NOM USPEC PR_AKT INF_AKT
sig om sine kerneopgaver...
nG_3S/P_ACC USPEC nG_P UTR_P_IDF_NOM
'... that the customers then may concentrate on their primary tasks...'
(Korpus90, online tagging)
- b. ... at kunderne så kan koncentrere
... XY:--:--:-- XY:--:--:-- XY:--:--:-- XY:--:--:-- XY:--:--:--
sig om sine kerneopgaver...
XY:--:--:-- XY:--:--:-- XY:--:--:-- XY:--:--:--
'... that the customers then may concentrate on their primary tasks...'
(Korpus90, offline tagging)

Another problem is that some sentences from online KorpusDK are apparently simply not included in the offline KorpusDK. I see no obvious reason for why this should be and can only observe that it is. One possible common characteristic is that the majority of these non-offline sentences are from either the very beginning (1984) or the very end (1992) of the data collection period for Korpus90 (there are none of these from Korpus2000), but that is not the case for all of the examples. The 66 examples that were only found in the online search, both those with tagging errors as in (188) and those not found in the offline corpus, are included in the final dataset.

I performed the chosen automatic search on Korpus90 and Korpus2000 separately because the files are in separate folders and because it could have been interesting to be able to compare the two subcorpora. I performed one search for *sin* (*sit*, *sine*) for each corpus and one search with the same parameters for *deres* for each corpus. I sorted the final four sentences.txt data files (one *sin* file and one *deres* file from Korpus90 and Korpus2000, i.e. four in all) manually. The exclusion/inclusion criteria for the sorting process are described in section 4.6. I used the custom sorting function in Excel to speed up the process by sorting the sentences according to the placement of *sin/deres* in the sentence and highlighting the instances of *sin* and *deres*. This had the advantage that I only had to manually locate the antecedent, as the reflexive was al-

		tilstandsf	på	hver	sin	side	af	ligevægtspilen		
		fingerbøl	minder	i	sin	vækstfor	lidt	om	kongelys	
	disse	guld fisk	i	hver	sin	blender	er	skyld	i	årets første
	pludselig	halede	de	hver	sit	gevær	frem	og	begyndte	at skyde
	ligesom	da	vi	i	sin	tid	gik	i	skole	sammen
	10	guld fisk	i	hver	sin	blender	blev	guld fiske	auktioner	bort på
	og	kursiverier	der	med	sin	vægtning	af	ordene	skal	sige det
	at	hver	femte	i	sin	frustration	protester	til	højre	er ikke
	studievæ	der	med	hver	sit	særpræg	gav	en	god	og vedkomr
	det	hjalp	os	på	sin	vis	for	så	havde	vi det
	ting	som	du	i	sin	tid	havde	lavet	sammen	med dine
	men	da	brødrene	i	sin	tid	for	første	gang	satte sig
	før	havde	alle	hver	sin	opfattels	af	kvalitet	nu	bliver det
	de	partier	der	i	sin	tid	satte	et	mål	op for
	derefter	gik	vi	hver	sit	enkel	middag	i	sit	syrenlilla federal
	to	begreber	der	på	hver	måde	kan	opfattes	lige	abstrakte
	den	68	generatio	som	i	tid	grundlag	gælden	i	dag sidder
	samtidig	wil	cykelstier	på	hver	side	af	rådhuspl.	blive	forbedret betydelig
	købsprisen	for	vestas	aktierne	i	tid	er	ukendt		
	mens	de	syv	temadisk	til	tid	skal	give	partiledel	inspiratio til
	alle	tre	arbejder	på	hver	del	af	markedet	for	finansiel opbaknir
	lizzi	fortæller	at	de	i	tid	kun	havde	et	meget mindre

Figure 4.1: Screenshot from Excel of the automatic sorting with highlighting of *sin*

ready highlighted and lined up with other sentences with the same amount of words before the reflexive. Figure 4.1 is a screenshot of an Excel sorting file with *sin* highlighted and sentences sorted according to distance between *sin* and the first word in the sentence. In the subsequent manual sorting process, I would locate and highlight the plural antecedents in sentences with plural antecedent *sin* and delete sentences without plural antecedent *sin*.

At a later stage in the analysis process, after having coded the majority of the data, I belatedly realised that my code would not be able to include plural antecedent *sin* with antecedents consisting of two or more coordinated singulars, such as the example given in (190). The English translation here is slightly misleading, as the Danish equivalent of *narcotics*, *narko*, is non-count and formally functions as a singular noun (it takes singular adjective inflection, for instance: *Narko er farlig-t*, En. gloss *narcotics.SG is dangerous-SG.N*). I have tried to remedy this translation detail slightly by marking *narcotics* as singular in the gloss.

- (190) a. **Alkohol og narko₁** satte **sine₁** dybe spor i gruppens
 alcohol and narcotics.SG sat REFL's deep marks in group.DEF's
 medlemmer...
 members
 'Alcohol and narcotics made their deep marks in the members of the
 group.' (Korpus90)

I performed the search for coordinated singulars in the online version of KorpusDK with an extra requirement of a maximal distance of eight words between the second coordinate and the reflexive *sin*. This corresponds to the distance between antecedent and reflexive of 75 % (the third quartile) of my plural reflexive *sin* data and I judged that the benefit of including a longer distance would be very small: With a greater distance comes a much larger number of false positive results, and it is quite possible that at least one of the 9+ words between second coordinate and reflexive would be plural-marked, in which case the sentence would have been picked up by the original automatic search by accident. In short, I judge the amount of missed relevant sentences with *sin* with coordinated singular antecedents to likely be very small.

4.5 Data sets for comparison: plural antecedent *deres*, any antecedent *sin* (random 5 %)

One research question that I want answered is whether plural antecedent *sin* behaves distributionally and syntactically as plural antecedent *deres*, as *sin* in general, or as a thing of its own. A follow-up question is in which domains plural antecedent *sin* differs from or is similar to the other two categories. In order to answer this question, I sampled random selections of sentences with *deres* and *sin*.

4.5.1 Search strategy for plural antecedent *deres*

From the offline versions of Korpus90 and Korpus2000, I extracted all instances of sentences containing *deres*, using the same R script that I used for extracting all sentences containing *sin*. Call this file *deres_1*. Subsequently on the basis of *deres_1*, I automatically reduced the number of sentences to only those containing a *deres* preceded by a plural word. Call this file *deres_2*. I wrote a Python script to randomize and extract a specific proportion of sentences (available in appendix B as *python-corpus-randomisation.py*). With this script, I extracted a random 10 % of the *deres* sentences from *deres_2* to create the final file, *deres_final*. I sorted *deres_final* manually to remove all instances of unbound *deres*. I followed this procedure for Korpus90 and Korpus2000. From each of the two sorted files I extracted a random selection of sentences that corresponded proportionally to the amount of plural antecedent *sin* found for each of the two corpora: 500 *deres* sentences from

Korpus90 (corresponding to approximately 450 plural antecedent *sin* sentences) and 700 *deres* sentences (corresponding to approximately 650 plural antecedent *sin* sentences). The 1200 *deres* sentences were coded using the same categories as for plural antecedent *sin* (see section 4.7).

4.5.2 Search strategy for all *sin*

To create a basis for comparing plural antecedent *sin* with *sin* in general, I extracted a random selection of 500 sentences (250 from Korpus90 and 250 from Korpus2000) from the file containing every *sin* in the offline corpus. These sentences were coded with the same categories as plural antecedent *sin* and *deres*. The COORD and the distributive category are only relevant for the plural antecedent reflexives and were not filled in for the examples with singular antecedents.

4.5.3 Number validation for plural antecedent *sin*: Checking a random 5 % from each file

I took a random sample of 5 % of the sentences from each of the two full files file (the full file with all instances of *sin* from Korpus90, and the full file with all instances of *sin* from Korpus2000). I read through the sentences in these two files and marked all instances of plural antecedent *sin*, including *sin* with a plural antecedent that consists only of coordinated singulars. I found 22 examples of plural antecedent *sin* in the Korpus90 sample of 4207 sentences. I found 25 examples of plural antecedent *sin* in the Korpus2000 sample of 4070 sentences. I multiply those numbers by 20 because the sample represents 5 % of the full data. This amounts to 440 and 500. This approximation tells me that I can expect find around 440 plural antecedent *sin* examples in Korpus90 and around 500 plural antecedent *sin* examples in Korpus2000.

4.6 Inclusion and exclusion criteria for examples with plural antecedent *sin*

I manually sorted the final versions of the automatically sorted *sin* data sets in order to find the examples of plural antecedent *sin*. In the initial rougher manual sorting process, I removed all clear instances of singular antecedent *sin*. Approximately 5 MB

of singular antecedent *sin* examples were removed in this way for Korpus90 (29502-1742=27760 removed sentences for Korpus90, and 1742 potentially relevant ones that I retained) and 4.5 MB for Korpus2000 (corresponding to 26378-1618=24760 removed sentences for Korpus2000, and 1618 potentially relevant ones that I retained). This process was both extremely time-consuming and carried with it some risk of human error in e.g. missing a relevant sentence here and there. I had two safe-guards in place to minimize the risk of erroneously deleting a relevant example. The initial online search provided me with a subset of the plural antecedent *sin* examples. I could use these to compare with the offline data to check whether sentences were missing from the offline data that were included in the online data. The random number validation described in section 4.5.3 performed the same function with a different subset of the data.

The remaining 1742+1618 sentences were sorted manually, leaving 1217 examples of plural antecedent *sin*. This second manual sorting process was more fine-grained, meant for identifying and including all relevant examples of plural antecedent *sin* and excluding and removing those that should not count as relevant for the investigation. The following section outlines the exclusion criteria.

4.6.1 Exclusion criteria

All instances of fixed expressions that can only be used with *sin* are excluded. This includes, but is not limited to, expressions such as *i sin tid* (En. lit. *in REFL's time* meaning *at that time*), *sine steder* (En. lit. *REFL's places* meaning *some (particular) places*) and *(gå) hver til sit* (En. lit. *(go) each to REFL's* meaning *walk away from each other, e.g. to go home*)³.

Only examples that are actually fixed expressions are excluded. I differentiate between fixed expressions and non-fixed expressions that accidentally use the same sequence of words on the basis of (perceived) intonation, syntactic configuration and semantic content. In the fixed expressions, *sin* is typically emphasized through intonation and stress, *sin* is not bound, and the meaning of the expression is more than the sum of its parts, as illustrated in the attempted translations into English above. However, several of the fixed expressions mentioned can also appear as non-fixed expressions where *sin* is typically unstressed, *sin* is bound, and the meaning of the words

³Other expressions that I have treated as fixed are *svare enhver sit*, *på sin vis*, *i sin (skønneste) orden*, *på sin plads*

is closer to the sum of their parts. The contrast is shown in the examples in (191) that both contain *i sin tid*, but only the first one is a fixed expression.

(191) ***I sin tid* as a fixed expression**

- a. ... var vi i sin tid tilfredse med beviser...
 were we in REFL's time content with evidence.PL
 '... we were content with evidence at the time...' (Korpus90)

***I sin tid* as a non-fixed expression**

- b. I **sin**₁ tid som sygeplejerske havde **hun**₁ set masser af blod...
 in REFL's time as nurse had she seen lots of blood
 'In her time as a nurse she had seen lots of blood...' (Korpus90)

A number of expressions with *sin* seem semi-fixed, meaning that they have some of the characteristics of fixed expressions while allowing more than one type of pronoun. These are expressions like *sætte sit præg* (En. *leave REFL's mark*), *holde sin mund* (En. *hold REFL's mouth*, meaning *hold one's tongue*) and *give sit besyv med* (En. lit. *give REFL's opinion with*). These may all have e.g. *deres* substituted for *sin* and as such I include these semi-fixed examples as relevant data.⁴

I exclude the very frequent construction *hver sin* (*hver(t) sit*, *hver sine*) (En. *each REFL's*). Plural antecedent *sin* seems to be completely conventionalised in this context where *sin* simply only takes plural antecedents and is unambiguously distributive (Hansen and Heltoft 2011, 572). The reason for excluding *hver sin* is that not only does *sin* here require plural antecedents, it also allows first and second person plural antecedents which *sin* otherwise never does, (192).

- (192) a. **Vi**₁ bliver ensomme på **hver sin**₁ måde...
 we become lonely on each REFL's way
 'We each become lonely in our own way...' (Korpus90)
- b. ... at **I**₁ opretholder **hver sin**₁ bolig.
 that you.PL retain each REFL's home
 '... that you each retain your own home' (Korpus90)
- c. **De**₁ har **hver sin**₁ hjerne...
 they have each REFL's brain
 'They each have their own brain...' (Korpus90)

⁴I include the following in the data set as semi-fixed expressions: (*sætte*) *sit præg*, (*sætte*) *sit/sine spor*, *i sin vorden*, *holde sin mund*, *give sit besyv med*, (*tale*) *sit tydelige/tungtvejende/eget sprog*, *kræve sit*, *i al sin enkel(t)hed*, *være på sit højeste*, *have sin pris*, *stå sin prøve*, *komme til sin ret*, *på sin side*, *sig og sine*, *tage sin tid*, *gøre sit*.

Table 4.6: *Hver sin/deres/vores/jeres* in KorpusDK (each REFL's/their/our/your)

hver vores	89
hver jeres	10
hver deres	530
hver(t) sin/sit/sine	1934

The addition of *hver* changes the antecedent options so much that it would greatly skew the results of the investigation if they were to be included. Additionally, *hver sin* is a reasonably frequent construction and occurs one and a half times as often in KorpusDK as plural antecedent *sin* does. Table 4.6 reports the frequency of *hver vores* (each our), *hver jeres* (each your), *hver deres* (each their) and *hver sin* in KorpusDK. *Hver vores*, *hver jeres* and *hver deres* only occur with antecedents of the corresponding person (i.e. first person for *hver vores*, second person for *hver jeres*, and third person for *hver deres*). *Hver sin* occurs with antecedents of all three persons but most frequently with third person antecedents.

Examples where the antecedent is an object (*object binding*) are excluded. Eight examples were excluded from the final data set for this reason. (193a) is ambiguous: Either the antecedent is the singular subject *man*, in which case the sentence is simply *sin* with a singular antecedent and should be left out, or the antecedent is the plural indirect object *nye indfaldsvinkler* and should be left out. 193b is a straight-forward example of object binding and is excluded.

- (193) a. ... og **man**_{1?} må som ansat give **nye indfaldsvinkler**_{1?} **sin**₁
 and one must as employee give new approaches REFL's
 chance.
 chance
 '... and one must as an employee give new approaches their chance.'
 (KorpusDK)
- b. Den skandaleramte olympiske komite, IOC, har ikke kunnet
 the scandal-ridden olympic committee IOC has not could
 fratage **de olympiske lege**₁ **sin**₁ popularitet...
 deprive the olympic games REFL's popularity
 'The scandal-ridden Olympic committee, IOC, has not been be able to
 deprive the Olympic Games of their popularity...' (KorpusDK)

Examples that are ambiguous due to lack of overt number marking are excluded. This is the case for *hjul* (En. *wheels*) in (194) where neither context nor the word itself disambiguates its number.

- (194) a. ... men i fritiden er det **TV 2-konkurrentens lette hjul**₁,
 but in free-time.DEF is it TV 2-competitor's light wheel.SG/PL?
 der drejer DRs gamle sprog-røgter om **sine**₁ farverige
 that turn.SG/PL? DR's old language-tender about REFL's colourful
 eger.
 spokes
 '... but in the spare time it is the light wheel(s) of the competitor TV 2
 who turn(s) DR's old language-tender about its/their colourful spokes.'
 (KorpusDK)

Examples that are ambiguous between having an arbitrary PRO antecedent or a plural antecedent are excluded. The subject of the embedded clause *at tilgive sine fjender* in (195a) is non-overt PRO. This PRO could be co-indexed with the matrix subject *de* in which case the sentence is about some specific people's forgiveness of their own specific enemies. *Sin* would in this case be bound by PRO and in turn by the matrix subject *de*, which would mean that the sentence would be a relevant example of plural antecedent *sin*. The reference of the non-overt PRO could also be arbitrary in which case the sentence is about a more abstract discussion about forgiveness. In this case *sin* would be bound by PRO but not by the matrix subject *de*, and the sentence would consequently not be an example of *sin* with a plural antecedent. It is not completely clear which reading that is intended and I exclude the sentence for that reason (and sentences similar to it).

- (195) a. ... og **de**₁ taler om en international indbyrdes afhængighed og
 and they talk about an international mutual dependence and
 om at tilgive **sine**₁ fjender.
 about to forgive REFL's enemies
 '... and they talk about an international interdependence and about for-
 giving their enemies.' (KorpusDK)

The data set contains several examples of sentences that contain a partitive directly followed by a relative clause. They are typically of the type "en af dem der..." (En. *one of those who...*) or "en af de X der..." (En. *one of the X's who...*) where X is a plural nominal. The relative clause can either be interpreted as modifying the full partitive phrase, (196a), or it can be interpreted as modifying only the final plural nominal, (196b).

- (196) a. [DP en af dem]_[CP der...]
 b. [DP en af [DP dem]_[CP der...]]

Two real examples of the kind can be seen in (197). Both have been excluded from the data set because of their potential for ambiguity, and so have a number of similar sentences with the same structure.

- (197) a. Bjørn fik fat i en af dem₁, der henviste til sin₁
 Bjørn got hold in one of them/those who referred to REFL's
 overordnede, som beredvilligt demonstrerede et lille, men brugbart
 superior who readily demonstrated a small but useful
 ordforråd på engelsk...
 vocabulary in English
 'Bjørn got a hold of one of them/those, who referred to his/her superior,
 who readily demonstrated a small but useful vocabulary in English...'
 (KorpusDK)
- b. Armando er en af New Yorks barske cab-drivers₁, der var
 Armando is one of New York's tough cab-drivers who was/were
 ved at miste livet på sit₁ job.
 about to lose life.DEF on REFL's job
 'Armando is one of New York's tough cab drivers, who was/were about
 to lose his/their life on his/their job.' (KorpusDK)

(197a) contains the partitive *en af dem* and the relative clause *der henviste til sin overordnede som...* In the reading where the relative clause modifies the full partitive, (197a) means that Bjørn finds one person who then points him to the superior. In this case the antecedent of *sin* would be *en af dem*, i.e. not a relevant example of *sin* with a plural antecedent. In the, probably less likely, reading where the relative clause modifies only the final plural *dem*, (197a) means that Bjørn finds a number of people who point him to their superior and that Bjørn chooses to talk to one of them. This second reading would be a relevant example of *sin* with plural antecedent *dem*.

There are also two possible readings of (197b). The partitive *en af New Yorks barske cab-drivers* can be modified by the relative clause *der var ved at miste livet på sit job*. In this case only one cab driver is close to risking his life, and the *sin* in the relative clause has a singular antecedent. The relative clause could also modify only *New Yorks barske cab-drivers* in which case the sentence means that many (or perhaps all) cab drivers are risking their life. This would mean that *sin* in the relative clause has a plural antecedent in *New Yorks barske cab-drivers*. These and similar examples are discussed further in section 4.8.3.

Examples where the plural antecedent is the name of an organisation as in (198a), a movie as in the movie "Tegninger" (En. "Drawings") in (198b), a book, or similar

items, are excluded. I do include examples where the name of the organisation is used to refer to several members of the given organisation. An example of this could be sentences where the plural antecedent is the name of a political party such as *Socialdemokraterne* (En. *The Social Democrats*).

- (198) a. Sjette Frederiks kro i Risskov, der hvor **Musiske Studenter**₁ havde holdt **sine**₁ særlige fester.
held REFL's special parties
'Sjette Frederik's inn in Risskov, the place where Musical Students had thrown its/their special parties.' (KorpusDK)
- b. "**Tegninger**"₁ vokser i **sin**₁ enkelthed fra barnets mikrokosmos til at fylde et større rum...
micro-cosmos to to fill a bigger space
'"Drawings" grows in its simplicity from the child's micro cosmos to filling a bigger space.' (KorpusDK)

Examples with *enhver*, *ingen*, *hver* as antecedents are excluded. The variation in agreement with these is discussed further in section 2.5.4 from page 109 and onwards.

- (199) a. ... og de blev dømt, **enhver**₁ efter **sine**₁ gerninger.
and they became judged each after REFL's deeds
'... and they were judged, each after his/her deeds.' (KorpusDK)

In a wide range of example types, the reflexive is contained within a modifying phrase where it is unclear whether the antecedent is a plural or a singular nominal. This is similar to the example types in (197) because of the same mechanism, i.e. ambiguity in which phrase the PP modifies. Some specific examples of this are given in (200) and these and similar sentences are all excluded due to this ambiguity.

- (200) a. ... der nu tog revanche mod **Palle Günther og Ambition Express**₁, som i **sine**₁ forrige starter har sænket Juuls favoritter.
who now took revenge against Palle Günther and Ambition Express who in REFL's previous starts has sunk Juul's favourites
'... who now took revenge against Palle Günther and Ambition Express who in its/their previous races has/have sunk Juul's favourites.' (KorpusDK)
- b. ... sådan som både **Socialdemokratiet** og **Regeringen**₁ i **sine**₁ finansredegørelser forfægter.
just like both Social-Democratic-Party.DEF and Government.DEF in REFL's economic-statements assert

'... just like the Social Democratic Party and the Government assert(s) in its/their economic statements.' (KorpusDK)

Duplicate examples are excluded, and in the cases where a sentence contains more than one plural antecedent *sin*, only the first *sin* (in linear order) is recorded and coded.

The last types of excluded sentence type are more varied and less frequent. These include coordination structures where two epithets describe the same person or entity, (201a); a specific type of time antecedents which are structurally plural but must be read as one entity, (201b) and (201c); and finally sentences with potentially misleading textual omissions built into the corpus such as (201d) where the [...] indicates places where a number of words have been taken out of the corpus, perhaps for reasons of anonymity.

- (201) a. Samlet forventer **banken og realkreditte**₁ at tjene altogether expect bank.DEF and mortgage-provider.DEF to earn mellem 2,1 og 2,3 mia. kr. på **sin**₁ basisforretning i hele år between 2.1 and 2.3 bill. kr. on REFL's main-business in all year 2000...
2000
'Altogether the bank and mortgage provider expect(s) to earn between 2.1 and 2.3 bill. kr. on its/their main business in all of 2000...' (KorpusDK)
- b. Samtidigt har **18 år med det samme parti ved magten**₁ gjort meanwhile have 18 years with the same party at power.DEF done **sit**₁ ved de britiske vælgere.
REFL's to the British voters
'At the same time, 18 years with the same party in power has/have done its/their (part) to the British voters.' (KorpusDK)
- c. ... og **to måneder på feltration**₁ sætter **sit**₁ præg på and two months on field-rations set REFL's mark on samtalerne.
conversations.DEF
'... and two months on field rations leave its/their mark on the conversations.' (KorpusDK)
- d. Der er to linier: dem der prøver at opnå magt [...] og **dem**₁ there are two lines those who try to obtain power [...] and those der vil bibeholde den [...] eller **sin**₁ magt ved hjælp af den... who will retain it [...] or REFL's power by help of it

'There are two lines: those who try to obtain power [...] and those who want to retain it [...] or his/their power by means of it...' (KorpusDK)

4.7 Coding

Table 4.7 provides an overview of the categories used in coding the corpus data. The first column, *Term*, is the name of the coding category I use in the dataset itself. The second column, *Values*, gives the possible values for the category in question. The third column, *Explanation*, is a more explicit explanation of the meaning of the shorthand in the first column. I discuss the coding and the specific choices made in more detail in the rest of this section.

Starting from the top, the category `semantic_antecedent` picks out the lexical word that provides the semantic content of the antecedent. This will not necessarily be the same as the syntactic antecedent, which I define solely in structural terms. The semantic antecedent must still be in a position that licenses it being the antecedent, i.e. it must be part of a chain of nominal elements that c-command the reflexive and where the reflexive is the lowest element. The semantic antecedent is the highest element in this chain, the one that in the terms of Reinhart and Reuland (1993) must be fully referential.

The `syntactic_antecedent` is the lowest element in this potential chain of co-referent elements. I illustrate the difference between `semantic_antecedent` and `syntactic_antecedent` with the example in (202).

- (202) Mange **kommuner**₁ benytter sig af [**PRO**₁ at sende **sine**₁ dårlige
 many municipalities use REFL of PRO to send REFL's bad
 betalere til inkasso,]
 payers to debt-collection
 'many municipalities make use of debt collectors to take care of their bad pay-
 ers'

Mange kommuner is the semantic antecedent. The embedded non-overt PRO subject in the infinitival clause is the syntactic antecedent, the lowest element in the antecedent chain.

The syntactic antecedent can be the PRO subject of an infinitival clause as in (202). The syntactic antecedent can also be the trace of the moved subject of a relative clause if the reflexive is embedded in a relative, as in (203).

Table 4.7: Coding of corpus data

Term	Values	Explanation
semantic_antecedent	various	the lexical word that <i>sin</i> gets its reference from
syntactic_antecedent	PRO	the syntactic placeholder for the antecedent
	relative pronoun no placeholder	
antecedent_is_pronoun	yes	is the antecedent a pronoun without noun complement?
	no	
subj_in_coord	no	is the antecedent a coordination structure?
	yes, only sg. items in COORD	
	yes, only pl. items in COORD	
	yes, sg. and pl. items in COORD	
animate	yes	antecedent is an animate noun
reflexive	no	
	sin	what is the form of the reflexive?
	sit	
	sine	
object	various	the complement of the reflexive
	local	local or non-local binding?
num_of_object	non-local	
	singular	number of object
	plural	
type_of_embedded	infinitive	type of embedded clause that the reflexive is contained in
	relative	
	not embedded	
complement/adjunct	complement	is the reflexive in a complement or modifier relation with the verb?
	adjunct	is the reflexive the complement of a preposition?
in_PP	yes	
	no	
verb	various	the main verb in the clause containing the reflexive
	yes	is the main verb a distributive verb?
verb_distributive?	no	
	yes	is there a finite verb in the clause that contains the reflexive?
verb_complex	yes	any auxiliary verbs in the clause that contains the reflexive?
	no	
intervening_singular	yes	do any singular nominals intervene between reflexive and antecedent?
	no	

- (203) De østjyske **musikere**₁ [_{CP} _{ec}₁ **der** gennem årene t₁ har givet
 the east-Jutlandic musicians ec REL through years.DEF t have given
sit₁ stampublikum mange sjove oplevelser]...
 REFL's regulars many fun experiences
 'the east-Jutlandic musicians who through the years have given their regulars
 many fun experiences'

The lowest trace is the trace of the moved subject which is moved out of the relative clause by way of an operator in CP-spec (see e.g. Vikner (1991, 125), or alternatively the highest projection in the relative clause is cP as in Nyvad, Christensen, and Vikner (2017) but the point about extraction by way of an operator in the highest projection remains the same). Finally, the semantic and the syntactic antecedent can be the same entity, as in (204) where *lydene* is both the semantic and the syntactic antecedent. I count both pronouns and larger DPs as possible semantic antecedents.

- (204) **lydene**₁ havde **sin**₁ forklaring i et gammelt hus
 sounds.DEF had REFL's explanation in an old house
 'there was a reason for the sounds in an old house'

The category *antecedent_is_pronoun* codes for whether the (semantic) antecedent is a bare pronoun or a full DP. For the purpose of analyzing the data on plural antecedent *sin*, I count as pronouns the plural personal pronouns *de* and *dem*, the universal quantifiers *alle* (En. *all*) and *begge* (En. *both*), as well as the indefinite plural *nogle*. I exclude *nogen* (En. *any/some*), *hver* (En. *each*), and *ingen* (En. *no one*) as possible plural antecedents. This topic is discussed in more detail in section 2.5.4 but I sum up the main points here, as well. *Nogen* is the common gender singular form that corresponds to *nogle*. In the spoken language, the most common pronunciation of *nogle* is exactly the same as the pronunciation of *nogen* (possibly differentiated through inflection, as argued in Hansen and Heltoft (2011, 584)), and some written instances of *nogen* could be intended as *nogle*. To be on the safe side, all instances of *nogen* are excluded. Hansen and Heltoft (2011, 571) describes *hver* as neutral in regard to number and *ingen* is precisely the zero sum, which in a certain sense could also be analyzed as a neutrality to number. *Ingen* may occur as the antecedent of either *sin* or *deres*, (205a)-(205b), and with either singular or plural inflection on a predicative adjective, (205c)-(205d). Due to this apparently frequent optionality, sentences with antecedents *hver* and *ingen* are excluded from the data.

- (205) a. **Ingen** undgår **sin**₁ skæbne.
 noone avoids REFL's destiny

'No one avoids his/her destiny.'

- b. **Ingen** gav **deres**₁ liv forgæves.
 noone gave their life.SG/PL? in vain
 'No one gave their life/lives in vain.'
- c. Ingen er perfekt-Ø, heller ikke en international skønhedsdronning...
 noone is perfect-DEF, also not an international beauty-queen
 'No one is perfect, not even an international beauty queen...'
- d. Ingen er skræmt-e.
 noone is scared.PL
 'No one is scared.'

I follow Hansen and Heltoft (2011, 565) in viewing *andre* and *mange* not as pronouns but rather as adjectives and they are coded as full nouns in the data set in order to differentiate them from the pronoun category. Hansen and Heltoft argue that a reason to not group *andre* and *mange* with the quantitative pronouns is that they may co-occur with other D-elements such as definite articles, which is never the case for the other members of the pronoun category. The contrastive examples in (206) are from Hansen and Heltoft (2011, 565) but slightly modified and expanded.

- (206) a. *de alle deltagere
 the all participants
 'the all participants'
- b. *de begge deltagere
 the both participants
 'the both participants'
- c. de mange deltagere
 the many participants
 'the many participants'
- d. de andre deltagere
 the other participants
 'the other participants'

The category `subj_in_coord` codes for whether the antecedent is a coordinated structure, and if it is coordinated, which elements the coordinated structure consists of. The coordination may involve two (or more) singular entities, two (or more) plural entities or two (or more) entities where one is plural and the other is singular. I have not coded for the order within the coordinated structures that contain both singular and plural elements.

The category `reflexive` simply denotes the form of the reflexive, which in turn depends on the number and gender of the complement of the reflexive. The reflexive is *sin* when its complement is singular common gender. The reflexive is *sit* when its complement is singular neuter gender. The reflexive is *sine* when its complement is plural.

The category `object` is the nominal complement of the reflexive. *Sin* combines with the same range of nominals as determiners do, and its complement may even be empty (in the case of ellipsis).

Local binding is by far the most frequent type of binding for *sin* in general, with non-local binding occurring in as low as 0.9 % and as high as 31 % of the cases investigated, where both local and non-local binding are possible. See Vikner and Ehlers (2017) for specific data on the distribution of local and non-local *sin*. This means that overall, for every instance of *sin* and not just *sin* in contexts that allow both local and non-local binding, non-local binding of *sin* can be expected to be very infrequent. The category `type_of_binding` codes for whether *sin* is bound locally (by far the most frequent) or non-locally (almost non-existent in the data).

The category `num_of_object` is a binary that codes for whether the complement of *sin* is singular or plural.

The category `type_of_embedded` codes the type of embedded clause that the reflexive is contained within (provided that the reflexive is embedded at all). There should be a full overlap in this category with the category of syntactic antecedent, i.e. a PRO antecedent should entail an infinitive embedded clause and a relative pronoun antecedent should entail a relative embedded clause, so this is essentially a double-checking category that could easily be dispensed with in a similar study.

The category `complement/adjunct` codes whether the reflexive is contained within a complement or an adjunct relative to the main verb of the clause.

The category `in_PP` is a binary that indicates whether or not the reflexive is the complement of a preposition. Tingsell (2013) finds much more variation in the scope

of reflexive use if the pronoun is embedded in a PP and I wanted to investigate if that is also the case with this data.

The category `verb` marks the main verb in the clause that the reflexive is contained within. The intention of this is to investigate whether certain verbs or types of verbs are more likely to occur with *sin* with a plural antecedent.

It has been suggested (see section 2.5.5 on page 117) that a distributive reading makes it more likely that *sin* will be acceptable with plural antecedents. The category `verb_distributive?` is an attempt to divide the sentences into whether or not the plural antecedent *sin* is used in a distributive reading. The coding of this category has to be somewhat subjective since it is not always obvious if the meaning of the sentence is intended to be distributive or not. Furthermore, the same verb can be used both in a distributive and collective meaning, depending on the wider context. The two sentences in (207) exemplify how the same verb, *pynte* (En. *decorate*), may be used in both a distributive and a collective sense, only dependent on the complement of the reflexive. (207a) is collective (the scouts own the cabin as a group), (207b) is distributive (the scouts each own a uniform which they each decorate). Both sentences are made up to illustrate how distributivity depends on the entire context of the sentence.

(207) a. **Spejderne**₁ pynter **sin**₁ hytte med pels.
 scouts.DEF decorate REFL's cabin with fur
 'The scouts decorate their cabin with fur'

b. **Spejderne**₁ pynter **sin**₁ uniform med pels.
 scouts.DEF decorate REFL's uniform with fur
 'The scouts decorate their uniforms with fur'

`verb_finite` and `verb_complex` are two more verb-based categories, both purely structural. `verb_finite` indicates whether the clause that contains *sin* contains a finite verb. `verb_complex` is positive if the clause contains one or more auxiliary verbs.

The last category, `intervening_singular`, codes for the presence of a nominal with singular number between the reflexive and its antecedent. It may be the case that speakers are more likely to use *sin* with a plural antecedent if the sentence also contains a singular element, just as intervening elements with contrasting grammatical number have been noted to interfere with verbal or adjectival agreement (see e.g. Christensen and Nyvad 2019).

4.8 Results

The following sections explore the research questions below in the eight separate sections in section 4.8.1 to section 4.8.8. In each section I outline a syntactic or semantic factor that the data is coded for, provide relevant examples from KorpusDK, and finally present the frequencies of occurrence in KorpusDK. I compare the numbers for plural antecedent *sin* with similar numbers for plural antecedent *deres* and for *sin* with all kinds of antecedents. It is important to note that the numbers are not comparable in magnitude but only as relative distributions. The 1218 examples of plural antecedent *sin* represents the entire selection of plural antecedent *sin* that I could find in KorpusDK. The numbers for *sin* in general and for bound *deres* represent only a small sample of the *sin* and *deres* in KorpusDK. There are 188585 examples of *sin* in KorpusDK all in all, which means that plural antecedent *sin* make up 0.65 % of the full sample of *sin*. Similarly, there are 89232 examples of *deres* in KorpusDK (not all bound, should be noted). The graphs that are shown, then, do not say that there are as many instances of plural antecedent *sin* in the corpus as there are instances of plural antecedent *deres*. They only show the relative distribution of each bound form within each of the coded factors. The last section of the chapter, section 4.9, is a discussion and a summary of the results.

1. How frequently do instances of plural antecedent *sin* occur in the corpus?
2. Do the instances of plural antecedent *sin* pattern similarly to regular instances of plural antecedent *deres* or *sin* with any kind of antecedent?
3. Which syntactic factors favour plural antecedent *sin*?
4. Which semantic factors favour plural antecedent *sin*?

4.8.1 Type of subject (full noun or pronoun)

The antecedents in the data set are coded according to which type of nominal that they consist of. I differentiate between subjects that are pronouns (and universal quantifiers) such as *de*, *begge*, and *dem* in (208a)-(208c) and full noun subjects such as *pigerne* in (208d) and *kortenes motiver* in (208e).

- (208) a. ... og **de**₁ har derfor fået **sin**₁ egen side.
 and they have therefore got REFL's own page
 '... and they got their own page accordingly.'

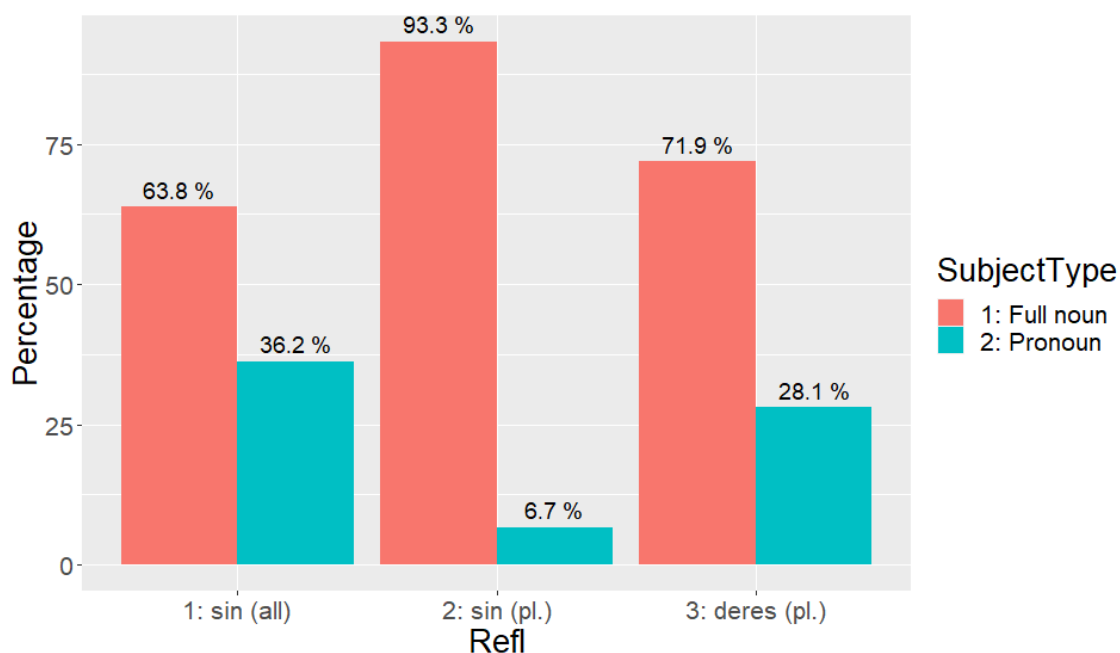


Figure 4.2: Subject: Full noun or pronoun

- b. ... og **begge**₁ så **sin**₁ egen frygt genspejlet i den andens øjne.
 and both saw REFL's own fear mirrored in the other's eyes
 '... and both saw their own fear mirrored in the other's eyes.'
- c. For han er blandt **dem**₁ der i dag holder **sin**₁ mund.
 for he is among them who in day hold REFL's mouth
 'For he is one of those who hold their tongue today.'
- d. **Pigerne**₁ har vundet **sine**₁ to hidtidige udekampe i
 girls.DEF have won REFL's two previous away-games in
 turneringen.
 tournament.DEF
 'The girls have won their two previous away games in the tournament.'
- e. **Kortenes motiver**₁ udtrykker med **sine**₁ kønne og varme farver...
 cards.DEF' motifs express with REFL's pretty and warm colours...
 'The motifs of the card express with their pretty and warm colours...'

Figure 4.2 plots the distribution of full noun antecedents against pronoun antecedents across all three data groups: *sin* with any type of antecedent, *sin* with plural antecedent and reflexive *deres* with plural antecedent. All antecedents *sin* and plural antecedent *deres* are reasonably similar: approximately two thirds of the examples (between 63.8 % and 71.9 %) have a full noun antecedent and the last third of the examples have pronoun antecedents. Plural antecedent *sin* patterns very differently with 93.3 % full noun antecedents and 6.7 % pronoun antecedents. This means that

plural antecedent *sin* overwhelmingly occurs in contexts where the plural antecedent is a full noun. The big question, then, is why this would be the case. Pronouns typically denote previously introduced entities while full nouns may introduce new material to the discourse. Perhaps it could feel more relevant for a writer to underscore the reflexive possessive relationship with a new entity (by using *sin* rather than *deres*) than with one already established.

4.8.2 Number of the nominal that contains the reflexive

An early intuition about the occurrence of *sin* with plural antecedents was that its use seems more acceptable when *sin* itself is part of a plural nominal phrase. The corpus examples in (209) show various examples of plural antecedent *sin* contained within singular nominal phrases.

- (209) a. **Alt for mange**₁ undlader at pleje **sin**₁ krop...
 all too many fail to tend REFL's body
 'Far too many fail to care for their body.'
- b. ... så **hundeejere**₁ kan tage **sin**₁ hund med...
 so do-owners can take REFL's dog with
 '... so dog owners can bring their dog..'
- c. ... så vi tror at **nordjyderne**₁ med opbakning fra **sit**₁
 so we think that north-Jutlanders.DEF with support from REFL's
 store publikum klarer opgaven
 large audience manage task.DEF
 '... so we think that the north Jutlanders will manage the task with support
 from their large audience.'
- d. **Bombardementerne**₁ går ind i **sit**₁ sjette døgn
 bombardments.DEF go in in REFL's sixth day
 'The bombardments enter their sixth day.'

The corpus examples in (210) show various examples of plural antecedent *sin* contained within plural nominal phrases.

- (210) a. ... er **skuespillerne**₁ tvunget til at indtage **sine**₁ måltider på offentlige
 are actors.DEF forced to to eat REFL's meals on public
 restauranter...
 restaurants
 '... are the actors forced to eat their meals in public restaurants.'
- b. Mens **de øvrige syv regioner**₁ skal lave **sine**₁ daglige 25 minutters
 while the other seven regions must make REFL's daily 25 minutes
 fjernsyn året rundt...
 television year.DEF around
 'While the other seven regions must make their 25 minutes of television all
 year round...'
- c. **Alle krige**₁ har **sine**₁ årsager og konsekvenser.
 all wars have REFL's causes and consequences
 'All wars have their causes and consequences.'
- d. **Anders Westers forældre**₁ har netop udlejet **sine**₁ herlige
 Anders Westers parents have just rented REFL's delightful
 træningsfaciliteter...
 exercise-facilities
 'Anders Wester's parents have just rented out their delightful exercise facil-
 ities.'
- e. Det er **de nye tanker**₁ der larmer og varsler **sine**₁ indtog i min
 it is the new thoughts that roar and predict REFL's entries in my
 indre bevidsthedscomputer.
 inner consciousness-computer
 'It is the new thoughts that make noise and predict their entries in my inner
 consciousness computer.'

The two examples in (211) show that *deres* may occur as a locally bound pronoun within singular as well as plural nominals.

- (211) a. **Alkoholikerne**₁ har **deres**₁ eget territorium på pladsen.
 alcoholics.DEF have their own territory on square.DEF
 'The alcoholics have their own territory on the square.'
- b. **Mange forældre**₁ har svært ved at bede **deres**₁ børn om penge.
 many parents have difficult by to ask their children for money

Table 4.8: *Sin* and *deres* in singular and plural DPs in KorpusDK

KorpusDK	<i>sin</i>	(pl. antecedent)	<i>deres</i>	(reflexive)
In singular DP	807	66.3 %	599	52.0 %
In plural DP	411	33.7 %	491	42.6 %
Unclear # (only <i>deres</i>)			62	5.4 %
All	1218	100 %	1152	100 %

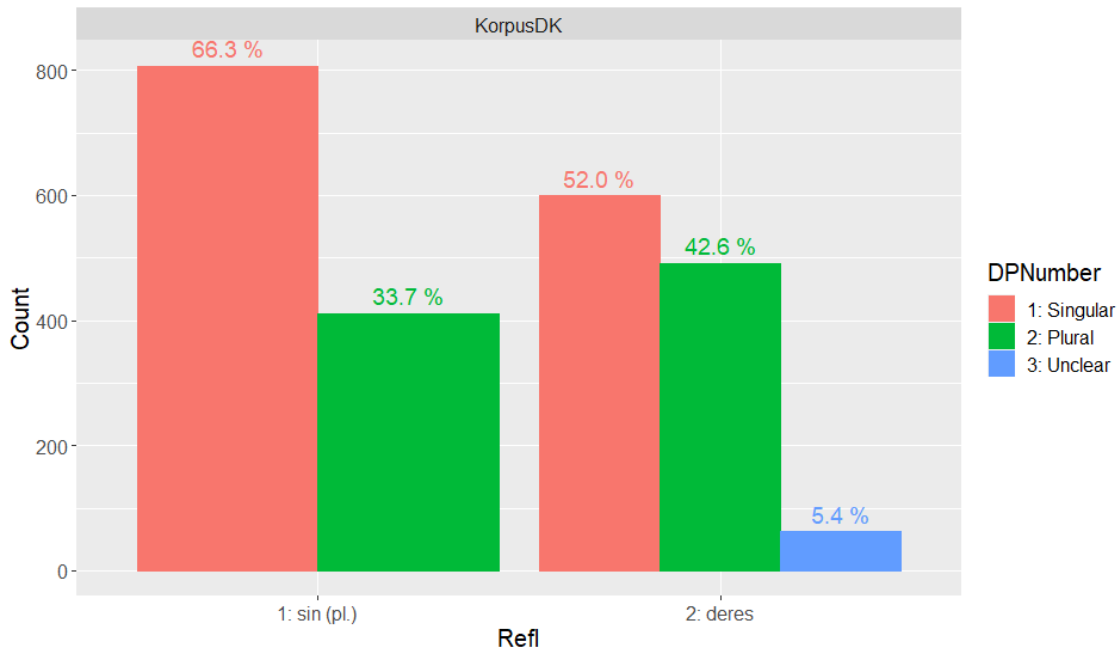


Figure 4.3: Reflexives in singular and plural DPs in KorpusDK

'Many parents have a hard time asking their children for money.'

Informal acceptability judgments from linguist colleagues and students tend to agree that the sentences in (209) are worse than the sentences in (210) (and the sentences in (211) are fine no matter what). The question is whether the corpus data back up the intuition that plural antecedent *sin* is better within a plural nominal than within a singular nominal. When I write *sin*, I refer to all three inflectional forms *sin/sit/sine*. Technically, then, the wording should be "the intuition that *sine* with a plural antecedent is better than *sin/sit* with a plural antecedent".

At a first glance, this hypothesis seems to not be borne out in the corpus. The raw occurrence numbers are shown in table 4.8. The numbers are plotted as bars in fig. 4.3.

Table 4.8 and fig. 4.3 indicate that plural antecedent *sin* and locally bound *deres*

Table 4.9: Singular and plural complements in KorpusDK

KorpusDK	Verb ^a		Preposition ^b		All words ^c	
Singular comp	2,062,282	73 %	3,146,994	74 %	13,315,939	72 %
Plural comp	760,136	27 %	1,114,670	26 %	5,253,801	28 %
All	2,822,418	100 %	4,261,664	100 %	18,569,740	100 %

^a Search strings:

Verb followed by singular: [pos="V"][morph=".*_S_.*"]

Verb followed by plural: [pos="V"][morph=".*_P_.*"]

^b Search strings:

Preposition followed by singular: [pos="PRP"][morph=".*_S_.*"]

Preposition followed by plural: [pos="PRP"][morph=".*_P_.*"]

^c All tagged words:

Singular: [morph=".*_S_.*"]

Plural: [morph=".*_P_.*"]

pattern fairly differently. Plural antecedent *sin*, perhaps counter to the initial intuition, occurs almost twice as frequently in singular DPs (66.3 %) as in plural DPs (33.7 %). *Deres* occurs in a slightly higher proportion of singular DPs (52.0 %) than plural DPs (42.6 %).

In 5.4 % of the occurrences of *deres*, the number of the DP is unclear. This is the case with nominals that are invariant in singular and plural, such as *får* (En. *sheep*) or *krav* (En. *demand(s)*), and where the immediate context does not offer any clarifying clues. This never occurs with *sin*, since *sin* itself inflects according to the number of the DP and consequently disambiguates the number of the nominal. *Deres* is invariant in form across all uses, reflexive or not, plural DP or not.

Plural antecedent *sin* occurs most often in singular DPs but it could still be the case that plural antecedent *sin* is favoured when it occurs in plural DPs. Singular complements (of verbs and prepositions) are far more frequent than plural complements in the corpus in general. Table 4.9 shows the distribution of plural and singular words in KorpusDK in three specific contexts. The data are from searches in KorpusDK specifically for verbs directly followed by singular/plural words, for prepositions directly followed singular/plural words, and for singular/plural words in general. This is necessarily a rough estimate as it leaves out all cases where the verb or preposition is linearly separated from its complement, and it wrongly includes cases where the verb or preposition does not actually have a relation to the following word. A skim of the first few pages for each search shows very few errors, however.

Table 4.10: Singular and plural complements in KorpusDK: with *sin* and *deres*

KorpusDK	<i>sin</i>	(all antecedents)	<i>sin</i>	(pl. antecedents)	<i>deres</i>	(pl. antecedents)
Singular comp	400	80 %	807	66.3 %	599	52.0 %
Plural comp	100	20 %	411	33.7 %	491	42.6 %

In all three conditions (complements of verbs, complements of prepositions, all words tagged as singular or plural), singular words are much more frequent (72-74 %) than plural words (26-28 %).

The next question, then, is whether plural antecedent *sin* occurs in plural nominals more often than the overall relative frequency of singular and plural words in the corpus in general would bring about. In table 4.10, I compare the distribution of *sin* with all antecedents with plural antecedent *sin* and locally bound *deres*.

For *sin* with all types of antecedents, 80 % of the instances occur in a singular DP and 20 % in a plural. Compared to the general distribution of verbal and prepositional complements and all number-tagged words in KorpusDK where 72-74 % of words are singular, this means that *sin* occurs in singular DPs at a rate of 6-8 percentage points higher than the general number distribution in the corpus. It seems that *sin* is in fact favoured in singular DPs, or more specifically that *sin* and *sit* together occur much more frequently than *sine*, and even more so than what could be expected from the distribution of singulars and plurals in the corpus in general.

The distribution of *sin* in general can be compared to the distribution of plural antecedent *sin* and plural antecedent *deres*. The occurrence numbers are repeated from table 4.8 to table 4.10 to compare them directly with all antecedents *sin*. Plural antecedent *sin* does occur more frequently in singular DPs, 66.3 %, than in plural DPs, 33.7 %. However, it turns out that plural antecedent *sin* does seem to be comparatively favoured in plural DPs where it occurs 13-14 percentage points more often than *sin* in general and 5-7 percentage points more often than the singular/plural distribution in the corpus in general (where 26-28 % of the number-tagged words are plural). In short, it seems that *sin* in general strongly favours singular DPs and that plural antecedent *sin* by comparison can indeed be said to favour plural DPs, in line with the original intuition, even if plural antecedent *sin* does occur more frequently in singular DPs. Finally, it seems that locally bound *deres* strongly favours plural environments, both compared to plural antecedent *sin* and to the corpus in general. At least 42.6 % of the instances of locally bound *deres* (9 percentage points more than the 33.7 % of plu-

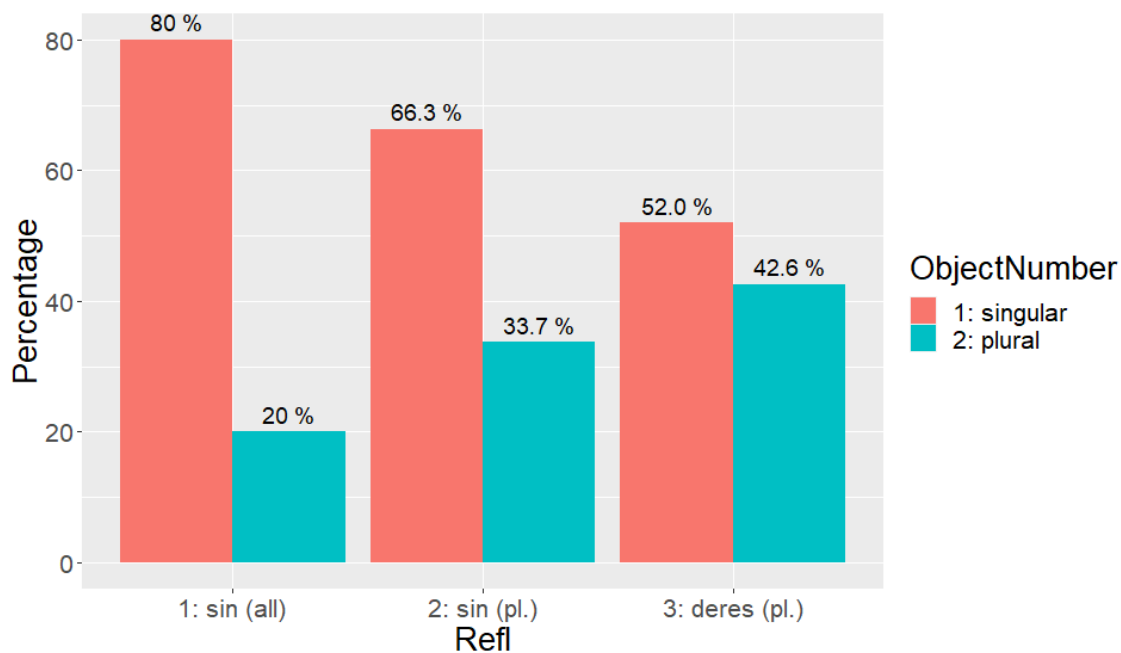


Figure 4.4: Relative occurrence of all *sin*, plural antecedent *sin* and plural antecedent *deres* by number of the DP

ral antecedent *sin*) occur in plural DPs. Plural antecedent *sin* seems to fall somewhere in the middle between all antecedents *sin* and locally bound *deres*: It is not as strongly favoured in plural DPs as *deres*, but still more so than *sin* in general. The percentages for all antecedents *sin*, plural antecedent *sin*, and locally *deres* are plotted side by side in fig. 4.4 for visual comparison. The columns show the frequency of occurrence of all antecedents *sin*, plural antecedent *sin* and plural antecedent *deres* (unclear cases are not included, which is why the *deres* numbers do not add up to 100 %).

4.8.3 Sentence type (simple or complex)

The corpus data is coded according to sentence type as simple or complex. Simple sentences are the ones where the topmost element in the reflexive chain (the most referential antecedent head) and the reflexive pronoun occur within the same minimal clause. Examples of this are given in (212).

- (212) a. ... det var som om **hans hænder**₁ levede **sit**₁ eget liv...
 it was as though his hands lived REFL's own life
 '... It was as though his hands lived their own life...'

- b. ... før **andre**₁ har sagt **sin**₁ mening.
 before others have said REFL's opinion
 '... before others have voiced their opinion.'
- c. ... hvordan **FN-styrkerne**₁ kan udføre **sin**₁ opgave mere rationelt.
 how UN-troops.DEF can perform REFL's task more rationally
 '... how UN troops can perform their task more rationally.'
- d. **Alle instrumenter**₁ har **sine**₁ kendetegn og **sine**₁ svagheder...
 all instruments have REFL's characteristics and REFL's weaknesses
 'All instruments have their characteristics and their weaknesses...'

The complex sentences are those that consist of (at least) a matrix clause with (at least) an embedded clause where the most referential antecedent DP is contained within the matrix clause and the reflexive is inside the embedded clause. In these cases, the local antecedent of the reflexive can be a PRO controlled by the most referential antecedent as in the examples with infinitival embedded clauses in (213a) and (213b). Participial non-finite embedded clauses are coded as non-finite, just like the infinitival embedded clauses.

- (213) a. ... og tvang **tre regioner**₁ til **PRO**₁ at tilpasse **sine**₁ love til
 and forced three regions to to adjust REFL's laws to
 forfatningen.
 constitution.DEF
 '... and forced three regions to adjust their laws to the constitution.'
- b. ... men **gæsterne**₁ formåede ikke **PRO**₁ at omsætte **sine**₁ mange
 but guests.DEF managed not to convert REFL's many
 dødbolde...
 set-pieces...
 '... but the guests did not manage to convert their many set pieces...'

Another option is that the reflexive is contained within a relative clause that modifies the most referential antecedent nominal. The most local antecedent is then the subject trace in the embedded IP, which is further moved to CP-spec from where it is available for binding by the item modified by the relative clause (as argued e.g. in

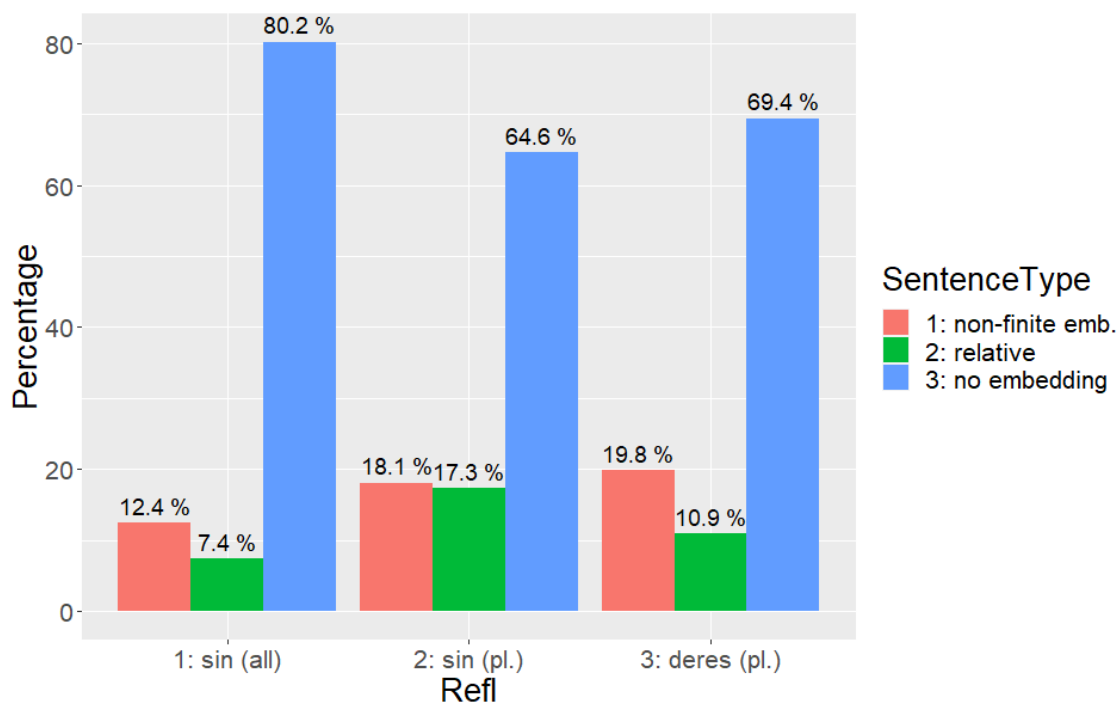


Figure 4.5: Sentence type: Simple or complex

Vikner 1991). (214a) and (214b) are examples of embedded relative clauses with reflexives.

(214) a. **Værker**₁ **ec**₁ der **t**₁ inddrager **sin**₁ beskuer.
works that draw-in REFL's viewer.
'Works that draw in their viewer.'

b. ... og **dem**₁ **ec**₁ der **t**₁ har **sin**₁ ansættelse hos privatbanerne.
and them that have REFL's employment at private-railways.DEF
'... and those who have their employment at the private railways.'

Figure 4.5 plots the distribution of simple and complex sentences in the three data groups. In the data set of *sin* with both singular and plural antecedents, the simple sentences account for 80 % of the data set. This is 15 percentage points more than in the plural antecedent *sin* group and 10 percentage points more than in the reflexive *deres* group.

All antecedents *sin* and reflexive *deres* show a similarity in their relative distribution of non-finite embedded and relative clauses: The non-finite embedded clauses outnumber the relative clauses almost two to one. The relative amount of embedded

sentences in the *deres* group is still rather higher than in the all antecedents *sin* group overall. Plural antecedent *sin* and reflexive *deres* are more similar in terms of the ratio of simple to complex sentences than plural antecedent *sin* and all antecedents *sin*. It seems that the plural antecedent groups in general contain a higher amount of complex sentences.

The plural antecedent *sin* is different from the two other groups particularly in the amount of relative clauses and the corresponding ratio of non-finite embedded to relative embedded. In the plural antecedent *sin* group the non-finite embedded clauses make up 18.1 % of the data. The relative clauses are practically as frequent at 17.3 %. One reason for this comparative majority of relative clauses in the plural antecedent *sin* group is likely that it contains a particular type of partitive relative clauses that are completely absent from the reflexive *deres* group. A few examples of the kind are shown in (215). These are all included in the final data set but several of their kind have been excluded because it was less clear whether the antecedent is the plural nominal prepositional complement or the singular partitive element (*one of...*) that is the antecedent.

- (215) a. Hun er 34 år, ugift og en af **de piger**₁ der altid er
 she is 34 years unmarried and one of those girls who always are
 smart klædt, dufter friskt og som tiltrækker alle med **sin**₁
 smartly dressed smell freshly and who attract all with REFL's
 charme...
 charm
 'She is 34 years old, unmarried and one of those girls who are always nicely
 dressed, smell freshly and who attract everyone with their charm.'
- b. Han er en af **syv nuværende og tidligere dopingbrugere**₁ der i
 he is one of seven current and former doping-users who in
 idrætssociolog Lisbeth Wahréns specialeafhandling Kropsmodifikation
 sports-sociologist Lisbeth Wahréns thesis body-modification
 gennem anabole steroider fortæller om **sine**₁ erfaringer med
 through anabolic steroids tell about REFL's experiences with
 dopingmidler.
 doping
 'He is one of seven current and former doping users who tell about their ex-
 periences with doping in Lisbeth Wahren's thesis "Body modification through
 anabolic steroids".'

- c. ... men hun er en af **de få forfattere**₁ der har opnået at få en
 but she is one of the few authors who have succeeded to get one
 af **sine**₁ fortællinger omsat til film.
 of REFL's stories turned into movies
 '... but she is one of the few authors who have succeeded in getting one of
 their stories turned into movies.'

These sentences can be found in KorpusDK when searched for specifically – as attested by the examples in (216) which were found by searching specifically for the desired construction in KorpusDK – but are seemingly fairly rare since they do not show up in my reflexive *deres* data set. These types of sentences are discussed further on page 186 because some of them were excluded from the data.

- (216) a. ... endelig har vi fået hapsed en af **dem**₁, der ødelægger Statens
 finally have we got grabbed one of them who destroy State's.DEF
 ejendom med **deres**₁ beskidte skriverier.
 property with their dirty scribblings
 '... finally we have managed to grab one of those who destroy the property
 of the State with their dirty scribblings.'
- b. FBU-formanden er en af **de fire**₁, der har meddelt **deres**₁
 FBU-chairman.DEF is one of the four who have announced their
 kandidatur i formandskapsløbet.
 candidacy in chairman-race.DEF
 'The FBU chairman is one of the four who have announced their candidacy
 in the race for chairman.'
- c. ... var Miller en af de **15 mordere**₁, som fra **deres**₁ skjul bag en
 was Miller one of the 15 murderers who from their cover behind a
 varevogn fyrede løs på demonstrationen.
 van fired loose on demonstration.DEF
 '... was Miller one of the 15 murderers who from their cover behind a van
 fired away on the demonstration.'

A specific question to discuss here is why this particular type of relative apparently mainly occurs with embedded *sin* rather than embedded *deres*, even when the antecedent in all 46 cases in the data set is plural. A more general question is why the

plural antecedent *sin* group overall contains a higher frequency of complex sentences – both non-finite and relative – than the *sin* group with all kinds of antecedents. One possible explanation is that the intervening element (be it PRO or an empty relative operator) is not in itself specified for features and could act as an intervener between the plural-specified eventual antecedent and the reflexive. Feature-lacking elements, such as arbitrary PRO, can bind *sin* and could impose an intervention effect that overrides the prominence of the plural feature of the antecedent. Another explanation could be that longer and more complex sentences are harder to process in general, which could bring about more agreement errors.

4.8.4 Argument or adjunct

The data is coded according to whether the reflexive appears in an argument or an adjunct, relative to the verb. I count as arguments both the complements of verbs such as the examples in (217a) and (217b) and complements of complex verbs that subcategorise for a preposition such as *betale for* in (217c).

- (217) a. ... intensiverer **PLOs væbnede styrker**₁ **sine**₁ aktiviteter...
 intensify PLO's armed forces REFL's activities
 '... PLO's armed forces intensify their activities.'
- b. ... så **børn**₁ kan lave **sin**₁ egen film med Legofigurer.
 so children can make REFL's own movie with Lego-figures
 '... so children can make their own movie with Lego figures.'
- c. ... at lade **folk**₁ selv betale for **sin**₁ knæskade efter en skitur til
 to let people self pay for REFL's knee-injury after a ski-trip to
 Østrig.
 Austria.
 '... to let people pay for their knee injury themselves after a skiing trip to
 Austria.'

The examples in (218) show *sin* in various adjuncts. I define adjuncts as phrases that are not subcategorised for in the sentences and which may consequently be left out.

- (218) a. **Kommunerne**₁ har i **sit**₁ forhandlingsoplæg krævet 4-500
 municipalities.DEF have in REFL's proposal demanded 4-500
 millioner ekstra...
 millions extra

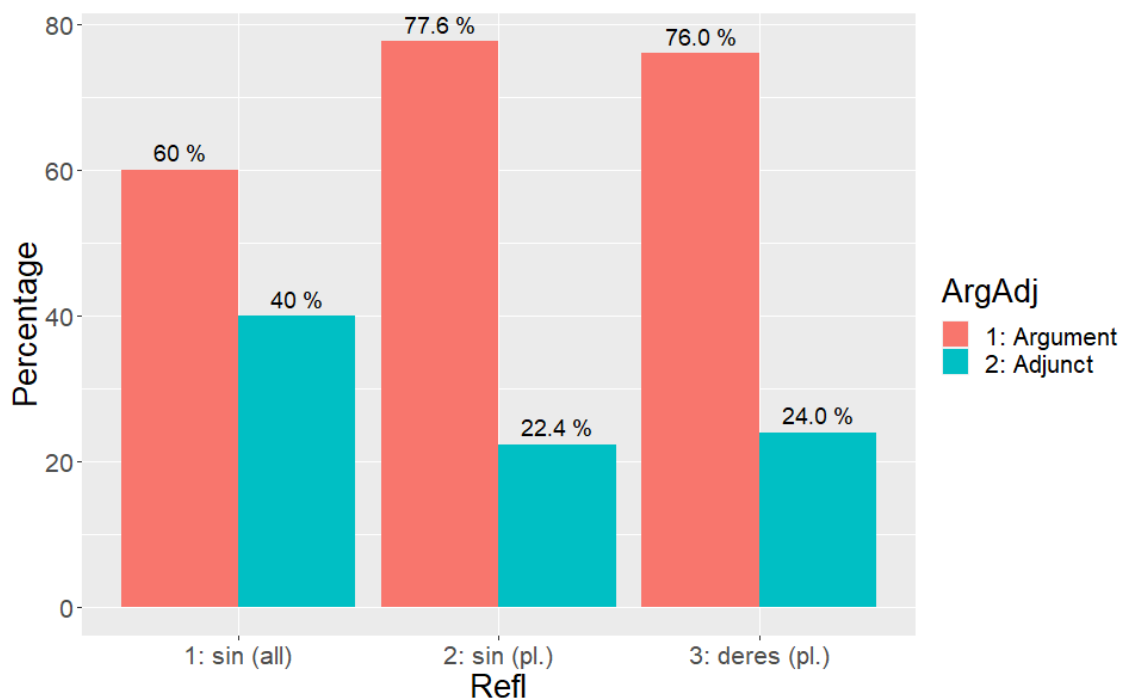


Figure 4.6: Argument or adjunct position

'The municipalities have demanded 4-500 millions extra in their proposal...'

- b. ... at **dioxiner**₁ har en effekt på mennesker ved **sin**₁ blotte
 that dioxins have an effect on humans by REFL's mere
 tilstedeværelse.
 presence
 '... that dioxins have an effect on humans by their mere presence.'

Figure 4.6 plots the distribution of arguments and adjuncts. The two plural groups, plural antecedent *sin* and reflexive *deres*, pattern remarkably similarly. Approximately one fourth of the reflexives occur in adjunct positions and three fourths in argument positions. The *sin* group with all types of antecedents is rather different from the two other groups with 60 % arguments and 40 % adjuncts.

4.8.5 Animate or inanimate

The data is coded according to the animacy of the antecedent, i.e. whether the antecedent is animate or inanimate. Several researchers have explored a connection between reflexive use and animacy. Lødorp (2009, 124) finds that inanimate subjects may bind pronominals locally in Norwegian, an option which is otherwise not

allowed. Tang (1989) describes that the Chinese simple anaphor *ziji* can only be used as an anaphor with animate antecedents. Reuland and Zubkov (2022) find, for Russian, that inanimates cannot bind reflexives non-locally and that inanimate plural antecedents are always interpreted as being distributive.

I have only distinguished between animates and inanimates in coding the data, and there is no coding subdivision between e.g. humans and non-humans. The examples in (219) have animate antecedents, human and non-human: Gulls, politicians, and children.

- (219) a. ... det gælder også for **sølvmåger**₁ som tager **sin**₁ del af
 this counts also for herring-gulls who take REFL's part of
 ællingerne.
 ducklings.DEF
 '... this is also the case the herring gulls who take their part of the ducklings.'
- b. Når **vore folkevalgte**₁ ikke lytter til **sine**₁ vælgere...
 when our electees not listen to REFL's voters
 'When our electees do not listen to their voters...'
- c. ... at **børn**₁ har en god kontakt med **sine**₁ forældre...
 that children have a good connection with REFL's parents
 '... that children have a good connection to their parents...'

The examples in (220) have inanimate antecedents: Reasonably tangible items like movies and poems, as well as corporations and sanctions, something largely intangible. I code corporations and other organisational collections of people as inanimate since the organisation as a whole can be viewed as an inanimate, even if it encompasses individual people.

- (220) a. ... og **gennemtænkte sjove film**₁, der henter **sin**₁ humor i
 and thought-out funny movies that fetch REFL's humour in
 nuet.
 now.DEF
 '... and thought-out funny movies that fetch their humour in the now.'
- b. ... hvorvidt **digtene**₁ kan have **sin**₁ oprindelse længere syd på...
 whether poems.DEF can have REFL's origin further south on

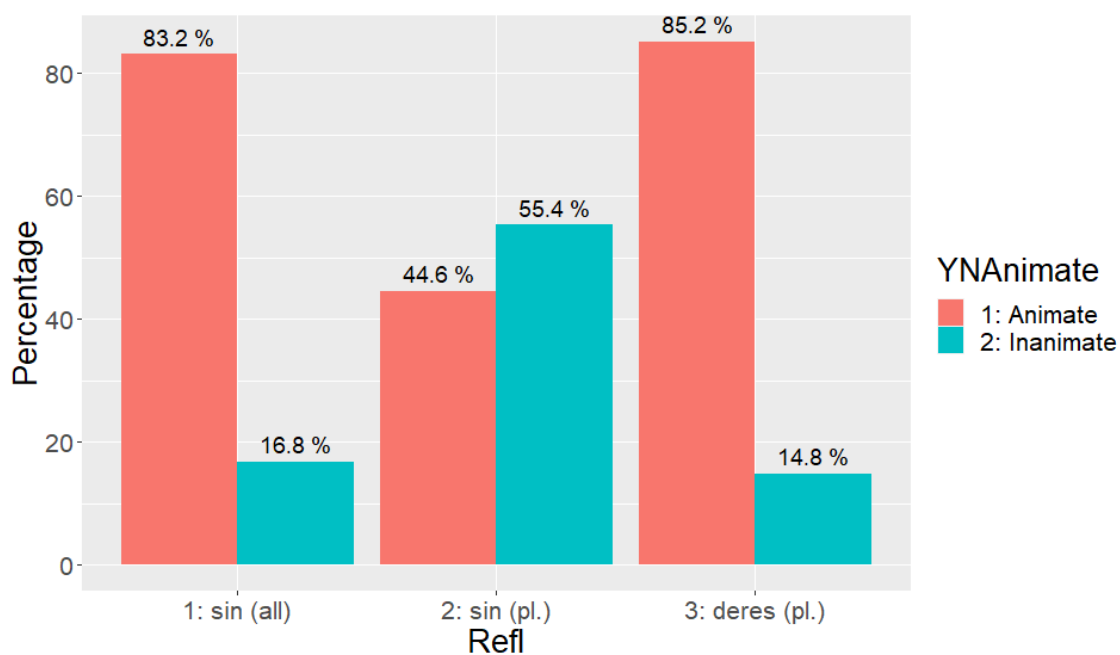


Figure 4.7: Animate or inanimate

'... whether the poems can have their origins further to the south...'

- c. Nu kan **industriens virksomheder**₁ teste **sine**₁ ufaglærte
 now can industry's.DEF companies test REFL's unskilled
 medarbejdere...
 workers
 'Now the companies of industry can test their unskilled workers...'
- d. Men **sanktioner**₁ har **sin**₁ pris...
 but sanctions have REFL's price
 'But sanctions have their price...'

Figure 4.7 plots the distribution of animates and inanimates across the three groups.

The plural antecedent *sin* group stands out notably from the two others, which pattern very much the same with 83-85 % animate and 14-17 % inanimate. The plural antecedent *sin* group contains only 44.6 % animates and 55.4 % inanimates. The assumption must be that the all antecedents *sin* and reflexive *deres* groups both illustrate the standard usage, and the animacy distribution in the plural *sin* group differs a great deal from this standard.

4.8.6 Antecedents in coordinations

The plural antecedents data sets are coded according to whether the reflexive antecedent is a coordination structure. Those antecedents that are coordinations are further coded according to the number of their individual conjuncts. Antecedents where all conjuncts are singular are coded as singular. As discussed on page 180, the initial automatic search methodology was unable to identify antecedents that consist only of coordinated singulars. This lack was remedied in a later stage in the data collection. (221a)-(221c) show examples of singular coordinations.

- (221) a. ... så **bestanden og dens afkom**₁ sikres i **sin**₁ eksistens.
 so herd.DEF and its offspring secure.PASS in REFL's existence
 '... so the herd and its offspring are secured in their existence.'
- b. ... også **lønmodtageren og den lille landmand**₁ skal have **sin**₁
 also wage-earner.DEF and the small farmer shall have REFL's
 retmæssige del...
 rightful share
 '... also the wage-earner and the small farmer must have their rightful share...'
- c. ... men **et land og en regering**₁ skal dømmes på **sine**₁
 but a country and a government shall judge.PASS on REFL's
 handlinger...
 actions
 '... but a country and a government must be judged by their actions...'

Antecedents with at least one singular and one plural conjunct are coded as a mix. (222a)-(222c) are examples of the mix category and they further illustrate that the singular conjunct may both be the first and the last conjunct.

- (222) a. **Kampen og blodsudgydelserne**₁ sætter **sine**₁ tydelige spor i
 fight.DEF and bloodsheds.DEF set REFL's clear marks in
 Jonatan...
 Jonatan
 'The fight and the bloodshed leave their clear marks in Jonatan...'
- b. ... og **både husholdningerne og industrien**₁ har øget **sin**₁
 and both households.DEF and industry.DEF have increased REFL's
 andel.
 share

'... and both the households and the industry have increased their share.'

- c. ... at **visionerne og selvrealisationen**₁ har **sine**₁ bedste
that visions.DEF and self-realisation.DEF have REFL's best
vækstvilkår...
growth-conditions
'... that the visions and the self-realisation have their best growth conditions...'

Antecedents where all conjuncts are plural are coded as plural and (223a)-(223c) show examples of these purely plural conjuncts.

- (223) a. **Flosklerne og klicheerne**₁ taler **sit**₁ eget sprog om...
platitudes.DEF and clichés speak REFL's own language about
'The platitudes and the clichés speak their own language about...'
- b. ... men **tidligere rygskader og benskader**₁ har åbenbart
but previous back-injuries and leg-injuries have apparently
krævet **sit**₁.
demanded REFL's
'... but previous back injuries and leg injuries have apparently taken their toll.'
- c. ... om **hjerteløse mødre og tåbelige fædre**₁ som ikke forstod
about heartless mothers and foolish fathers who not understood
små børns sande behov og kun tænkte på **sit**₁ eget.
small children's true needs and only thought on REFL's own.
'... about heartless mothers and foolish fathers who did not understand the true needs of small children and only thought about their own.'

The working hypothesis for this coding category is that the plural antecedent *sin* data set would have a higher relative frequency of coordinations with one or more singular conjuncts than the reflexive *deres* data set. The reasoning behind this assumption is that a singular item in the coordinated subject could act as an agreement attractor that could make the choice of *sin* over *deres* more likely.

In the data sets, which are presented visually in fig. 4.8, the differences in distribution of the coordinate structures appear to lend their support to this hypothesis. The reflexive *deres* group overall has 9.3 % sentences with a coordinated antecedent while the plural antecedent *sin* group has 25.4 %. This difference is a consequence of

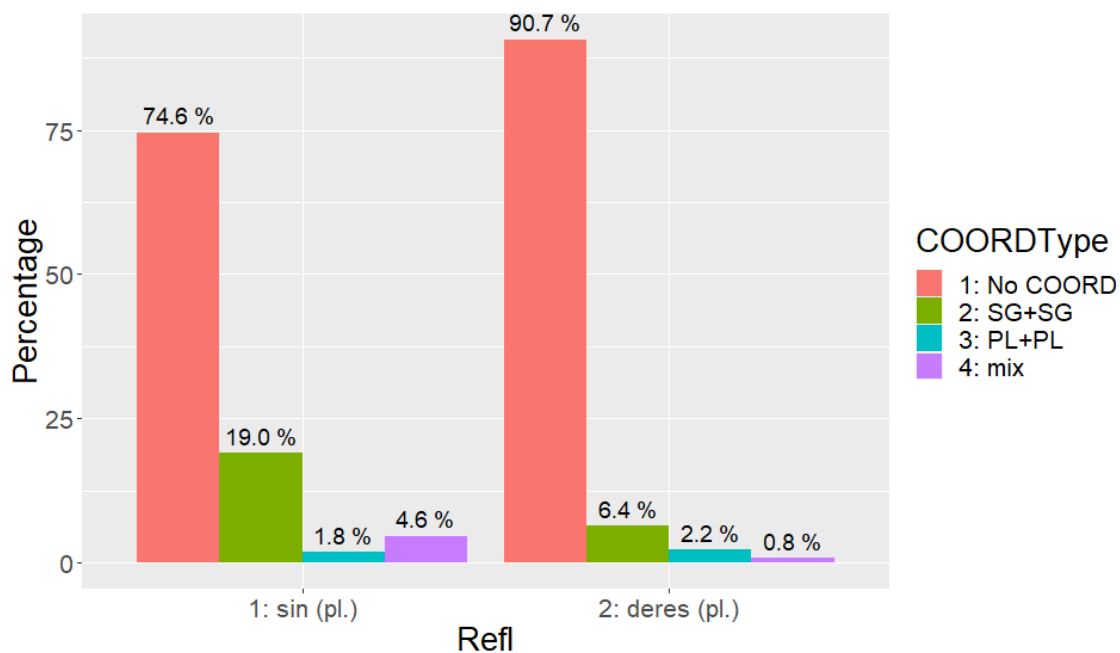


Figure 4.8: Coordinated antecedents (singular, plural, mix)

the much higher relative frequency of the two groups with coordinated singulars in the plural antecedent *sin* compared to the reflexive *deres* group. 19 % of the pl. *sin* sentences are coordinated singulars and 4.6 % contain at least one singular conjunct. The reflexive *deres* group contains 6.4 % coordinated singulars and 0.8 % sentences with at least one singular conjunct. The two groups contain approximately the same amount of coordinated plurals, ranging from 1.8 % in the plural antecedent *sin* group to 2.2 % in the reflexive *deres* group, so the difference seems to be in the singular and mix groups.

4.8.7 Distributivity

I have coded my plural antecedent *sin* and reflexive *deres* data for distributivity. The three categories I use are *distributive*, *non-distributive*, and *unclear*. The unclear cases are the sentences that are ambiguous between a distributive and a non-distributive reading and where the larger context does not disambiguate sufficiently. The examples in (224) are coded as distributive.

- (224) a. **Alt for mange**₁ undlader at pleje **sin**₁ krop...
 all too many fail to tend REFL's body
 'Far too many fail to care for their body.'

- b. ... der var jo **visse personer**₁ der altid skulle ha' **sin**₁ vilje
 there were yes certain people who always should have REFL's way
 '... there were after all certain people who always wanted to get their way.'
- c. ... at der er **mange fædre**₁ som gerne vil have **sine**₁ børn
 that there are many fathers who certainly will have REFL's children
 boende hos sig.
 living with REFL
 '... that there are many fathers who will certainly want to have their children living with them.'
- d. **Stadig flere firmaer**₁ vælger at sende **sine**₁ chauffører på køreteknisk
 still more companies choose to send REFL's drivers on driving
 kursus...
 course
 'Still more companies choose to send REFL's drivers on a driving course...'
- e. **Alle dele**₁ har **sin**₁ egen funktion...
 all parts have REFL's own function
 'All parts have their own function...'

Each sentence contains a plural subject with a relationship to the reflexive *sin* that must be read atomically from the point of the subject, i.e. as a one-to-one or one-to-many relationship between individual members of the subject and the reflexive nominal. Sentences with distributive readings can be found both with *sin* in singular and plural DPs. In (224a), each individual member of the group *alt for mange* has a body that they do not tend for. In (224b) certain individuals tend to get their (personal, individual) way. In (224c) the subject is a generalized group of fathers who each have one or more children (a slightly different, but still overall distributive, reading is possible where a child in some cases has more than one father). This is additionally an example of a *dependent plural* (discussed further in section 2.5.5). (224d) is also a dependent plural with a one-to-one-or-more reading, i.e. all companies in question have at least one driver but some or all may have more than one. Finally, (224e) is straight-forwardly distributive with one function for each part. It should be noted here that *egen* does not convey distributivity but rather a contrast of ownership (note e.g. that *Spejderne har fået sin egen hytte* (En. *The scouts have been given their own hut*) does not imply that each scout has their own hut).

The examples in (225) are examples of non-distributive use of plural antecedent *sin*. These provide readings where the plural subject is read as a group or a collective that shares the relationship to the reflexive.

- (225) a. **Irakerne**₁ solgte **sin**₁ råolie til Storbritannien...
 Iraqis sold REFL's crude-oil to Great-Britain
 'The Iraqis sold their crude oil to Great Britain...'
- b. **Fremad Valbys håndbolddamer**₁ måtte indkassere **sit**₁ femte
 Fremad Valby's handball-ladies must.PST rake-in REFL's fifth
 nederlag i træk...
 defeat in row
 'Fremad Valby's handball ladies had to rake in their fifth defeat in a row...'
- c. **Hitlers forbrydelser**₁ var på det tidspunkt kun i **sin**₁ vorden.
 Hitler's crimes were on that time only in REFL's infancy
 'Hitler's crimes were only in their infancy at the time.'
- d. ... som **museets forevisninger**₁ har oplevet i hele **sin**₁
 as museum's.DEF screenings have experienced in whole REFL's
 levetid.
 lifetime
 '... as the museum's screenings have experienced in their entire lifetime.'
- e. **Bombardementerne**₁ går ind i **sit**₁ sjette døgn.
 bombardments.DEF go in in REFL's sixth day
 'The bombardments enter their sixth day.'
- f. ... viser hvem **højreekstremisterne**₁ henter **sin**₁ støtte hos.
 shows who right-wing-extremists.DEF receive REFL's support from
 '... shows who the right-wing extremists receive their support from.'
- g. ... at **hans oprørere**₁ vil fortsætte **sine**₁ fremstød mens Mobutu
 that his rebels will continue REFL's offensives while Mobutu
 overvejer deres krav...
 considers their demands
 '... that his rebels will continue their offensives while Mobutu considers their demands...'

- h. ... da to af øens store selskaber₁ offentliggjorde sine₁
 as two of island's.DEF large companies announced REFL's
 fusionsplaner.
 merger plans
 '... as two of the island's large companies announced their merger-plans.'
- i. ... om bakterierne i gyllen₁ giver sine₁ resistensgener videre...
 if bacteria.DEF in manure.DEF give REFL's resistance-genes on
 '... if the bacteria in the manure pass their resistance genes on...'
- j. **Konflikterne**₁ har sine₁ rødder i kolonitiden...
 conflicts.DEF have REFL's roots in colonial-time.DEF
 'The conflicts have their roots in the colonial period...'

In (225a) a group of Iraqis sell oil, not the individual members. The handball ladies in (225b) lose the game together as a team. (225c)-(225e) are similar in kind in that they all express their plural, inanimate subjects as a group whose relationship to the reflexive is more in terms of a feature of the subject rather than something that the subject (or individual members of the subject) may possess. (225f) and (225g) have animate plural subjects which together receive support or continue battles. The subject of (225h) explicitly consists of two entities which together share merger plans. This is coded as non-distributive in spite of the fact that the subject is clearly atomized, since the individual members of the subject possess the reflexive nominal together. The bacterial subject of (225i) share resistance genes which can be passed on: Even in a reading where individual bacteria pass on their genes, the genes that are passed on are presumably common to all the bacteria, making the sentence non-distributive. The conflicts in (225j) should probably be read not as a collective but as a group of individual conflicts. The reading becomes non-distributive nonetheless as they all stem from the same situation of colonial repression, rather than each conflict having an individual colonial root.

The examples in (226) show a selection of sentences with unclear or ambiguous distributivity. Neither the sentence itself nor the wider context of its paragraph has been sufficient to disambiguate the readings completely. Some of the example lean towards distributivity and some of them lean towards non-distributivity but not enough to make a completely clear choice.

Unclear distributivity

- (226) a. ... især i weekenderne har **adresserne**₁ svært ved at
 particularly in weekends.DEF have addresses.DEF difficult by to
 rumme **sit**₁ publikum...
 contain REFL's audience
 '... in the weekends in particular the addresses have trouble containing
 their audience...'
- b. Nu kan **industriens virksomheder**₁ teste **sine**₁ ufaglærte
 now can industry's.DEF companies test REFL's unskilled
 medarbejdere...
 workers
 'Now the companies of industry can test their unskilled workers...'
- c. ... som også **danske tv-stationer**₁ fylder **sine**₁ seere med.
 that also Danish TV-stations fill REFL's viewers with
 '... that also Danish TV stations fill their viewers with.'
- d. ... med **pragfulde smykker**₁ der ofte tager **sit**₁ udgangspunkt i
 with magnificent jewelry.PL that often take REFL's origin in
 organiske former.
 organic shapes.
 '... with magnificent pieces of jewelry that often have their origin in organic
 shapes.'
- e. Aftalen er at **de medvirkende**₁ får **sin**₁ del så snart filmen
 deal.DEF is that the cast-members get REFL's share so soon movie.DEF
 går i overskud.
 goes in profit
 'The deal is that the cast members get their share as soon as the movie
 yields a profit.'
- f. ... at **højskolerne**₁ har udspillet en del af **sin**₁ rolle...
 that folk-schools have out-played a part of REFL's role
 '... that the folk schools are partly out of the game...'
- g. ... så **børn**₁ kan lave **sin**₁ egen film med Legofigurer.
 so children can make REFL's own movie with Lego figures
 '... so that children can make their own movie with Lego figures.'

- h. ... tvang **byerne**₁ til at rense **sit**₁ spildevand.
 forced cities.DEF to to clean REFL's waste water
 '... forced the cities to clean their waste water.'
- i. **Universiteterne**₁ lader **sine**₁ kandidater i stikken...
 universities.DEF let REFL's candidates in lurch.DEF
 'The universities leave their candidates in the lurch...'
- j. ... før **rockerne**₁ må smide **sine**₁ læderveste med mexicanere og
 before bikers.DEF must shed REFL's leather-vests with Mexicans and
 dødningshoveder...
 skulls
 '... before the bikers must shed their leather vests with Mexicans and skulls...'

The example in (226a) centers around three night clubs on the same street and it is not clear whether the night club audience in question is meant to split into groups that individually only go to one night club (the distributive reading) or whether the audience in question is shared between the clubs (the non-distributive reading). The unskilled workers in (225b) could either be read as a group of subgroups where each subgroup of workers belongs to one company (the distributive reading) or they could be read as a more general mass of those workers who work in industry (the non-distributive reading). The viewers in (226c) could be read as individual subgroups, one for each TV station (distributive), or as a generalized mass of TV viewers (non-distributive). The pieces of jewelry in (226d) may have individual origins (distributive) or a shared one (non-distributive). The share to be meted out in (226e) could be read as a bit for each individual cast member (distributive) or as a larger pool of money that belongs to the cast as a group (non-distributive). The folk schools in (226f) could have individual roles to play, one for each particular school (distributive), or they could have a common role to play in society as a school concept (non-distributive). Each individual child in (226g) could make a movie (distributive), small groups of children could make a movie together (still somewhat distributive), or all the children could make a movie together (strongly non-distributive and probably a less likely reading). The waste water in (226h) could be read as individual pools, each belonging to one city (distributive), or as a shared pool of waste water from all the city (non-distributive). The candidates in (226i) could be read as groups of candidates, one group for each university (distributive), or as one general group of university candidates (non-distributive).

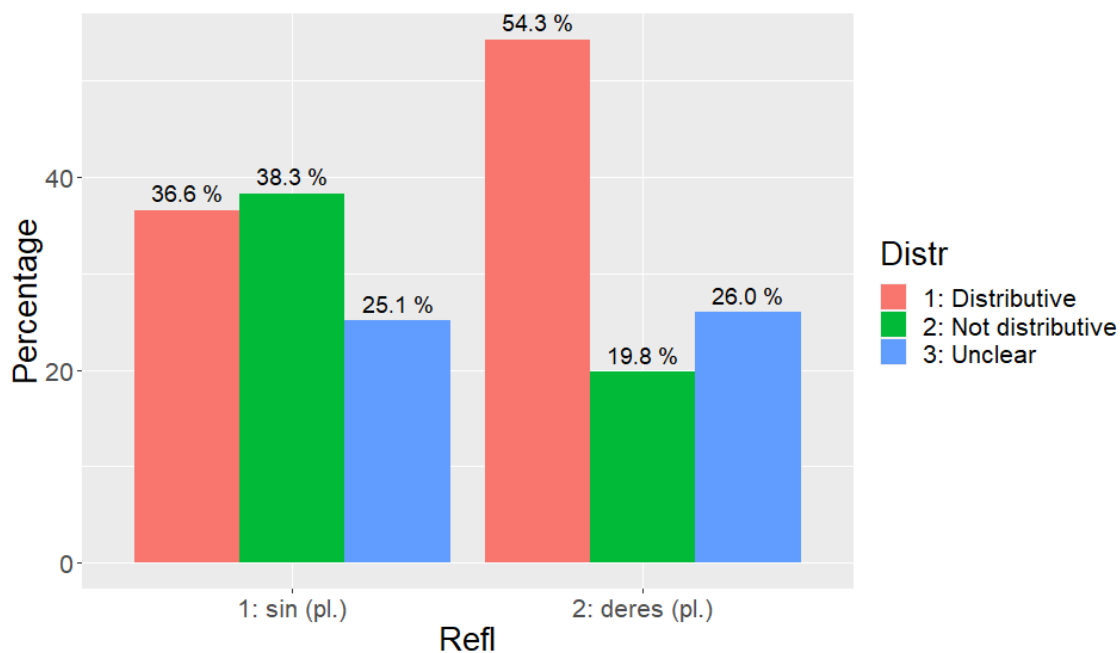


Figure 4.9: Distributivity in the KorpusDK data

(226j) could be read as distributive where each individual biker takes off a vest, or it could be read as a more general and perhaps more abstract statement about a group of bikers having to shed all their vests. A point in favour of the latter, more general, interpretation is that presumably not all individual vests would feature both Mexicans and skulls.

Figure 4.9 plots the data for distributivity in KorpusDK for the plural antecedent *sin* and reflexive *deres* data sets.

The *sin* group and the *deres* group have essentially the same amount of sentences with unclear distributivity at around 25 %. The initial hypothesis was that distributivity improves the acceptability of plural antecedent *sin* and the data seems to be a clear argument *against* this hypothesis, against my own initial expectations.

4.8.8 Distance between reflexive and antecedent

The examples in (227) and (228) show the two extremes on the spectrum: Two sentences where antecedent and reflexive are right next to each other in the linear order and two sentences where the overt antecedent and the reflexive are very far apart.

Short distance between reflexive and antecedent

- (227) a. I dette perspektiv fik **fagene**₁ **sin**₁ ofte uudtalte begrundelse.
 in this perspective got subjects.DEF REFL's often unspoken reason
 'In this perspective the subjects got their often unspoken reason.'
- b. I det katolske land har **alle byer**₁ **sine**₁ moteller..
 in the Catholic country have all towns REFL's motels
 'In the Catholic country all towns have their motels...'

Long distance between reflexive and antecedent

- (228) a. Han hører til blandt **de garvede dommere**₁, der har overstået
 he hears to among the experienced judges who have done
 tiden som førstevoterende og derfor kan nøjes med at
 time.DEF as first-voting and consequently can do.PASS with to
 komme i retten hver anden dag og da normalt gå klokken 14
 come in court.DEF every other day and then normally leave o'clock 14
 for at passe **sine**₁ bijob.
 for to tend REFL's extra jobs
 'He is among the experienced judged who have done their time as first-
 voting and can consequently make do with coming to court every second
 day and then normally leave at 14 o'clock to take care of their extra jobs.'
- b. I kvartfinalen venter **de forsvarende mestre**₁, Hypo
 in quarter-final.DEF wait the defending champions Hypo
 Niederösterreich, som i gruppe mod Bækkelaget fra Oslo, Krim
 Niederösterreich who in group against Bækkelaget from Oslo, Krim
 Electa fra Ljublana i Slovenien og Zaporoshje fra Ukraine har
 Electa from Ljublana in Slovenia and Zaporoshje from Ukraine have
 fejjet al modstand til side med lutter sejre i **sine**₁ fem første
 swept all resistance to side with sheer victories in REFL's five first
 kampe.
 matches
 'In the quarter final the defending champions, Hypo Niederösterreich, await
 who in the group against Bækkelaget fra Oslo, Krim Electa from Ljublana
 in Slovenia and Zaporoshje from Ukraine have swept aside all resistance
 with sheer victories in their five first matches.'

I have counted the number of words between the semantic antecedent (defined as the highest element in the antecedent chain, which may also consist of one or more

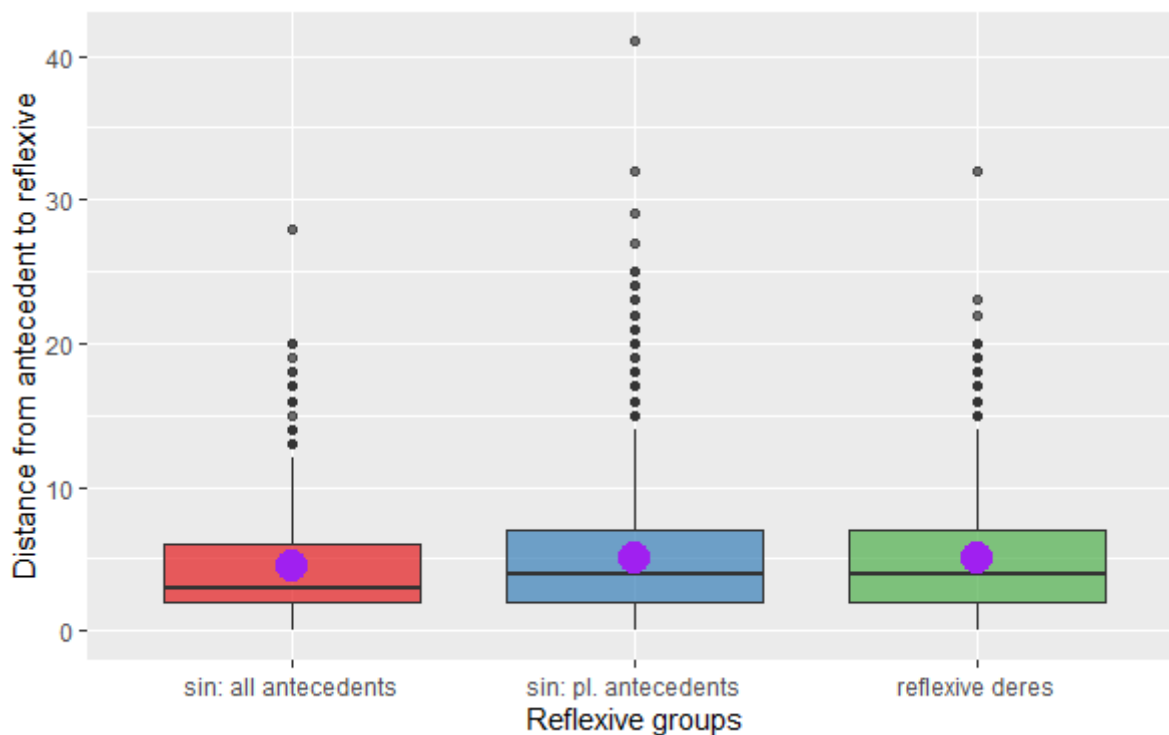


Figure 4.10: Linear distance between antecedent and reflexive

local, non-overt antecedent phrases) and reflexive for all sentences in the data sets. The data is visualised as box-and-whiskers plots in (4.10) for all three groups.

The data points are the counts, one for each sentence in the data set, of the linear distance between the topmost antecedent phrase and the reflexive. The number of words between antecedent and reflexive fall within the range of 0 to 41.⁵

The relevant numbers from fig. 4.10 are summarized in table 4.11. All antecedents *sin* stands out in having a lower second quartile (Q2), third quartile (Q3), and mean than the two other groups. The range from Q1 to Q2 is lower, as well, which indicates that the all antecedents *sin* groups has a greater proportion of sentences with a shorter distance than the two other groups. It should, however, be noted that this difference is very small: It is the difference between having three or four words between

⁵The horizontal lines in the boxes represent – from bottom to top – the first quartile (Q1, the highest data point of the first 25 % of the data, sorted from smallest to largest), the median (Q2, the highest point of the first 50 % of the sorted data), and the third quartile (Q3, the highest data point of the first 75 % of the sorted data). The purple dot inside the boxes shows the mean of the data points. The vertical line at the top (and bottom) of the boxes is the "whisker" part of the box-and-whisker plot. The top vertical line represents the data points above the third quartile which fall within 1.5 times the interquartile range, i.e. 1.5 times the distance between Q1 and Q3. Anything in the data set above the whiskers is an outlier by convention and these are represented by the small black dots. The bottom vertical line represents the data points from 0 to the value of Q1.

Table 4.11: Numerical values for the plot in (4.10)

Data set	Min.	Q1	Q2	Mean	Q3	Max
All <i>sin</i>	0	2	3	4.5	6	28
Pl. antc. <i>sin</i>	0	2	4	5.13	7	41
Reflexive <i>deres</i>	0	2	4	5.03	7	32

antecedent and reflexive. The mean is higher than the median in all three groups, likely a consequence of the outliers, which particularly for plural antecedent *sin* are quite a bit higher than the great majority of the data.

The hypothesis to be tested is that there is a greater frequency of sentences in the plural antecedent *sin* group with a long distance between antecedent and reflexive than in the two other reflexive groups. The data overall does not support this hypothesis as the three groups pattern much the same. The main difference is that the plural antecedent *sin* group contains more outliers that are also further away from the main body of the data than in the two other groups. It does not seem that distance between reflexive and antecedent is a strong factor in the use of plural antecedent *sin*.

4.9 Discussion

The corpus study described in this chapter was aimed at answering the four research questions below and I summarize what the data shows for each question. The study is mainly explorative and descriptive out of some necessity, simply because there are not that many claims about the syntactic or semantic behaviour of plural antecedent reflexives that I managed to find in the literature. The few exceptions to this are the factors *distributivity*, *animacy*, and *number of the nominal that contains sin*. The explorative nature of the study also means that I may have missed relevant factors or put too high a significance on others. One missed factor that I would like to code for in a follow-up to this study, is the natural and grammatical gender of the antecedent. I hypothesize in section 2.5.3 that gender, more so than animacy, could be an explanatory factor. This hypothesis emerged at a later stage of the project, which means that I have not had the opportunity to code the data for gender and consequently that I do not (yet) have the data to support or dismiss that claim.

1. How frequently do instances of plural antecedent *sin* occur in the corpus?

2. Do the instances of plural antecedent *sin* pattern similarly to regular instances of plural antecedent *deres* or *sin* with any kind of antecedent?
3. Which syntactic factors favour plural antecedent *sin*?
4. Which semantic factors favour plural antecedent *sin*?

4.9.1 RQ1: How frequently do instances of plural antecedent *sin* occur in the corpus?

I found 1218 examples of *sin* with plural antecedent in KorpusDK. A search for *sin*, *sit*, and *sine* in KorpusDK returns 188,404 examples. The plural antecedent *sin* examples make up 0.6 % of this data. This is a very small subset, by all measures. Plural antecedent *sin* is also much less frequent in the written corpus than in the spoken LANCHART corpus, where the plural antecedent examples make up as much as 7 % of all the examples with *sin*. An interesting and still-unanswered question is whether the frequency of plural antecedent *sin* is increasing. My original intention was to compare the frequencies of plural antecedent *sin* in Korpus90 with the frequencies of plural antecedent *sin* in Korpus2010 in order to get as large a time span as possible with reasonably similar data sets. This unfortunately turned out to be infeasible within the current study due to the size of (and amounts of errors in) Korpus2010. It should be possible to do this comparison on a random subset of e.g. 10 % of the items from each corpus and that is certainly a topic for a follow-up study.

4.9.2 RQ2: Do the instances of plural antecedent *sin* pattern similarly to regular instances of plural antecedent *deres*? or *sin* with any kind of antecedent?

I compare the data set of 1218 examples of plural antecedent *sin* with a similar data set of 1152 examples of bound *deres* and another data set of 500 examples of *sin* with any kind of antecedent. I call bound *deres* and *sin* with any antecedent the *standard forms*.

Plural antecedent *sin* generally does not pattern like the standard forms, i.e. like plural antecedent *deres* or like *sin* in general. It is clearly a thing of its own rather than just a randomly occurring alternative to plural antecedent *deres*. The standard

forms pattern more alike for some factors while plural antecedent *sin* patterns more like plural antecedent *deres* for other factors.

Type of subject (full noun or pronoun)

For the factor `subject type`, i.e. whether the antecedent is a full noun or a pronoun, the two standard forms are quite similar in having approximately two thirds full noun antecedents (63.8-71.9 %) and one third pronoun antecedents (36.2-28.1 %). Plural *sin* has a much higher occurrence of full noun antecedents at 93.3 % and only 6.7 % pronouns.

Number of the nominal that contains *sin*

An early, informal observation about plural antecedent *sin* is that it seems to be more acceptable when it is contained within a plural nominal. The corpus data supports this to some extent but not entirely. *Sin* in general occurs slightly more often within singular nominals than the overall distribution of singular and plural nominals in the corpus would lead us to expect. *Sin* in general occurs in 80 % singular nominals and 20 % plural nominals, and the corpus contains 72 % singular nominals overall. Compared to all instances of *sin*, plural antecedent *sin* does occur relatively more frequently in plural nominals: 33.7 % of all plural antecedent *sin* are contained within plural nominals. However, the bound instances of *deres* are to an even greater extent contain with plural nominals, namely 42.6 % of the occurrences of plural antecedent *deres*. Plural antecedent *sin* apparently falls in between the two standard forms in this factor.

Sentence type (simple or complex)

Plural antecedent *sin* differs somewhat, and in different ways, from the two standard forms in terms of the clause types that they occur in. Plural antecedent *sin* and plural antecedent *deres* both occur slightly more often in embedded clauses (30.6-35.4 %) than *sin* does in general (19.8 %). Plural antecedent *sin* occurs more often in relative clauses than plural antecedent *deres* (6.4 percentage points which corresponds to approximately 60 % more in relative terms). This is most likely due to a specific type of clause where a relative clause modifies a particular kind of partitive (see more on page 186). These clauses are much more common in the plural antecedent *sin* data

set than in the plural antecedent *deres* data set and likely explain most of the observed difference.

Argument or adjunct

Plural antecedent *sin* and plural antecedent *deres* essentially have the same distribution of how frequently they are contained within an adjunct (a modifier relative to the verb) or in an argument (a complement relative to the main verb). 22.4-24 % of plural antecedent *sin* or *deres* are found inside verbal modifiers and the rest inside verbal complements. *Sin* in general is found in 40 % modifiers. Some of this difference could very well be explained by the search setup for the two plural antecedent groups. The automatic search misses most examples of modifiers that are topicalisations of *sin* or *deres*. This has the effect that there is one example of a topicalised bound *deres* and five examples of a topicalised bound plural antecedent *sin* in the data (out of 1152 plural antecedent bound *deres* and 1218 *sin*). There are 21 examples of topicalised *sin* with all kinds of antecedents, by comparison, out of a sample of 500 sentences. It is unfortunately hard to say with the chosen search setup whether the different amounts of topicalisations are due to the search strategy alone or whether there is a more substantive reason.

Animate or inanimate

The factor where *sin* with plural antecedents stands out the most is in the animacy of the antecedent. All antecedents *sin* and bound *deres* have approximately 15 % inanimate antecedents and 85 % animate antecedents. The picture for plural antecedent *sin* is very different in that there are 55 % inanimate antecedents and only 45 % animate antecedents in the data. There is a tendency in Danish to use locally bound pronouns – rather than reflexives – if the antecedent is animate and is overtly gendered as masculine or feminine. Inanimates are not gendered (discounting grammatical gender) so one hypothesis is that the lack of natural gender on inanimates could make the choice of *sin* with a plural inanimate antecedent more likely. This must, however, remain an unsubstantiated hypothesis as I have not coded my data for natural gender. Another impression from the data is that the relation of the *sin* nominal to the inanimate antecedents is more like a property or a feature of the antecedent rather than something that belongs to the antecedent. The fact that the inanimate antecedents are also relatively often collective could point to *semantic agreement* a

la Corbett (discussed from page 105 and onwards) as a possible mechanism behind some of the data. The animate antecedents tend to have a more classic ownership relationship to the *sin* nominal.

Antecedents in coordinations

The antecedents of plural antecedent *sin* are fairly frequently (23.6 %) coordinations of singular nominals or of a singular nominal and a plural nominal. This number is much higher than for bound *deres* where 7.2 % of the antecedents are coordinated singulars or singular coordinated with a plural. This difference seems like a clear case of agreement attraction where the singular coordinand(s) are treated like the agreement target rather than the whole coordination. This is also known as closest conjunct agreement (CCA, also approached as first or last conjunct agreement) (see e.g. the review in Nevins and Weisser 2019). Some analyses of CCA argue that the coordinated structure is ϕ -deficient either in terms of number or gender, and this fits well with an analysis where the Danish *sin* is more likely to be allowed to agree with plural antecedents if they are not overtly specified for plural or natural gender.

Distributivity

The second-to-last factor that the data is coded for is distributivity. This is one of the only syntactic factors touched upon in the literature on plural antecedent *sin* and the prediction is that plural antecedent *sin* will be more likely to occur if the plural antecedent is also distributive. This is not obviously what the data shows, however. Distributive and non-distributive plural antecedents are approximately equally frequent with *sin* (36.6-38.3 %). Comparing these numbers to bound *deres*, we see a rather higher proportion of distributive antecedents (54.3 %) compared to non-distributive antecedents (19.8 %). Looking a bit further out in the linguistic landscape, some researchers have suggested a close connection between animacy and distributivity with reflexives. Reuland and Zubkov (2022, 13) write about the Russian reflexives *sebe* and *sebjja* that "binding by plural inanimates is always interpreted distributively". Huang (2002) reports about Chinese that the simple reflexive *ziji* either forces a distributive reading (when locally bound) or requires distributive marking (when non-locally bound). I split my corpus data according to distributivity and animacy in order to see whether anything similar is the case for plural antecedent *sin* and *deres* in Danish. The results are reported in table 4.12.

Table 4.12: Animacy and distributivity with *sin* and *deres* with plural antecedents, animate or inanimate

	<i>sin</i>		<i>deres</i>	
Animate antecedent	N = 543	(100 %)	N = 982	(100 %)
Collective	N = 145	(26.7 %)	N = 188	(19.1 %)
Distributive	N = 315	(58.0 %)	N = 565	(57.5 %)
Unclear	N = 83	(15.3 %)	N = 229	(23.3 %)
Inanimate antecedent	N = 675	(100 %)	N = 170	(100 %)
Collective	322	(47.7 %)	N = 40	(23.5 %)
Distributive	130	(19.3 %)	N = 60	(35.3 %)
Unclear	223	(33.0 %)	N = 70	(41.2 %)

The table shows how sentences with plural antecedent *sin* and *deres* pattern according to animacy and distributivity. The *sin* sample has a much larger proportion of sentences with an inanimate antecedent (55.4 % inanimates, compared to 14.8 % inanimates in the *deres* sample). The numbers in the two columns pattern in quite the same ways with animate antecedents: A smaller proportion of collective sentences (20-25 % of the sample) and a larger proportion of distributive sentences (around 60 % of the sample). The sentences with inanimate antecedents are quite different from each other across the two samples and quite different from the sentences with animate antecedents. The inanimate *sin* sentences show a fairly strong preference for collective readings (50 % collective compared to 20 % distributive), which is almost exactly the opposite of how the sentences with animate antecedents are distributed. The *deres* sentences pattern more like the animate sentences with a larger proportion of distributive (35 %) compared to collective (25 %). Both inanimate groups have a fairly high proportion of sentences with unclear distributivity. Danish clearly allows both distributive and collective readings with animate and inanimate antecedents, and plural antecedent *sin* even shows a preference for collective readings with inanimate antecedents. This is, perhaps not too surprisingly, very different from the results reported for Russian and Chinese.

Distance between reflexive and antecedent

There was no strong tendency in either direction brought about by the linear distance between the reflexive and the antecedent.

4.9.3 RQ3: Which syntactic factors favour plural antecedent *sin*?

The details of this research question are large discussed in the sections above. Summarizing the summary, plural antecedent *sin* occurs slightly more often (than *sin* with any kind of antecedent) in plural nominals. Plural antecedent *sin* occurs more often in complex sentences, and if the plural antecedent is a coordination with one or more singular nominals. It seems likely that agreement attraction plays a part here, i.e. that the reflexive possessive ends up agreeing with a singular coordinand, with a number-neutral non-overt subject, or with the plural antecedent in a relationship reminiscent of a dependent plural.

4.9.4 RQ4: Which semantic factors favour plural antecedent *sin*?

Plural antecedent *sin* has a large overrepresentation of inanimate antecedents. These inanimate antecedents, in addition, are more often collective (rather than distributive). This is a very different pattern from both *sin* with all antecedents (only in terms of animacy) and from bound *deres*. A possible mechanism here is semantic agreement where *sin* and *deres* are both available as bound forms with collective singulars as the antecedent. A similar mechanism in the other direction could be at play where the possessive reflexive takes part in a kind of semantic agreement with the collective inanimates.

4.9.5 Final note

KorpusDK does not have a great deal of background data for texts and the data consequently is not very well suited to investigate the sociolinguistic or dialectal factors that could have an impact on the use of plural antecedent *sin*. This is possible to a much greater extent with the acceptability data in the study that is the topic of chapter 5. The specific syntactic and semantic factors that make up the acceptability judgment study are further informed by the results from the corpus study in this chapter.

Chapter 5

Acceptability judgment data

5.1 Introduction

In this chapter, I look specifically at how a relatively homogenous sample of young Danish speakers rate the use of possessive reflexive *sin* (*sit, sine*) used with a plural subject. This is a usage which is normally taken to be ungrammatical in Danish. An example of this is *Forældrene elsker sine børn*, En. *The parents love their children*, which in prescriptively correct Danish should be *Forældrene elsker deres børn*. In the corpus study described in chapter 4, I investigated the occurrence of plural antecedent *sin* in the Danish written corpus KorpusDK. The acceptability study discussed in this chapter is a different empirical approach to the topic and can be used to answer different questions. The distributional results from the corpus study allowed me to build the experimental items in the acceptability study on a foundation of attested material. The format of the acceptability judgment study enables a more fine-grained study of the factors that turned out to be relevant in the corpus study. It additionally provides access to sociolinguistic factors such as regional or socio-economic differences between speakers. This type of information is not available in the corpus, and e.g. dialectal differences could very well have an impact on how speakers rate the non-standard uses of *sin* in the experiment. One hypothesis in my study is that speakers from the Western Jutland dialect area may hyper-correct their use of *sin* due to prescriptivist pressure from the standard dialect. Hyper-correction could lead speakers to overgeneralize and accept *sin* with plural antecedents to a greater extent than other speakers.

It is a basic assumption within my branch of linguistics that language change hap-

pens when children acquire language and in some cases reanalyze the input from the language they hear. Consequently, a difference in language use from one generation to a younger one can potentially be a sign that a change is taking place in the language. In a previous, unpublished, pilot study, I investigated the acceptability of *sin* with plural antecedents in a sample of 500 Danish speakers of ages 25 to 50+ which did show a marked age effect, younger speakers accepting *sin* with plural subjects more easily than older speakers. In the study reported here, I investigate a reasonably homogeneous sample of younger speakers in order to investigate the extent that linguistic structure and regional dialect features may influence the acceptability of *sin* with plural subjects. My original intention was to recreate this experiment with older subjects in order to make a case for whether the use of *sin* with plural antecedents shows signs of intergenerational differences as an apparent time study (e.g. Bailey et al. 1991). I unfortunately did not manage to do so for this thesis but it would be a very interesting and obvious follow-up study.

5.2 Experimental design and methodology

The experiment discussed in this chapter is an acceptability judgment study. This type of study is widely used and generally accepted as a valid and useful methodology within the field of generative experimental syntax (see e.g. Sprouse (2015), Christensen (2019), A. Chacón (2021), and Goodall (2021) for various reviews). See also e.g. Schindler and Brøcker (2020) for a non-generative outline of the arguments against acceptability judgments as the better source of intuition data (compared to more informal approaches to introspection).

The aim of the study is to investigate the acceptability of reflexive possessive *sin* with plural antecedents in a reasonably homogenous sample of young Danish speakers. The experiment is created in a way that lets me investigate whether a number of specific morphosyntactic factors improve (or decrease) the acceptability of plural antecedent *sin*. It also allows me to investigate whether other non-standard uses of reflexive pronouns may influence the acceptability of plural antecedent *sin*. Finally, the size and regional variety of the sample population allows me to investigate whether there are specific regional, social, or dialectal differences at play in the acceptability of plural antecedent *sin*.

593 people participated in the study. The participants are all Danish speakers and

they were all students at various Danish upper secondary education programmes (Da. 'gymnasium', henceforth in the text simply singular *gymnasium* or plural *gymnasiums*). Approximately 70 % of students who finish *grundskolen* choose the gymnasium as their first priority for continued youth education¹. Students most often attend the gymnasium for three years approximately between the age of 16 and 19. The oldest speakers in the sample are in their early 20s and the youngest are 15. A pilot study with a more diverse sample population was performed prior to the main study, which helped improve the final design of the experiment.

5.2.1 Distribution of the questionnaire

The participants in the study are all students at Danish gymnasiums. They were recruited through their teachers (who, in turn, were recruited mainly from my personal or professional networks). The data for the study were collected online through the free survey tool Google Forms between 2016 and 2019. The data collection process took place in the classroom during school hours. The participants were asked to answer the questionnaire on their own, i.e. specifically without consulting their classmates. They were asked to use their own devices (computer, smartphone, tablet) to answer the questionnaire. The students were not rewarded in any way for participating or for finishing the questionnaire. In all but one instance, the teacher in question was responsible for initiating, explaining, and overseeing the experiment. In the one instance, one of the first times that the experiment was run, I was present myself in the classroom to distribute and introduce the questionnaire.

Participants were asked to read 100 different sentences and to provide acceptability judgments for each sentence on a scale from 1 to 5. The sentences were presented in an online questionnaire through Google Forms. The questionnaire itself spans 20 pages with five sentences on each page. All sentences are marked as obligatory, which means that the survey software will not allow a participant to go on to the next page without having provided an answer to each of the five sentences on the page. Figure 5.1 shows an example of the first page of a questionnaire. This page follows a page of background (demographic) questions.

The teacher was given a link to the online questionnaire to share with the students and a thorough description of what to say (and preferably not to say) to the stu-

¹According to a 2020 report from Gymnasieskolernes Lærerforening, <https://www.gl.org/nyt/Documents/S%C3%B8getal%202020%20til%20de%20gymnasiale%20uddannelser%20-%20GL.pdf>

The image shows a screenshot of a Google Form questionnaire page. At the top, a purple header bar contains the text "(1/20)". Below this, there are five question items, each with a five-point Likert scale. The questions are in Danish and marked with an asterisk to indicate they are required. The scales are labeled "Helt uacceptabel" (completely unacceptable) on the left and "Helt acceptabel" (completely acceptable) on the right. The response options are numbered 1 through 5, with radio buttons for each. At the bottom of the form, there are three buttons: "Back", "Next", and "Clear form".

(1/20)

Tastaturet ikke larmer når man bruger det. *

1 2 3 4 5

Helt uacceptabel Helt acceptabel

I går har jeg kage med fra arbejde. *

1 2 3 4 5

Helt uacceptabel Helt acceptabel

De køber en sofa i genbrugsbutikken. *

1 2 3 4 5

Helt uacceptabel Helt acceptabel

Manden som at der gik på vejen havde en hund med. *

1 2 3 4 5

Helt uacceptabel Helt acceptabel

Vi så ham ramme ham selv med malingen. *

1 2 3 4 5

Helt uacceptabel Helt acceptabel

Back Next Clear form

Figure 5.1: Example of the first (non-background) page of the questionnaire itself on Google Forms.

dents both before and after running the experiment. The specific instructions given to teachers who run the experiment and the participating students are provided in the Danish original and English translation in appendix A. Teachers were encouraged to try out the questionnaire themselves before giving it to their students.

The information sheet specifically asks the teachers to not explain too much about the purpose of the questionnaire. This is an attempt to prevent bias arising from the individual teachers' take on this kind of linguistic experiment. I have tried to anticipate a handful of questions that might arise while students fill out the questionnaire by including these in the teacher information sheet. This is also in order to mini-

mize bias from varying spontaneous explanations from individual teachers. Some of these questions and answers were inspired by the pilot study where participants gave feedback in various ways (in the fora where the questionnaire was shared; via e-mail directly to me). Others were inspired by written or oral comments from the first batches of participating students. The teacher instruction sheet also includes a section that explains the purpose and motivation behind the survey, which teachers were encouraged to share with their students after running the experiment. A number of teachers e-mailed me after the experiment to say that the questionnaire and the post-experiment text inspired interesting discussions about language change and reflexive use in the classroom.

The questionnaire begins with a page with an introductory instruction sheet for the students. The participant instruction sheet is reproduced in the Danish original and English translation in appendix A. The instruction sheet briefly explains the format of the questionnaire and shows a specific example of what the test sentences will look like: "Kaniner elsker gulerødder." (En. 'Rabbits love carrots.'). Below the sentence is the same 5-point rating scale that is used in the rest of the questionnaire. The text introduces the rating scale from 1 (*Helt uacceptabel*, En. 'Completely unacceptable') to 5 (*Helt acceptabel*, En. 'Completely acceptable'). This is followed by the point that there are no right or wrong answers in this setting and that I encourage participants to go with their first impulse for their ratings (rather than overthinking and overanalyzing). This point is expanded upon with a few lines that say that the interesting point, for the purposes of the study, is the language that people of the participants' own age group would actually use, rather than the rules of proper grammar that they have been taught in school. The instruction sheet ends with a statement that makes explicit that the data collected in the questionnaire is anonymous, cannot be used to identify individual students, and that the only purpose of the data collection is scientific research. This also includes a final statement that participants by filling out the questionnaire consent to their fully anonymous data being saved and stored.

5.2.2 Materials

The test items are Danish written sentences presented in an online questionnaire. The questionnaire consists of 100 sentences and a number of demographic questions. Respondents are asked to rate the sentences on a Likert scale from 1 (Not acceptable) to 5 (Acceptable). There are four slightly different versions of the questionnaire. Par-

ticipants were routed into the different questionnaires via a randomization function provided by Google Forms (this is an alternative to e.g. asking participants to pick a questionnaire based on their month of birth). Figure 5.2 shows the proportion of respondents for each of the four questionnaires. This approach seems to have resulted in a reasonably even distribution across the questionnaires: 25.7 % of the participants answered questionnaire 1, 23.5 % of the participants answered questionnaire 2, 22.3 % of the participants answered questionnaire 3, and 28.5 % of the participants answered questionnaire 4.

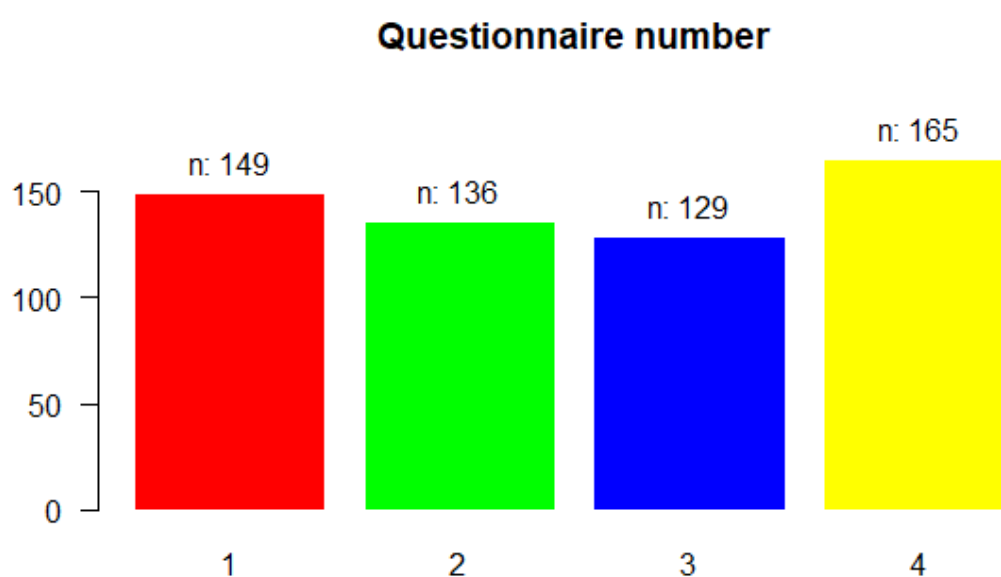


Figure 5.2: Barplot of distribution across the four different questionnaires.

Every questionnaire contains 32 **primary sentences** which vary between each questionnaire. The questionnaires all contain the same 30 **secondary** sentences, the same 30 **filler** sentences, the same four "**awake**" sentences, and the same four **training** sentences. The sentences are pseudo-randomized within each of the questionnaires. I randomize the sentences in each questionnaire with a Python script and the sequence that the sentences are presented in is consequently different for each of the four questionnaires. The sequence is only pseudo-random because I arrange the randomized sentences so that the primary items are always spaced out with two non-primary items in between. The four training sentences are always the first four sentences in the questionnaires. The sentence types are explained in greater detail in the following subsections.

Primary sentences

Each questionnaire contains 32 **primary sentences**. They represent 16 unique sentences structures which are all repeated once in each questionnaire. Participants in the study see and rate the same sentence structure twice throughout the questionnaire. The two instances of the same structure are presented with different lexical content, meaning that the structure is the same but the sentences are (lexically) different. The pair in (229) shows an example of this, i.e. two primary sentences with the same structure but different lexical content. The specific structure of the shown examples is *plural antecedent – verb – plural nominal with sin – final prepositional phrase*.

(229) **Two primary sentences with the same relevant structure but different lexical content**

- a. **Medlemmerne**₁ sender **sine**₁ forslag til bestyrelsen.
Members.DEF send REFL's suggestions to board.DEF
'The members send their suggestions to the board.'
- b. **Tvillingerne**₁ taber **sine**₁ nøgler i bussen.
Twins.DEF lose REFL's keys in bus.DEF
'The twins lose their keys in the bus.'

The primary sentences all contain a *sin* (*sit*, *sine*) which is bound by a plural subject antecedent. They are all structurally grammatical, apart from the fact of the (presumed) unacceptability of having *sin* bound by a plural antecedent in Danish. The format of the experiment is a 2×2×2×2 factorial design. This means that the sentences are built up according to four independent variables with two levels each.

The first independent variable is the number of the DP that contains *sin*. The number can be *plural* (e.g. *sine børn*, En. *REFL's children*) or *singular* (e.g. *sit barn*, En. *REFL's child*). The number of the DP is determined by the number of the head noun.

(230) **Independent variable 1: Number of the DP that contains *sin***

- a. Forældrene₁ henter **sine**₁ **børn**. (plural)
Parents.DEF collect REFL's children
'The parents collect their children.'
- b. Forældrene₁ henter **sit**₁ **barn**. (singular)
Parents.DEF collect REFL's child
'The parents collect their child.'

The second independent variable is sentence complexity. The sentence can be *simplex* or *complex*. A simplex sentence in this context is a sentence with a subject, a main verb, and a direct object with a *sin* (*sit*, *sine*). A complex sentence in this context contains the *sin* in an embedded infinitival clause with a non-overt subject. The sentence structures with the relevant clausal boundaries and non-overt subject are shown in the examples in (231).

(231) **Independent variable 2: Complexity of sentence**

- a. [CP Forældrene₁ **henter sit₁ barn**]. (simplex)
 Parents.DEF collect REFL's child
 'The parents collect their child.'
- b. [CP Forældrene₁ husker [IP PRO₁ **at hente sit₁ barn**]]. (complex)
 Parents.DEF remember to collect REFL's child
 'The parents remember collecting their child.'

The third independent variable is the type of subject. The subject can be a *pronoun* or a *full noun*.

(232) **Independent variable 3: Type of subject**

- a. **De₁** henter sit₁ barn. (pronoun subject)
 They collect REFL's child
 'They collect their child.'
- b. **Forældrene₁** henter sit₁ barn. (full noun subject)
 Parents.DEF collect REFL's child
 'The parents collect their child.'

The fourth and last independent variable is the animacy of the subject. The subject can be animate or inanimate. The content of the sentence varies according to the animacy of the subject in order to make the scenarios more plausible. This necessarily means that the animate and inanimate sentences are quite different from each other in lexical content but similar in structure.

(233) **Independent variable 4: Animacy**

- a. **Forældrene₁** henter sit₁ barn. (animate)
 Parents.DEF collect REFL's child
 'The parents collect their child.'
- b. **Træerne₁** mister sin₁ farve. (inanimate)
 Trees.DEF lose REFL's colour
 'The trees lose their colour.'

Table 5.1 sums up the four independent variables. The examples show the two levels for each of the four factors. The bolded part of the examples shows the relevant level, e.g. the difference between having a plural or a singular nominal that contains *sin*.

Table 5.1: Linguistic variables, 2⁴ combinations

# of DP that contains <i>sin</i>	plural	Forældrene henter sine børn . The parents collect REFL's children.
	singular	Forældrene henter sit barn . The parents collect REFL's child.
Complexity of sentence	simplex	Forældrene henter sit barn . The parents collect REFL's child.
	complex	Forældrene husker at hente sit barn . The parents remember to collect REFL's child.
Type of subject	pronoun	De henter sit barn. They collect REFL's child.
	full noun	Forældrene henter sit barn. The parents collect REFL's child.
Animacy	animate	Forældrene henter sit barn. The parents collect REFL's child.
	inanimate	Træerne mister sin farve. The trees lose REFL's colour.

The combination of the four independent variables yield 16 unique sentence structures or 16 *conditions*. Together, the 16 conditions make up an *item*. An example of a full item is shown in table 5.2.

In order to increase the reliability of the acceptability judgments (this is one among several suggestions for improving reliability of acceptability judgments of reflexives made by Strahan (2011)), each condition is repeated once in the experiment. Each item is shown once in the questionnaire so in order to provide two examples of each of the 16 conditions, I created 32 items. The items are all unique in terms of lexical words but (attemptedly) identical in terms of grammatical structure according to the

Table 5.2: An experimental item, the four factors that combine to form the 16 different conditions

Factor	Levels (binary)	
# of DP that contains <i>sin</i>	Plural:	0XXX
	Singular:	1XXX
Complexity of sentence	Simplex:	X0XX
	Complex:	X1XX
Type of subject	Pronoun:	XX0X
	Full noun:	XX1X
Animacy	Animate:	XXX0
	Inanimate:	XXX1

16 different conditions.

A questionnaire is made by randomly selecting two instances of each of the 16 conditions in a way so that each of the 32 items provides one and only one sentence. These 32 sentences are the **primary** sentences. All 32 items can be found in appendix C.

Secondary sentences

The **secondary test items** are 30 sentences with various other kinds of reflexive structures. The secondary test items are the same across all four questionnaires. The purpose of these secondary test items is to test whether the acceptability of the primary test items correlates with the extent that the respondents accept other standard and non-standard uses of reflexives. Among the secondary test items are also examples of what can be analyzed as hypercorrected reflexives, i.e. a reflexive pronoun used in a context where the non-reflexive pronoun is required.

All the 30 secondary sentences with mean ratings from the experiment can be found in the tables in appendix C.3 on page 391. One example from each category is also given in table 5.4.

The secondary sentences are divided into categories with two or more example sentences. *Local sin* consists of sentences that contain a possessive (either a reflexive or a non-reflexive) and which obey the standard rules of binding. This means that the sentences either contain a locally bound *sin* or a locally free non-reflexive singular possessive. The ratings given to this category are predicted to be high. *Local hans* contains two sentences with a locally bound non-reflexive pronoun. Standard Danish

Table 5.3: An experimental item, examples of all 16 different conditions and a four digit place-code that indicates which factor levels that are used in the condition. A code of 0000 means a sentence with a plural DP that contains *sin* (0XXX), a simple sentence (00XX), and a pronoun subject (000X) that is also (intended to be) animate (0000).

Sentence	Condition code
De strækker sine vinger i solen. Gloss: they stretch REFL's wings in sun.DEF	0000
De strækker sine grene ud over åen. Gloss: they stretch REFL's branches out over stream.DEF	0001
Hønsene strækker sine vinger i solen. Gloss: hens.DEF stretch REFL's wings in sun.DEF	0010
Træerne strækker sine grene ud over åen. Gloss: trees.DEF stretch REFL's branches out over stream.DEF	0011
De løber ud for at strække sine vinger i solen. Gloss: they run out for to stretch REFL's wings in sun.DEF	0100
De får lys ved at strække sine grene ud over åen. Gloss: they get light by to stretch REFL's branches out over stream.DEF	0101
Hønsene løber ud for at strække sine vinger i solen. Gloss: hens.DEF run out for to stretch REFL's wings in sun.DEF	0110
Træerne får lys ved at strække sine grene ud over åen. Gloss: trees.DEF get light by to stretch REFL's branches out over stream.DEF	0111
De strækker sit næb ud efter maden. Gloss: they stretch REFL's beak out after food.DEF	1000
De strækker sin krone ud over åen. Gloss: they stretch REFL's crown out over stream.DEF	1001
Hønsene strækker sit næb ud efter maden. Gloss: hens.DEF stretch REFL's beak out after food.DEF	1010
Træerne strækker sin krone ud over åen. Gloss: trees.DEF stretch REFL's crown out over stream.DEF	1011
De løber hen for at strække sit næb ud efter maden. Gloss: they run over for to stretch REFL's beak out after food.DEF	1100
De får lys ved at strække sin krone ud over åen. Gloss: they get light by to stretch REFL's crown out over stream.DEF	1101
Hønsene løber hen for at strække sit næb ud efter maden. Gloss: hens.DEF run over for to stretch REFL's beak out after food.DEF	1110
Træerne får lys ved at strække sin krone ud over åen. Gloss: trees.DEF get light by to stretch REFL's crown out over stream.DEF	1111

Table 5.4: One secondary sentence from each category

Subcategory	Example	Mean rating
Local <i>sin</i>	Hun hentede sine pakker på posthuset. She collected REFL's parcels at post-house.DEF	4.60909
Local <i>hans</i>	Magnus havde en kat med et fjollet navn. Nu vil han kalde hans nye kat det samme. Magnus had a cat with a silly name. Now will he name his new cat the same.	4.07818
No binder	* Det var sin underbo, der gjorde Peter sur. It was REFL's downstairs-neighbour who made Peter mad.	1.88909
Non-local <i>sin</i>	Hun bad mig passe hendes kat. She asked me look-after her cat.	4.49818
DP-spec binder	Vi elsker Magnus' tegning af sin kat. We love Magnus' drawing of REFL's cat.	3.89454
<i>sig</i> good	Hun bad mig hjælpe sig med lektierne. She asked me help REFL with homework.PL.DEF	2.29636
<i>sig</i> bad	Hun bad mig hjælpe sig selv med lektierne. She asked me help REFL self with homework.DEF	1.77090

supposedly disallows this usage but it is a standard feature in a number of traditional Jutlandic dialects (see more in section 2.4 from page 82 and onwards). Ratings, consequently, are predicted to vary from high to low with a great deal of inter-speaker variation. *No binder* contains sentences with *sin* in various non-bound configurations. They are all predicted to receive low ratings. The category *non-local sin* consists of two sentences with non-local binding, one with a bound pronominal and one with a bound reflexive. Both sentences are grammatical but the complexity of the syntax leads me to expect ratings in the middle of the scale (given the results argued in e.g. Fanselow and Frisch (2006) that complex sentences tend to be rated lower just by virtue of their higher complexity and associated higher processing costs). The category *DP-spec binder* contains sentences with either a reflexive or a pronominal bound by a specifier in DP-spec rather than the standard clausal subject. The sentence with the supposedly bound pronoun, *Jeg så Peters fotografi af hans forældre og var meget imponeret.*, is slightly problematic in that participants are not explicitly instructed to read *Peter* and *hans* as coreferent, and so it is not possible to be certain whether participants rated on the basis of this intended binding configuration. The clausal subject

in all three sentences is a first or second person pronoun in order to let the specifier in DP-spec be the only possible binder. This precaution is inspired by Keller and Asudeh (2001) whose experimental results indicate that DPs with possessors in the specifier may very well be binding domains (at least in English) but also that they are not impenetrable to binding by the clausal subject. The overall expectation for this category is that participants would rate the sentences overall as acceptable with a penalty for the complexity of the sentence. The category *sig good* consists of a handful of sentences with non-possessive reflexives and pronominals. They are all grammatical and the prediction is that they will all be rated as acceptable. The category *sig bad* consists of a handful of sentences with non-possessive reflexives and pronominals that are all constructed to be ungrammatical in standard Danish. The ratings for this category are expected to be low overall.

Filler sentences

The **fillers** are 30 sentences that vary from 1 to 5 in acceptability, in an attemptedly even spread. None of the fillers contain reflexive structures. The filler sentences are the same across all four questionnaires. The sentences themselves, and the approximated acceptability judgments used to select them, are borrowed from the Danish part of The Nordic Dialect Database (Lindstad et al. 2009). I searched the database for sentences with a typical rating of 1 (and 2, 3, 4, and 5, respectively) and selected six sentences from each acceptability category. I provide an example of a filler from each of the five acceptability categories in (234). The number to the right of the example indicates the intended acceptability level (on a scale from 1 to 5).

- (234) a. Guldmedaljer har han mange. (Intended rating: 1)
 gold-medals has he many
 'Gold medals has he many.'
- b. Gæsten blev bagt en kage. (Intended rating: 2)
 guest.DEF was baked a cake
 'The guest was bought a cake.'
- c. Jeg har kufferterne pakket. (Intended rating: 3)
 I have suitcases.DEF packed
 'I have the suitcases packed.'
- d. Jeg fik min cykel stjålet. (Intended rating: 4)
 I got my bike stolen
 'I got my bike stolen.'

- e. Bo havde aldrig læst bøgerne. (Intended rating: 5)
 Bo had never read books.DEF
 'Bo had never read the books.'

The selection seems to have worked fairly well since the average rating of all the fillers is around 3 (the middle of the scale) across the final data set. Respondents use the whole scale from 1 to 5 to judge the fillers.

"Awake" sentences

All four questionnaires contain the four "**awake**" items. Of these, two are perfectly acceptable, simple Danish sentences. The other two are complete gibberish, syntactically speaking, but with Danish words. The specific sentences are given in (235).

- (235) a. Mia elsker is.
 Mia loves icecream
 'Mia loves icecream.'
- b. Børnene spiller bold i frikvarteret.
 Kids.DEF play ball in break.DEF
 'The kids play ball in the break.'
- c. * Isen på glammer af grønlig.
 Ice.DEF on barks(V) of greenish
 'The ice on barks of greenish.'
- d. * At købte i morges vi.
 To bought in morning we
 'To bought in morning we.'

The intention behind including these is to catch those respondents who simply click through the questionnaire at random, or who intentionally judge sentences oddly. I exclude respondents who do not judge the bad "awake" items as 1 or 2 as well as respondents who do not judge the good "awake" items as 4 or 5.

The setup of the data collection does make it likely that a number of participants would complete the questionnaire without actually engaging with the sentences. The participants did not technically volunteer to participate themselves, as it were their teachers that choose to volunteer them. There was no direct motivation for the students to actually finish the questionnaire, and I do not know how many started the questionnaire and gave up along the way. However, in order to find those people who did actually finish the questionnaire by clicking through at random, the "awake" sentences are an easy litmus test. I have included a bit of leeway for misclicks and similar

issues in the "awake" exclusion criteria, so that a participant is only considered for exclusion if he/she gives odd ratings to at least two of the four "awake" sentences. The expectation is that everyone should rate the good "awake" sentences as 4 or 5 and the bad "awake" sentences as 1 or 2. The participants who rated differently for two or more "awake" sentences were identified with a few lines of R code and excluded from the data set.

Training sentences

The first four sentences in all four questionnaires are training items. The intention behind these is that participants get used to the format of the study before they engage with the main material. The participants do not get feedback on their judgments of the training items and are not told explicitly that the first items are intended as training. I do not use the judgments of the training items in the analysis or visualizations.

The four training sentences are shown in (236). (236a) is intended to be fully acceptable. (236b) is intended to be fully unacceptable. (236c) and (236d) fall somewhere in the middle of the scale, neither fully unacceptable nor fully acceptable.

- (236) a. De køber en sofa i genbrugsbutikken.
They buy a couch in charity-shop.DEF
'They buy a couch in the charity shop.'
- b. Tastaturet ikke larmer når man bruger det.
Keyboard.DEF not make-noise when one uses it
'The keyboard not makes noise when one uses it.'
- c. I går har jeg kage med fra arbejde.
I yesterday have I cake with from work
'Yesterday I bring cake from work.'
- d. Manden som at der gik på vejen havde en hund med.
Man.DEF who that who walked on road.DEF had a dog with
'The man who that walked on the road had a dog with him.'

5.3 The final data set

The full data set consists of acceptability judgment data from 593 people studying at various Danish gymnasiums. The 593 students come from 16 different gymnasiums across Denmark and one gymnasium in Greenland. The data from the gymnasium in Greenland was not included in the final data analysis, making the final data set con-

sist of acceptability judgment data from 579 students from 16 Danish gymnasiums. See table 5.5 for the list of included gymnasiums and fig. 5.3 for a map of where the gymnasiums are placed in Denmark.

Table 5.5: Gymnasiums in the acceptability study

Gymnasium	No. of participants
Jylland	
Brønderslev	35
Esbjerg (HTX)	26
Herning	29
Kolding	49
Nykøbing Mors	46
Randers (HTX)	14
Ringkøbing (HHX)	54
Sønderborg (HTX)	31
Aabenraa	47
Aabenraa (HTX)	9
Aarhus	48
Fyn	
Odense	63
Sjælland	
Roskilde	51
Copenhagen (Rysensteen)	19
Rødovre	39
Slagelse	19
Greenland (excluded <i>a priori</i>)	
Aasiaat	14
All participants	593
Number of gymnasiums	17 (16 included)

The gymnasiums included in the study were from the offset selected with the intention of having as broad a sample as possible, primarily in terms of geographical and linguistic region and whether or not the gymnasium could be said to be an urban gymnasium or a rural gymnasium. The final range of gymnasiums in the study is for practical reasons less representative than would be ideal: The data is fairly dependent on where I or others in my professional network happened to personally know gymnasium teachers who were willing to let their students take part in the study. The gymnasiums included in the study are for the most part *STX gymnasiums* (11 out of 16, the most general type of *gymnasium*). Four gymnasiums are *HTX gymnasiums* (gymnasium with a focus on the technical and natural sciences) and one is an *HHX*

gymnasium (gymnasium with a focus on business economics and marketing). When no additional information is provided in the table, the gymnasium is an STX.



Figure 5.3: Map of Denmark with city names for the included gymnasiums and linguistic regions.

5.3.1 Background information provided by the study participants

Each participant is asked to fill in a number of demographic details on the first page of the questionnaire. I briefly go through these categories and how the answers are

distributed across the data. The numbers provided are from the full data set *without* the data from the 14 participants from Aasiaat, Greenland.

Age is a multiple choice list with options ranging from 15 to 20 in one-year increments, another option which is the broader 21-25, and finally an "Other/No response" option. The distribution of answers in the data set is shown in the barplot in fig. 5.4. The great majority of respondents fall within the age range between 16 and 19.

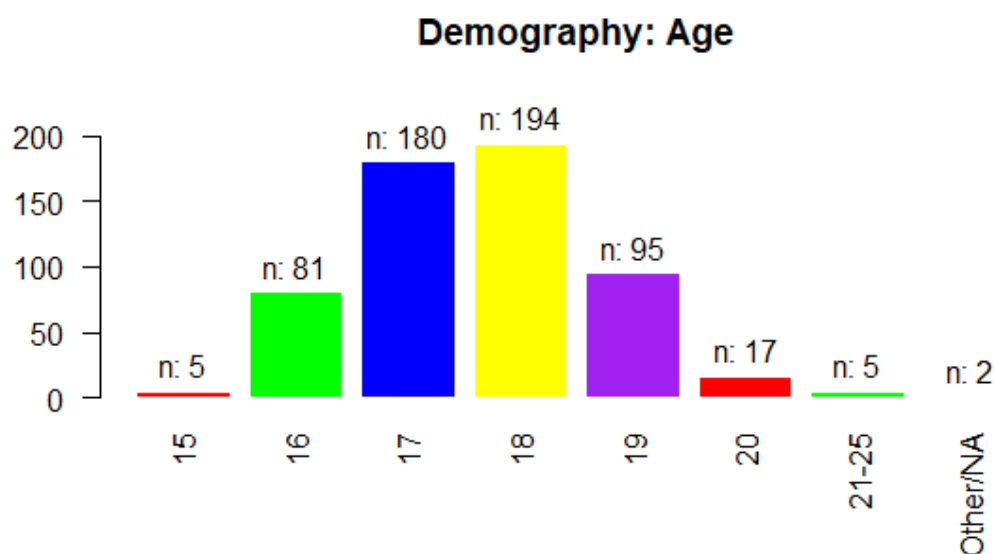


Figure 5.4: Barplot of distribution across the various age groups.

In the category **region**, respondents were presented with the map in fig. 5.5. They were asked to choose the region where they had lived the longest from the seven regions on the map, or to alternatively pick "Other". The specific lines drawn on the map for Jutland are a simplified version of the linguistic splits noted by *Jysk Ordbog* on their *Atlaskort K 7.1* (reproduced as fig. 5.6 on page 248). Regions 1 and 3 on the map in fig. 5.5 correspond to the areas in Western and Southern Jutland that housed traditional dialects with prenominal definite articles. Eastern and Northern Jutland (region 2 on the map in fig. 5.5) had postnominal articles, just like the rest of Denmark. I discuss in section 2.4 on page 82 why this particular traditional dialectal split in Jutland could be relevant to the topic at hand. The reason, in short, is that languages with prenominal articles tend to not have reflexive possessive pronouns, which gives a very different reflexive system for the traditional dialects in these parts of Jutland. Funen is given as a relevant dialect area in itself, region 4. Zealand is split in two: Re-

gion 6 which contains Copenhagen and the surrounding area (corresponding to the administrative area *Region Hovedstaden*) and region 5 is the rest of Zealand. The island Bornholm is region 7 but I did not manage to get any participants from there. Finally, participants had the option of choosing "Andet" (*Other*), which only one person did.

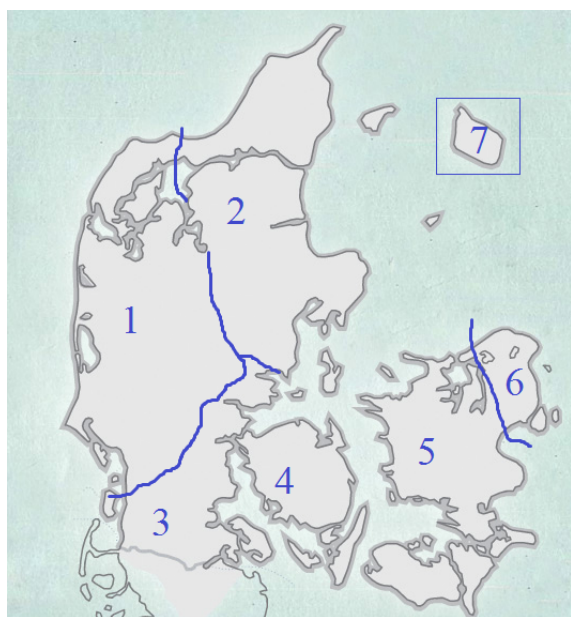


Figure 5.5: Map of Denmark with numbered linguistic regions.



Figure 5.6: Atlaskort K.7 from Jysk Ordbog. *Foranstillet artikel* means *prepositional article* and *efterhængt artikel* means *postnominal article*.

The number of participants are distributed somewhat unevenly across the seven regions. The distribution is shown as a barplot in fig. 5.7. Overall, the majority of participants (387 or 66.8 %) are from Jutland, 10.5 % are from Funen, 13.8 % are from Zealand, and 8.6 % are from the Copenhagen area. The large Jutland majority is likely an artefact of the sampling method which draws heavily on my personal (and largely Jutland-based) gymnasium network.

Participants are asked to choose their preferred **gender** category from three options: Boy, girl, and other. 237 participants (40.9 %) identify as boys, 341 participants (58.9 %) identify as girls, and one person (0.2 %) chose Other/NA. This corresponds

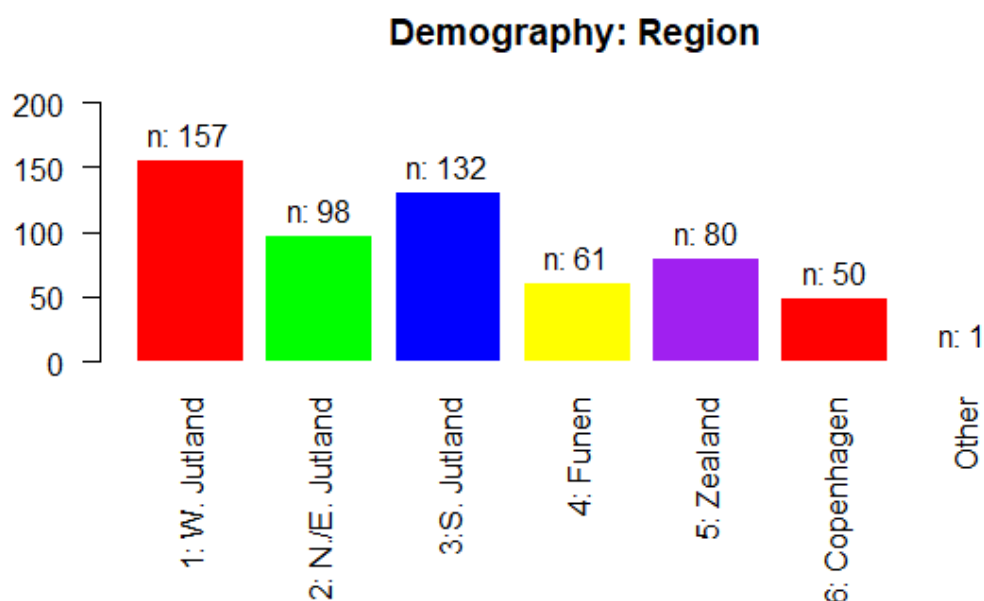


Figure 5.7: Barplot of participant distribution across the various regions.

closely to the overall gender distribution in STX gymnasiums².

The majority of the gymnasiums in the study are of the general type, STX. One gymnasium, Ringkøbing, is an HHX (a business gymnasium), and four gymnasiums are HTXs (STEM gymnasiums, Esbjerg, Randers, Sønderborg, Aabenraa HTX). The HHX and HTX gymnasiums differ from the STX gymnasiums in their gender distribution both in my sample and in the general gymnasium population. Figure 5.9 illustrates the gender distribution across the three gymnasium types in the data set (which aligns closely with the gender distribution in the general gymnasium population): The one HHX has 61.1 % boys and 38.9 % girls, the HTX's have 78.75 % boys and 21.25 % girls, and the STX's essentially have the reverse gender distribution with 31.7 % boys and 68.1 % girls. Since the majority of the gymnasiums in the data set are STX's, the overall gender distribution mirrors the STX gymnasiums in having a majority of girls.

The participants were asked to note whether **Danish is their first language**. "First language" may be defined in a number of ways. The specific wording of the questionnaire ("Er dansk dit modersmål?", En. "Is Danish your native language?") leaves the specific definition open, but the three options ("Yes", "Yes, among others", "No") allows the possibility of having more than one first language. The distribution is plotted

²61 % girls and 39 % boys, according to a 2017 report from the Danish Ministry of Children and Education, <https://shorturl.at/eoDQ4>

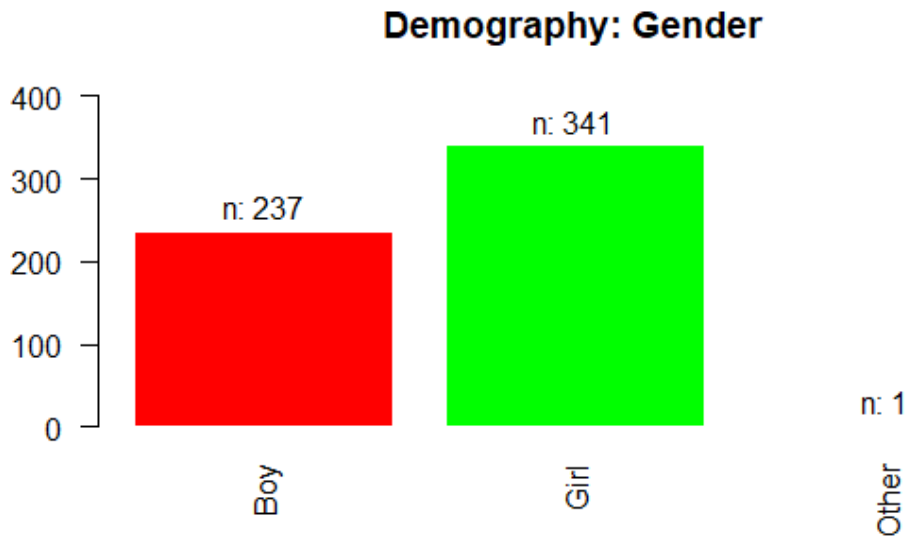


Figure 5.8: Barplot of distribution across the three gender options.

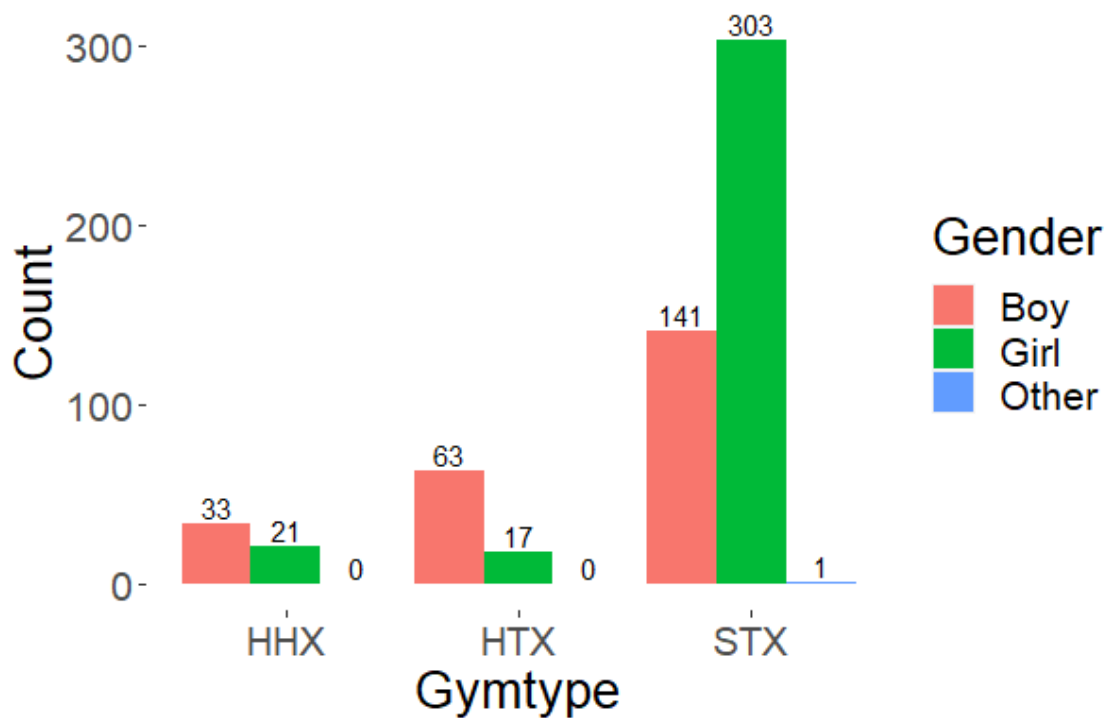


Figure 5.9: Barplot of gender distribution in the three types of gymnasium.

in fig. 5.10. 512 participants (88.4 %) have Danish as their only first language, 50 participants (8.6 %) have Danish as one first language among others, and 17 participants

(2.9 %) do not have Danish as their first language. The participants who answered "No" or "Yes, among others" are asked to note their other first language(s) in a follow-up text box. 27 different first languages³ are represented in the data set apart from Danish. The most frequent (other) first languages are English (11 participants), Turkish (9 participants), Arabic (7 participants), and German (5 participants).

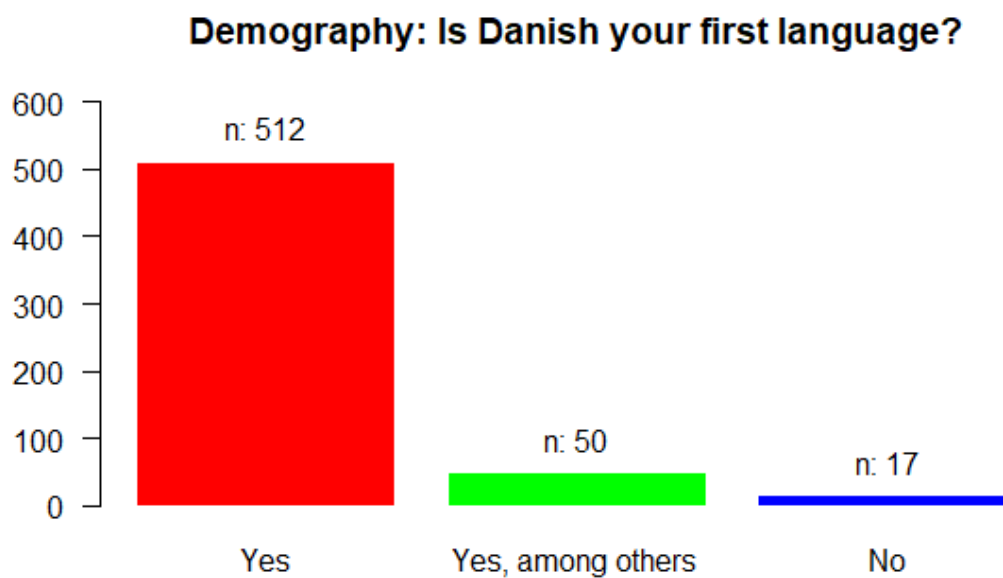


Figure 5.10: Barplot of distribution of respondents with Danish as their first language; Danish as one among other first languages; Danish as something other than a first language.

The questionnaire contains three questions that together form a picture of whether the participants have spent most of their life in **a more rural or a more urban** setting. Participants are asked to write down which *byer* (En. 'villages/towns/cities') they have lived in (or alternatively, to write that they live(d) in the countryside or abroad). Furthermore, they are asked to write down the place that they have lived the longest, and finally where they went to *folkeskole* (En. equivalent: Primary and lower secondary school). I condense these three questions into a single codeable category for doing the statistical analysis. The definitions of the various places of residence used in the coding process are provided in table 5.6. The categories themselves are adapted from

³Full list of other first languages, named as the participants themselves do and translated into English: Amharic (Ethiopian), Arabic, Azerbaijani, Bosnian, Bulgarian, Dari, English, Filipino, French, Faroese, German, Greenlandic/Kalaallisut, Iranian, Kurdish, Lithuanian, Montenegrin, Norwegian, Pashto, Persian/Farsi, Polish, Portuguese, Russian, Serbian, Somali, Spanish, Turkish, Vietnamese.

Table 5.6: Definition of places of residence used in the data analysis (coding categories)

Place of residence	Population size	Share of full pop. (2021, %)	Share of data set (%)
Countryside	<200	11.75	6.73
Village	200-1,000	6.74	8.80
Small town	1,000-10,000	21.57	30.74
Medium town	10,000-49,999	19.87	15.03
Large town	50,000-100,000	7.15	16.58
City	>100,000	32.93	19.86
Other	Unclear	-	2.25

Danmarks Statistik (a state institution, the main authority on statistics about Danish society) and their publication *Byopgørelsen*⁴, an annual census report. Participants were not introduced to these simplified categories but were asked the more unrestricted questions described above.

Most participants note only one place of residence. In the cases where a participant has noted more than one place, questions 2 and 3 (where they have lived the longest and where they went to *folkeskole*) help in determining how to code the participant. Most often, the place where they went to *folkeskole* and the place that they have lived the longest match, in which case the participant will be coded as belonging to the category that matches that particular place. In a few cases, I have not been able to determine the relevant category, even with the help of questions 2 and 3. In these cases the participant is coded as *Other* together with the participants who indicated that they had lived abroad the longest.

The final demographic question in the questionnaire asks participants about the educational background of their parents. This question is included as a measure of **socio-economic status (SES)**. The distribution of answers is illustrated in fig. 5.12. Participants were provided with a multiple choice list with seven general categories, reproduced in full in table 5.7 with English translations added. Each of the categories (apart from *Folkeskole* and *Gymnasium*) were supplemented with two examples of specific job titles that could fit into the particular level of education. The question as it is worded in the questionnaire does not specify whether participants should input every education for each parent, or just the highest one. Some participants have entered just one education, and some participants have entered an entire list, beginning

⁴<https://www.dst.dk/da/Statistik/dokumentation/statistikdokumentation/byopgoerelsen/indhold>

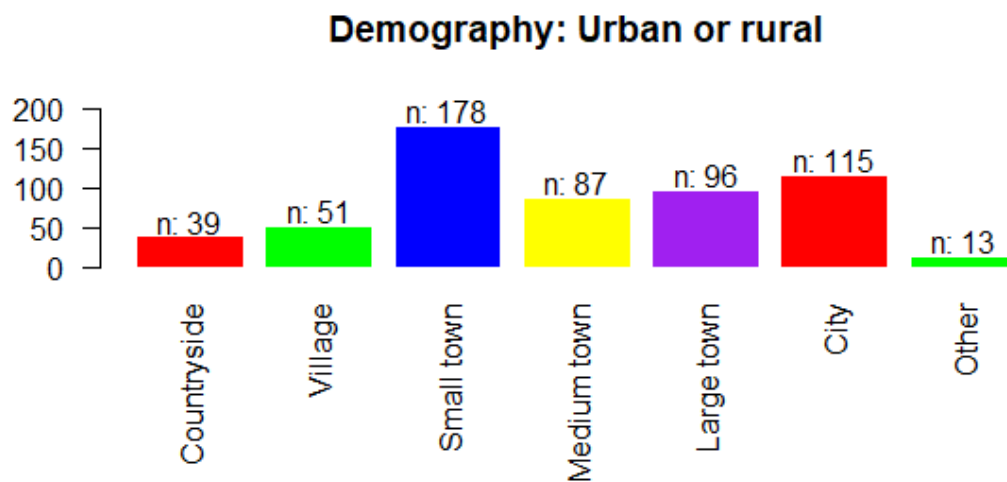


Figure 5.11: Barplot of distribution across the respondents' place of living (more rural or more urban).

with *Folkeskole*. In the latter cases, I code the response as the highest listed educational level in the participant's answer in order to only have one response for each participant. In some cases, a response will include e.g. both *Folkeskole* and *Other/NA*. These are coded as the highest listed educational level that is not *Other/NA*. The only responses that are coded as *Other/NA* are those where *Other/NA* is the only provided answer.

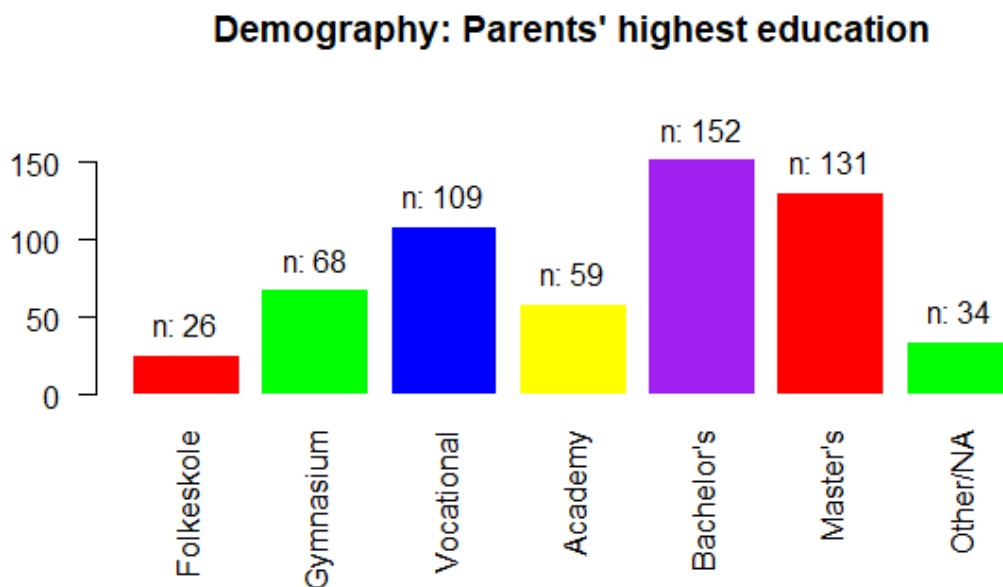


Figure 5.12: Barplot of distribution across highest achieved educational level among respondents' parents.

5.3.2 The full data set, background information summarized

The demographic information for all the participants is summarized in the tables on the following pages. The table is split according to the 16 gymnasiums.

Table 5.7: Parents' highest education: Description of categories

Danish term	Danish example
Folkeskole	
Gymnasial uddannelse	STX, HTX, HHX, HF
Erhvervsfaglig uddannelse	F.eks. murer eller lægesekretær
Kort videregående uddannelse	2-3 år, f.eks. tandplejer eller datamatiker
Mellemlang videregående uddannelse	3-4,5 år, f.eks. folkeskolelærer eller sygeplejerske
Lang videregående uddannelse	5-6 år, f.eks. en universitetsgrad
Ved ikke/Vil ikke svare	
English translation of term	English translation of examples
Primary and lower secondary education	
Upper secondary education	
Vocational education and training	E.g. mason or medical secretary
Academy Profession	E.g. dental therapist or information technologist
Bachelor's degree	E.g. primary school teacher or nurse
Master's degree	E.g. a university degree
Other/NA	

5.3.3 Data exclusion criteria and outliers

The full, final data set contains data from 593 gymnasium students. I quality checked the data in the early phases of the statistical analysis, which led me to exclude 43 participants (7.3 % of the full data set). The data set used for the statistical analyses consequently consists of data from the remaining 550 gymnasium students.

The 43 excluded participants includes all 14 students from Aasiaat in Greenland and 29 students from other gymnasiums. The 14 students from Aasiaat were all excluded *a priori* as they are too heterogenous as linguistic group from the rest of the surveyed students. 8 out of 14 of the students do not identify as native Danish speakers and the 6 others are Kalaallisut-Danish bilingual.

29 other students from different gymnasiums were removed from the data set. 25 of these rated the "awake" sentences in a way that suggests that the person might have simply clicked through the questionnaire at random. Four other participants were excluded as outliers later in the process, first identified on the basis of their filler ratings as their "awake" ratings were "normal enough" to avoid automatic detection. The ratings from five participants showed up as outliers on boxplots of the filler ratings for the data set as a whole. Four of these five participants were excluded as their ratings

Table 5.8: Full, unsimplified, data set (background variables, alphabetically according to gymnasium, continued in (5.9) and (5.10))

No. of students (N)	Brønderslev	Esbjerg	Herring	Kolding	Morsø	Odense	Randers	Ringkøbing
35	26	29	49	46	63	14	54	
Gender								
Boy (N =)	13	23	3	15	10	26	7	33
Girl (N =)	22	3	26	34	35	37	7	21
Other (N =)	0	0	0	0	1	0	0	0
Danish first language								
Yes (N =)	34	25	28	43	41	53	14	50
Yes+ (N =)	1	0	1	5	4	8	0	3
No (N =)	0	1	0	1	1	2	0	1
Region								
1 (N =)	0	26	28	2	45	2	0	53
2 (N =)	33	0	0	1	1	0	14	0
3 (N =)	2	0	0	45	0	1	0	0
4 (N =)	0	0	0	0	0	60	0	0
5 (N =)	0	0	1	0	0	0	0	0
6 (N =)	0	0	0	1	0	0	0	0
8 (N =)	0	0	0	0	0	0	0	1
Place of residence								
Countryside (pop. <200)	5	4	2	2	14	0	2	1
Village (pop. 200-1,000)	11	4	2	2	5	2	2	11
Small town (pop. 1,000-10,000)	15	3	15	5	20	6	2	40
Medium town (pop. 10,000-49,999)	1	5	0	0	1	1	0	0
Large town (pop. 50,000-100,000)	0	10	10	36	0	1	7	1
City (pop. >100,000)	1	0	0	3	0	53	1	0
Other (pop. unclear)	2	0	0	1	6	0	0	1
Parents SES								
1: Folkeskole	4	2	1	1	1	0	1	4
2: Gymnasium	0	6	1	5	8	5	0	11
3: Vocational	12	5	7	7	15	7	4	11
4: Academy	6	3	1	2	4	2	3	9
5: Bachelor's	10	6	13	14	16	21	4	12
6: Master's	1	3	6	15	1	26	1	4
7: Other/NA	2	1	0	5	1	2	1	3

Table 5.9: Full, unsimplified, data set (background variables, alphabetically according to gymnasium, contd. from (5.8) and in (5.10))

No. of students (N)	Roskilde	Rysensteen	Rødovre	Slagelse	Sønderborg (HTX)	Aabenraa (HTX)
51	19	39	19	31	9	9
Gender						
Boy (N =)	23	7	12	3	26	7
Girl (N =)	28	12	27	16	5	2
Other (N =)	0	0	0	0	0	0
Danish first language						
Yes (N =)	40	16	34	15	27	7
Yes+ (N =)	7	3	4	2	2	2
No (N =)	4	0	1	2	2	0
Region						
1 (N =)	0	0	0	0	0	0
2 (N =)	1	1	0	0	0	0
3 (N =)	0	0	0	0	28	9
4 (N =)	0	0	0	1	0	0
5 (N =)	49	1	12	17	0	0
6 (N =)	1	17	27	1	3	0
8 (N =)	0	0	0	0	0	0
Place of residence						
Countryside (pop. <200)	1	0	0	0	4	1
Village (pop. 200-1,000)	6	0	0	1	1	2
Small town (pop. 1,000-10,000)	19	2	0	3	19	2
Medium town (pop. 10,000-49,999)	5	3	25	14	6	4
Large town (pop. 50,000-100,000)	20	1	8	0	0	0
City (pop. >100,000)	0	13	5	1	0	0
Other (pop. unclear)	0	0	1	0	1	0
Parents SES						
1: Folkeskole	1	0	3	1	3	0
2: Gymnasium	11	0	5	3	6	0
3: Vocational	7	0	8	4	6	2
4: Academy	3	0	5	5	5	1
5: Bachelor's	16	3	8	3	7	2
6: Master's	10	16	6	2	2	2
7: Other/NA	3	0	4	1	2	2

Table 5.10: Full, unsimplified, data set (background variables, alphabetically according to gymnasium, contd. from (5.8) and (5.9))

	Aabenraa (STX)	Aarhus
No. of students (N)	47	48
Gender		
Boy (N =)	11	18
Girl (N =)	36	30
Other (N =)	0	0
Danish first language		
Yes (N =)	42	43
Yes+ (N =)	3	5
No (N =)	2	0
Region		
1 (N =)	1	0
2 (N =)	0	47
3 (N =)	46	1
4 (N =)	0	0
5 (N =)	0	0
6 (N =)	0	0
8 (N =)	0	0
Place of residence		
Countryside (pop. <200)	3	0
Village (pop. 200-1,000)	2	0
Small town (pop. 1,000-10,000)	23	4
Medium town (pop. 10,000-49,999)	18	4
Large town (pop. 50,000-100,000)	0	2
City (pop. >100,000)	0	38
Other (pop. unclear)	1	0
Parents SES		
1: Folkeskole	1	3
2: Gymnasium	4	3
3: Vocational	7	7
4: Academy	8	2
5: Bachelor's	14	3
6: Master's	9	27
7: Other/NA	4	3

almost solely consisted of 1's and 2's or 5's with no seeming relation to the rated sentences. One person among the four excluded had ratings that started out "normal" for the first sets of sentences, after which he/she may have lost patience and rated the rest of the sentences as 5. The overall ratings of the fifth participant identified in this way did not give reason for exclusion, and this person's data was kept in the data set.

5.4 Statistical models and model control

The remaining parts of this chapter walk through a statistical analysis of the data from the acceptability study on *sin* with plural antecedents described in the preceding sections. The statistical analyses are tailored to answer the following four research questions:

1. Is there a difference in the rating given to the first and second instance of the same grammatical structure, shown with different lexical content?
2. To what extent is the acceptability of *sin* with plural antecedent dependent on morphosyntactic factors such as the number of the object, the type of subject, and the sentence structure?
3. To what extent is the acceptability of *sin* with plural antecedent dependent on extralinguistic factors such as region, age, socio-economic status of parents, gender, and place of living (urban or rural)?
4. To what extent is the acceptability of *sin* with plural antecedent dependent on otherwise non-standard use of reflexives?

I use R (R Core Team 2021, versions 4.0.4-4.1.0) for all statistical analyses and data visualisations. I begin by outlining the statistical models that I describe my data with. An important part of this step is the model control, where I verify that I should be able to use the chosen inferential tests and get results out that are trustworthy. The conceptual point here is that if the assumptions behind a statistical model are not met, a statistical test based on the model with the un-met assumptions will not be valid.

5.4.1 Research question 1: Statistical model and model control

Statistical model (RQ1)

There are 16 different primary sentence structures in the questionnaires (see more on page 236). All 16 primary sentence structures are repeated once which gives two sets of 16 primary sentences or 32 primary sentences all in all. Participants in the study see and rate the same sentence structure twice within a questionnaire. The first research question asks whether the rating given to the 16 sentence structures in the first set is the same as the rating given to the same 16 sentence structures in the second set. I compare the ratings pairwise so that the rating comparison is between the 16 pairs of identical sentence structures.

I call the two groups group 1 and group 2. Group 1 is the mean rating given to the first instance of the 16 different sentence structures across the 16 gymnasiums. This sums up to $16 \times 16 = 256$ mean values. Group 2 is the mean rating given to the second instance of the 16 different sentence structures across the 16 gymnasiums. This also sums up to $16 \times 16 = 256$ mean values. I analyze the data with a paired t-test that tests whether there is a significant difference in means between group 1 and group 2. In formal terms, this corresponds to testing the hypothesis H_0 on the statistical model specified in (237).

(237) Mean rating for group 1: $X_i, i = 1, \dots, 256$,
 Mean rating for group 2: $Y_i, i = 1, \dots, 256$, and

$$\Delta_i = X_i - Y_i \sim N(\mu, \sigma^2)$$

$$H_0 : \mu_1 = \mu_2,$$

equivalent to

$$H_0 : \mu_1 - \mu_2 = 0$$

where μ_1 is the mean of group 1 and μ_2 the mean of group 2.

The notation

$$\Delta_i = X_i - Y_i \sim N(\mu, \sigma^2)$$

requires some unpacking. Δ_i is the difference in rating between data point X_i and data point Y_i , i.e. between the ratings for the first and second instance of the same structure. I specify the with the notation $N(\mu, \sigma^2)$ that these differences are distributed

according to a normal distribution with mean value μ and variance σ^2 . This type of model comes with a number of assumptions that should be checked in order to make sure that the chosen type of test will give a valid description of the data. This is the topic of the next section.

Model control (RQ1)

There are three assumptions that must be checked in order to make sure that a standard t-test can actually be performed on the data. The assumptions all pertain to the differences in means between the two groups. The assumptions to be checked are *independence*, *equal variance*, and *normal distribution*.

I assume that the differences in means between the groups are independent, based on the design of the questionnaire. There is only one observation from each person for each data point and there are no (or surely very few) siblings or similarly close relationships between participants.

To test the assumption of equal variance, I plot the differences against the averages in the groups. This plot is called a Bland-Altman plot and is shown in figure 5.13 (originally from Bland and Altman 1986). The plot is a scatterplot with the average ratings on the x-axis and the differences between each pair of data points on the y-axis. The x-value for a point in the plot represents the average rating for the two instances of the same structure: $x = (\text{rating of first instance} + \text{rating of second instance}) / 2$. The y-value for a point in the plot represents the difference in rating: $y = \text{rating of first instance} - \text{rating of second instance}$. For the plot to support the assumption of equal variances for the differences, the data points should spread out around 0 and the variance (i.e. the distance from 0 on the plot) should be the same all along the x-axis. The data points do spread out around 0 and there are no obvious trends in how the points cluster depending on where they are on the x-axis (an obvious trend to look for is the trumpet shape where the variance rises with larger x-values – this is not what we see here). The plot, in other words, supports the assumption of equal variances.

The third and last model check is to confirm that the differences are normally distributed. Judging from the qqplot of the differences in fig. 5.14, with 95% confidence bands in the plot, this assumption can also be upheld. The term qqplot is an abbreviation of quantile-quantile plot. It plots the quantiles (essentially, the data points arranged according to size) of the differences along the y-axis and the quantiles for a



Figure 5.13: Bland-Altman plot of the differences between means in the two groups (first and second instance of the structure in the questionnaire) plotted against the averages of the groups. This plot indicates that it is reasonable to assume equal variances of the differences between the two groups.

sample from a normal distribution with the same parameters along the x-axis. If the observed values (the y-values) are distributed in a similar way to the simulated values (the x-values), the points will follow a straight line (as drawn on the plot). This particular plot has added confidence bands (the dotted lines). If the plotted values largely fall within these bands, we can assume that the fit between observed and simulated data is close enough and we can confirm the assumption of normal distribution. The points cluster closely around the straight line and every point is within (or on top of) the 95% confidence bands. The qqplot supports the assumption that the differences are normally distributed.

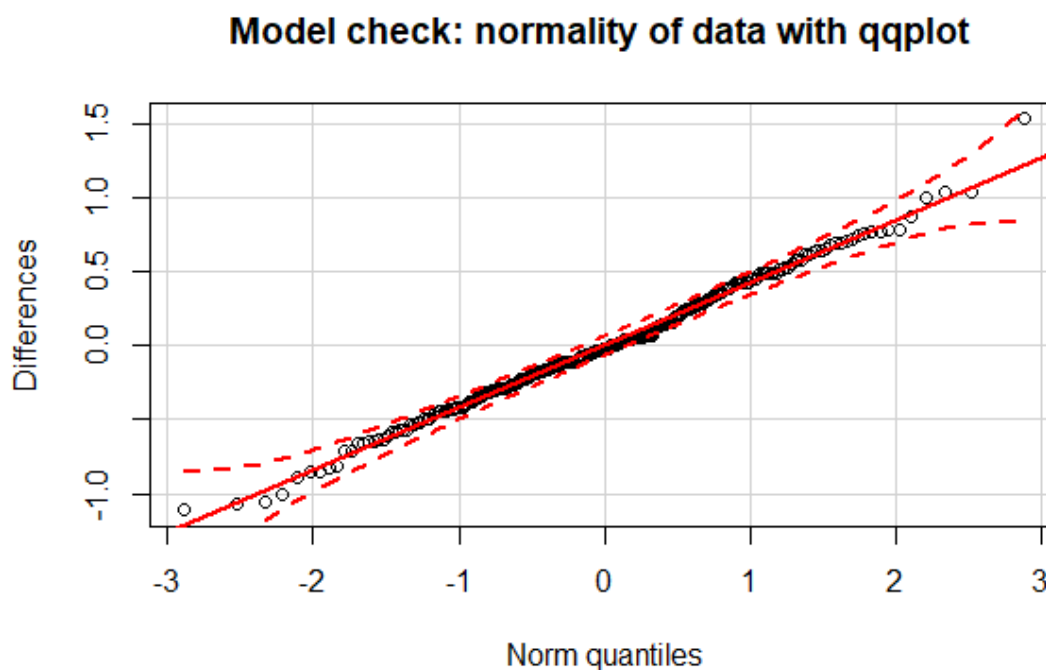


Figure 5.14: qqplot with 95% confidence bands of the differences between the two groups in the t-test. The linearity of the points indicates that it is reasonable to assume that the differences are normally distributed.

5.4.2 Research question 2: Statistical model and model control

Statistical model (RQ2)

2. To what extent is the acceptability of *sin* with plural antecedent dependent on morphosyntactic factors such as the number of the object, the type of subject, and the sentence structure?

To address RQ2, I fitted a linear model to the data and analyzed the data the model in R using analysis of variance (ANOVA). This type of analysis is appropriate for a data set with a continuous outcome variable and categorical explanatory variables. It compares the means in more than two groups. The method is called analysis of variance because it compares the variance within the groups (a *group* is e.g. the gymnasium) to the variance between the groups. The outcome variable is the mean rating given to the two instances of the 16 different primary structures for each of the 16 gymnasiums. The data set, which is the same as the one used to perform the t-test in the previous section, contains 32 outcome data points for each gymnasium, and the data

set with outcome variable and the various possible dependent variables is specified in table 5.11.

Table 5.11: Data set used to investigate research question 2

Variable	Data	Type	No. of data points
Outcome	Mean of primary ratings, 32 ratings per gymnasium	Continuous (mean)	16 * 32 = 512
Dependent	Gymnasium	Factor (16 levels)	
Dependent	Number (sg/pl)	Factor (2 levels)	
Dependent	Simple or complex	Factor (2 levels)	
Dependent	Subj. type (pronoun or full DP)	Factor (2 levels)	
Dependent	Animacy	Factor (2 levels)	
Dependent	Sequence	Factor (2 levels)	

I model my data as specified in (238).
(238)

$$Y_{nctagi} = \mu + \alpha_n + \beta_c + \gamma_t + \delta_a + \zeta_g + \beta\gamma_{ct} + \epsilon_{nctagi}$$

$$\epsilon_{nctagi} \sim N(0, \sigma^2)$$

Here, Y_{nctagi} is the mean rating of primary sentence i of object number n , sentence complexity c , subject type t , animacy a , gymnasium g , and ct is the interaction between sentence complexity and subject type. μ is the overall mean.

α_n is the effect of the number of the object.

β_c is the effect of sentence complexity.

γ_t is the effect of subject type.

δ_a is the effect of animacy.

ζ_g is the effect of gymnasium.

$\beta\gamma_{ct}$ is the effect of the interaction between sentence complexity and sentence type.

I specify the assumption that the residuals, ϵ_{nctagi} , are independent and identically distributed random variables, which I confirm in the Model control section below with a model check. The model specified here is the final result of a process of fitting linear models to the data. *Sequence*, which is specified as a possibly relevant dependent variable in table 5.11, is shown to be non-significant with the t-test in section 5.5.1 and the modelling process for this particular research question also indicated that it would be safe to leave out this factor from the final model.

Model control (RQ2)

For the linear model specified in (238), the model check is on the residuals. They are assumed to be identically (normal) distributed, independent random variables with equal variance. There is nothing to suggest that they are not independent: Every student makes one rating for each structure and there are presumably no (or at least very few) twins in the sample or anything else that could lead to non-independence. For the two other assumptions, I plot the standard graphs for model checking in the 2-by-2 plot in figure 5.15 and walk through them one by one.

The first plot in the upper left corner is the Residuals vs Fitted plot. This is a scatter plot (like the Bland-Altman plot described in section 5.4.1) where the residuals from the model are plotted on the y-axis and the *fitted values* on the x-axis. The residuals represent the deviance of the observed data points from the straight line that the data is modelled onto (in a linear model). The fitted values are simulated, normally distributed responses computed from the proposed linear model of the data. For the plot to support the assumption of equal variances, the data points should cluster randomly around 0 with no obvious non-linear trends (e.g. if the line around 0 had been a closer to a parabola) and no trumpet shapes. This is indeed what we see in the plot, and it consequently supports the assumption of equal variances.

The next plot, the Normal Q-Q in the upper right corner, checks for normal distribution of the residuals. The makeup of the plot is the same as the qqplot described in fig. 5.14 on page 263 (apart from the pedagogical confidence bands), and I refer the interested reader to that page for further details. The only difference is that the y-values represent the residuals here, rather than differences. The points cluster nicely around the straight line, which supports the assumption that the residuals are normally distributed.

The third plot, Scale-Location, is the same plot as the Residuals vs Fitted plot. The only difference is that the y-values are residuals that have been transformed (standardized and taken the square root of). The Scale-Location plot makes the same point, namely that we can assume equal variances.

The Cook's distance plot in the lower right corner shows the observations with the largest residuals, i.e. the greatest difference from the modelled straight line. If some of these are large, they could be outliers that should be evaluated. These are very small (compared to the rating scale from 1 to 5), however, so there is nothing in plot to suggest that there are suspicious outliers to take care of.

The plots, taken together, suggest that the assumptions of normality and equal variances of the residuals can be upheld.

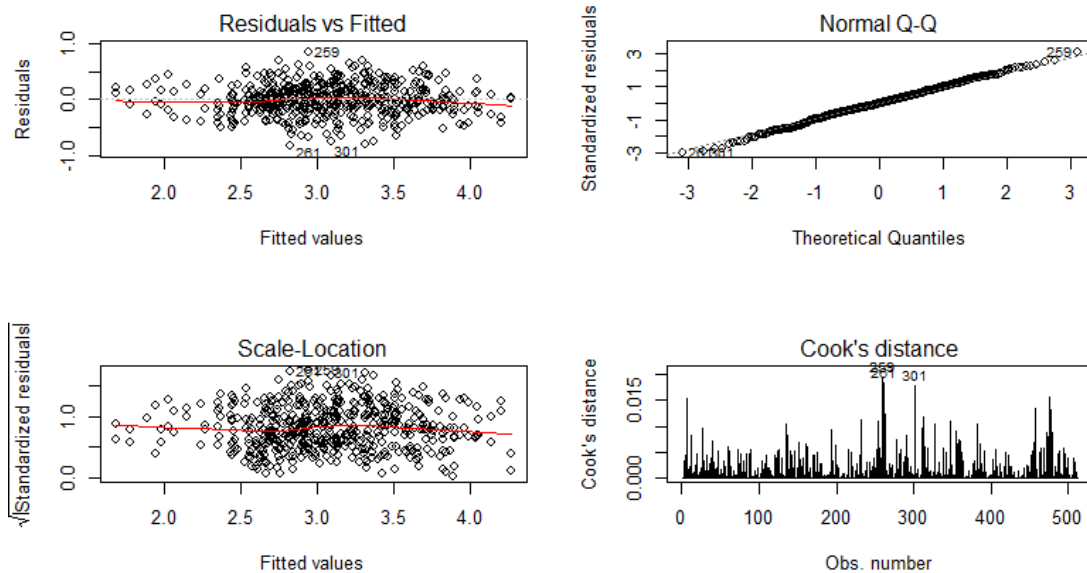


Figure 5.15: Standard plots for model checking for the linear model. Together, they indicate that assumptions of equal variance and normal distributions can reasonably be upheld.

5.4.3 Research questions 3 and 4: Statistical model and model control

Statistical model (RQ3 and RQ4)

Research questions 3 and 4 are modelled statistically within this section. I repeat them below.

3. To what extent is the acceptability of *sin* with plural antecedent dependent on extralinguistic factors such as region, age, socio-economic status of parents, gender, and place of living (urban or rural)?
4. To what extent is the acceptability of *sin* with plural antecedent dependent on otherwise non-standard use of reflexives?

To address these questions, I fitted a linear model and analyzed the data and the model in R using ANOVA. The major difference between this analysis and the analy-

sis of research question 2 is the way that the data set is structured. The data for this analysis takes as its outcome variable the primary mean rating for each participant in the study, i.e. the mean value of the participant's rating of the 32 primary sentences. This particular data set is summarized in table 5.12. The dependent variable categories that concern the background of the participants (from *Gymnasium* to *Urban or rural* in the table) are described in more detail in section 5.3.1. The dependent variables that concern the various linguistic categories are described in more detail in section 5.2.2.

For the purpose of the statistical analysis, I have simplified all of the dependent variable linguistic measures (from *Awake* to *sig bad* in the table) by converting them from mean values (which is, in itself, also a simplification) into factors with three levels: Low, Medium, High. The definition of what constitutes a "Low" or a "High" rating, e.g., is relative and thus different from category to category. A "Low" rating in all categories is a rating whose mean falls below the first quartile (the value that is the highest of the 25 % lowest ratings, Q1) of the category across all participants. A "High" rating is a rating whose mean is higher than the third quartile (the value that is the lowest of the 25 % highest ratings, Q3). A "Medium" rating is a rating whose mean falls between the first and the third quartile, i.e. within the range from 25 % (first quartile) to 75 % (third quartile) of the ratings for this particular category in the data set. This necessarily means that some information is lost in the conversion. The upside of doing this conversion, on the other hand, is a statistical model that is much easier to interpret and apply to the data at hand.

I model my data as specified in (239) which is a linear model with several categorical predictor variables. The model specified here is the final result of a process of fitting models to the data.

(239)

$$Y_{abcdefghjki} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \zeta_e + \eta_f + \theta_g + \iota_h + \kappa_j + \lambda_k + \epsilon_{abcdefghjki}$$

$$\epsilon_{abcdefghjki} \sim N(0, \sigma^2)$$

Here, $Y_{abcdefghjki}$ is the mean primary rating for participant i from gymnasium a with parents' SES b , filler rating c , local *sin* rating d , local *hans* rating e , *no binder* rating f , non-local *sin* rating g , *DP-spec binder* rating h , *sig good* rating j and *sig bad* rating k . μ is the overall mean.

Table 5.12: Data set used to investigate research questions 3 and 4

Variable	Data	Type	No. of data points
Outcome	Mean of primary ratings, 1 rating per participant	Continuous (mean)	550
Dependent	Gymnasium	Factor (16 levels)	Brønderslev, Esbjerg, Herning, Kolding, Morsø, Odense, Randers, Ringkøbing, Roskilde, Rysensteen, Rødovre, Slagelse, Sønderborg (HTX), Aabenraa (HTX), Aabenraa (STX), Aarhus
Dependent	Region	Factor (7 levels)	1 (Western Jutland), 2 (North and Eastern Jutland), 3 (Southern Jutland), 4 (Funen), 5 (Zealand), 6 (Copenhagen), 8 (Other)
Dependent	Parents' SES	Factor (7 levels)	Folkeskole, Gymnasium, Vocational, Academy, Bachelor's, Master's, Other/NA
Dependent	Questionnaire	Factor (4 levels)	1, 2, 3, 4
Dependent	Age	Factor (8 levels)	15, 16, 17, 18, 19, 20, 21-25, Other/NA
Dependent	Gender	Factor (3 levels)	Boy, Girl, Other/NA
Dependent	Danish L1	Factor (3 levels)	Yes, Yes, among others, No
Dependent	Urban or rural	Factor (7 levels)	Countryside, Village, Small town, Medium town, Large town, City, Other
Dependent	A: Awake	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	T: Training	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	F: Filler	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	S: Local <i>sin</i>	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	S: Local <i>hans</i>	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	S: No binder	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	S: Non-local <i>sin</i>	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	S: DP-spec binder	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	S: <i>sig</i> good	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)
Dependent	S: <i>sig</i> bad	Factor (3 levels)	Low (< Q1), Medium (Q1-Q3), High (Q3 <)

α_a is the effect of gymnasium.

β_b is the effect of parents' SES.

γ_c is the effect of filler rating.

δ_d is the effect of local *sin* rating.

ζ_e is the effect of local *hans* rating.

η_f is the effect of no binder rating.

θ_g is the effect of non-local *sin* rating.

ι_h is the effect of DP-spec binder rating.

κ_j is the effect of *sig* good rating.

λ_k is the effect of *sig* bad rating.

I specify the assumption that the residuals, $\epsilon_{abcdefghjki}$, are independent and identically distributed random variables, which I confirm in the Model control section below with a model check.

Model control (RQ3 and RQ4)

The assumptions behind the statistical model used to investigate research questions 3 and 4 and specified in (239) are checked using the same standard plots as in the model check on page 266. The assumptions to be checked are the same: The residuals are assumed to be identically (normal) distributed, independent random variables with equal variance. This means checking that the residuals can be assumed to be normally distributed; that the observations can be assumed to be independent; and that variance homogeneity can be assumed. The observations are the same as in the previous model check, and can be assumed to be independent based on the design of the data collection. The standard plots in fig. 5.16 check for normally distributed residuals and variance homogeneity. The qqplot in the upper right corner shows the observations clustering nicely around the diagonal, confirming that the assumption of normal residuals can be maintained. The two scatter plots (Residuals vs Fitted in the upper left corner and the Scale-Location in the lower left corner) show residuals clustering randomly around the central line and show no trumpet shapes in the Residuals vs Fitted plot, confirming that the assumption of variance homogeneity can be maintained. There are no obvious outliers to be seen in any of the plots. In conclusion, the model check showed that the assumptions behind the statistical model used to investigate research questions 3 and 4 can be maintained. (I elaborate more on the plots in the two previous model control sections on page 261 and 265).

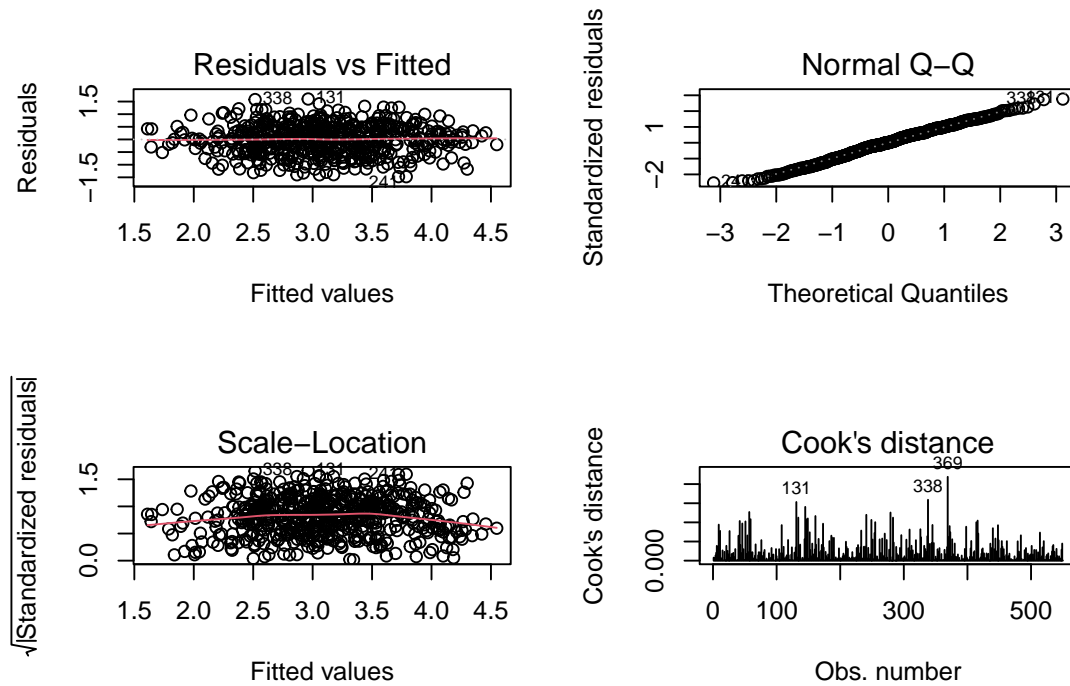


Figure 5.16: Standard plots for model checking for the linear model based on mean ratings for each individual participant. Together, they indicate that assumptions of equal variance and normal distributions can reasonably be upheld.

5.5 Results and discussion

5.5.1 Research question 1: Paired t-test

The first of the four research questions in this report is repeated here below.

1. Is there a difference in the rating given to the first and second instance of the same grammatical structure, shown with different lexical content?

Table 5.13: Results, paired t-test for sequence that addresses RQ1

Variable	Estimate	p	t	df	95% CI	
Δ mean rating	0.0002311975	0.993	0.00872	255	[-0.05198188,0.05244428]	***

The result of the statistical test of the hypothesis that goes with research question 1 is a t-value of 0.00872 which in a t-distribution with 255 degrees of freedom corresponds to a p-value of 0.993. Compared to the standard (significance) cut-off point

of $p = 0.05$, $p = 0.992$ means that I cannot reject the hypothesis that the mean in one group is the same as the mean in the other.

In more practical terms, the result of this test is that it is reasonable to assume that there is *no difference* between the rating given to the first and second instance of the same grammatical structure in the questionnaire. The plot in fig. 5.17 visualizes the means of each group (i.e. each set of 16 different sentence structures, two sets in all) for each gymnasium, and the fact that the two groups (solid and dotted lines, one for each group) clearly pattern very alike across all 16 gymnasiums supports the result that there is no significant difference in means between the two groups.

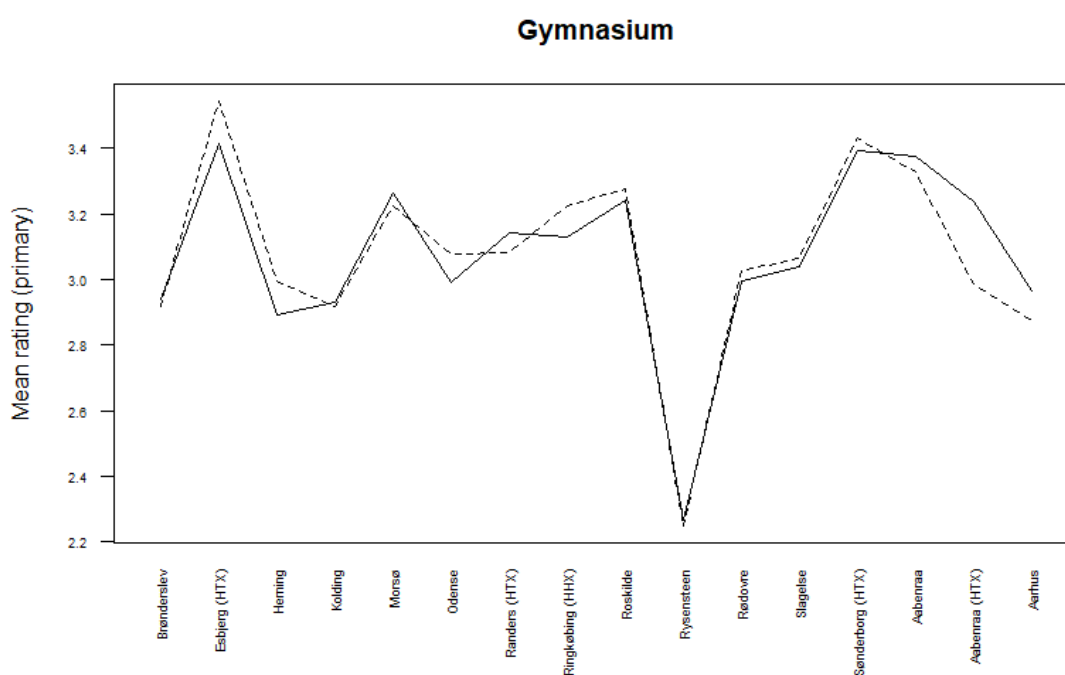


Figure 5.17: Interaction plot of the mean ratings for the two groups as they pattern across all 16 gymnasiums. The dotted line represents the first group, the ratings of the first instance of each primary structure in the questionnaire. The solid line represents the second group, the ratings of the second instance of each primary structure in the questionnaire. The plot supports the statistical test in the conclusion that it is reasonable to assume that there is no difference in mean between the two groups. The two groups follow each other closely across most gymnasiums with the largest difference (of approximately 0.2 points) seen in the ratings from Aabenraa (HTX).

The significance of this question is that it supports the hypothesis that it is the sentence structure overall, rather than the specific choices of lexical items, that determines how participants rate the sentences.

5.5.2 Research question 2: ANOVA

I repeat research question 2 (RQ2) below. I model the data as a linear model and test the hypothesis of differences in means between the groups with an ANOVA test. I interpret the results from the test below in order to address the research question.

2. To what extent is the acceptability of *sin* with plural antecedent dependent on morphosyntactic factors such as the number of the object, the type of subject, and the sentence structure?

Table 5.14: Output of ANOVA on linear model that addresses RQ2 (main effects and interactions)

Variable	df	F	p	
Gymnasium	15	32.00	$< 2.2 \times 10^{-16}$	***
Number of object	1	70.11	5.941×10^{-16}	***
Complexity of sentence	1	36.65	2.815×10^{-16}	***
Type of subject	1	567.13	$< 2.2 \times 10^{-16}$	***
Animacy	1	163.13	$< 2.2 \times 10^{-16}$	***
Type×Complexity	1	95.28	$< 2.2 \times 10^{-16}$	***
Type×Animacy	1	6.36	0.01198	*

Table 5.14 shows the output of the ANOVA itself, summarizing that all the terms included in the model have a significant effect in terms of describing the data. It should be noted here that the output shown is the result of a standard `ANOVA()` test on a linear model in R, which is a sequential test (Baayen 2008, 167), rather than a marginal test. A sequential test tests whether the particular variable contributes meaningfully to the statistical model, given the one or more explanatory variables that the model already contains (in practical terms: The variables in the table above the variable in question). This also means that the ANOVA output might give other test values (F and p-values) if the sequence of the explanatory variables were changed, and that the particular values should not be given too much interpretive weight. The relevant take-away is that all the variables in the model contribute meaningfully.

The output of the table in table 5.15 is from a `summary()` call on the same linear model. This function is a marginal – rather than sequential – test which tests the individual levels of all explanatory variables in the model with the assumption that every other explanatory variable in the model is in the model and is defined at a particular baseline value. Specifically, the `summary()` tests whether a level of an explanatory

Table 5.15: Estimates in final linear model that addresses RQ2

Variable	Estimate	SE	t	95% CI	p	
Baseline ^a	2.56	0.06	43.995	[2.45;2.68]	$< 2 \times 10^{-16}$	***
Number of object: singular	-0.21	0.02	-8.373	[-0.26;-0.16]	5.94×10^{-16}	***
Complexity of sentence: complex ^b	0.09	0.04	2.622	[0.02;0.16]	0.009	**
Type of subject: full noun ^c	0.77	0.04	17.929	[0.69;0.86]	$< 2 \times 10^{-16}$	***
Animacy: inanimate	0.25	0.04	7.248	[0.19;0.32]	1.66×10^{-12}	***
Interaction: full noun \times complex	-0.49	0.05	-9.761	[-0.58;-0.39]	$< 2 \times 10^{-16}$	***
Interaction: full noun \times inanimate	0.13	0.05	2.522	[0.03;0.22]	0.012	*
Gymnasium: Esbjerg	0.58	0.07	8.595	[0.41;0.69]	4.08×10^{-14}	***
Gymnasium: Herning	0.01	0.07	0.178	[-0.13;0.15]	0.86	
Gymnasium: Kolding	-0.005	0.07	-0.071	[-0.14;0.13]	0.94	
Gymnasium: Morsø	0.31	0.07	4.459	[0.18;0.45]	1.02×10^{-5}	***
Gymnasium: Odense	0.10	0.07	1.470	[-0.03;0.24]	0.14	
Gymnasium: Randers	0.18	0.07	2.578	[0.04;0.32]	0.010	**
Gymnasium: Ringkøbing	0.25	0.07	3.493	[0.11;0.38]	0.0005	***
Gymnasium: Roskilde	0.33	0.07	4.668	[0.19;0.47]	3.93×10^{-6}	***
Gymnasium: Rysensteen (CPH)	-0.67	0.07	-9.56	[-0.81;-0.53]	$< 2 \times 10^{-16}$	***
Gymnasium: Rødovre	0.08	0.07	1.150	[-0.06;0.22]	0.25	
Gymnasium: Slagelse	0.12	0.07	1.739	[-0.02;0.26]	0.08	
Gymnasium: Sønderborg	0.48	0.07	6.854	[0.34;0.62]	2.16×10^{-11}	***
Gymnasium: Aabenraa (STX)	0.42	0.07	6.012	[0.28;0.56]	3.59×10^{-9}	***
Gymnasium: Aabenraa (HTX)	0.18	0.07	2.571	[0.04;0.32]	0.010	**
Gymnasium: Aarhus	-0.009	0.07	-0.130	[-0.15;0.13]	0.897	

^a The baseline (read as (Intercept) in R) is the mean rating of structure i with values

gymnasium = Brønderslev,
number of object = plural,
complexity of sentence = simple,
type of subject = pronoun,
animacy = animate

and level of interaction corresponding to *simple* \times *pronoun*. The interpretation of these interaction terms will be elaborated further in the Discussion section.

^b Corresponds to interaction term *complex* \times *pronoun*.

^c Corresponds to interaction term *simple* \times *full noun*.

variable can be assumed to be 0, compared to the values of the baseline. I will walk through and interpret the results of the *main effects* in the model from table 5.15 from the top but I initially postpone the discussion of the *interaction terms*. The first row in the table is what I have called the *Baseline*. In R, this is printed as the (Intercept) in the output. The baseline corresponds to the estimated value of the outcome variable (here: Mean primary rating) given a model that by default takes the value of the first (numerically or alphabetically or coded) level of each of the factors of the model. In this case, it means that the baseline is the outcome of the model, given explanatory variables with values *Brønderslev* (alphabetically the first of the 16 gymnasiums), *plural object* (coded as the first level of the *number of object* factor), *simple* sentence type, *pronoun* type subject and *animate* subject.

The model includes two interactions: One interaction between *type of subject* and *complexity of sentence* and one interaction between *type of subject* and *animacy*. The value of an interaction in the baseline will correspond to the given values of the interaction terms as they are in the baseline, i.e. the interactions *simple* × *pronoun* and *simple* × *animate*.

It is typically not very illuminating in itself to look at the p-value for the baseline, since this p-value only tells us that the baseline outcome is significantly different from 0. However, the estimated value for the baseline tells us that this particular combination of factors gives a mean primary rating of 2.56, i.e. in the lower medium range on a scale from 1 to 5.

Research question 2 concerns the impact of (linguistic) structure on the rating of *sin* with plural antecedents. The means of the mean rating for each gymnasium and the 16 different structures are plotted in fig. 5.18. There, it is evident that some structures stand out as particularly good or bad in terms of high or low ratings, similarly across all gymnasiums, and it is reasonable to suggest on the basis of the plot that structure does make a difference. With the ANOVA, I investigate which levels of the four structural factors that contribute in which direction.

Returning to table 5.15, the only structural main effect that is not part of an interaction is *number of object*. The examples in table 5.16 illustrate the relevant difference between the plural object *sine børn* and the singular object *sit barn* which are the two levels that vary for this factor. The factor is described in more detail in section 5.2.2 on page 236.

The baseline value for the *number of object* factor is plural in the statistical model,

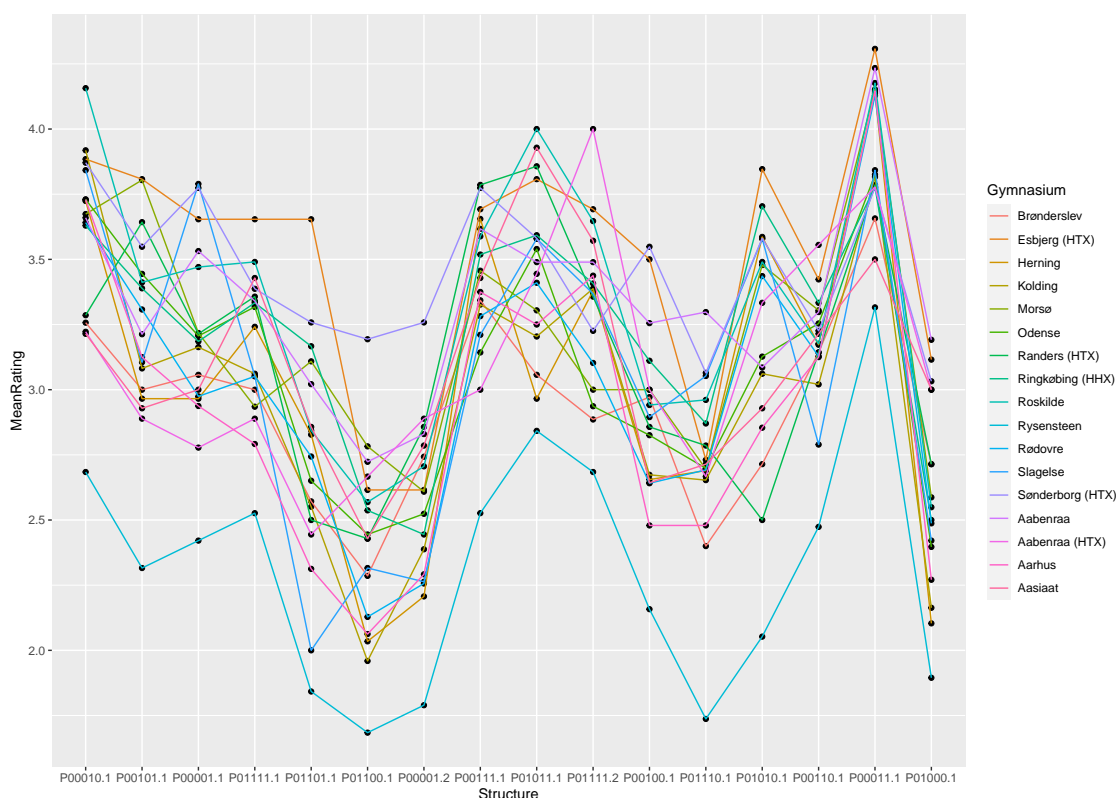


Figure 5.18: Plot of the mean of the mean ratings for the 16 unique different structures, grouped according to gymnasium. It appears that there is a difference in rating due to structure that seems to be more or less the same no matter the gymnasium.

Table 5.16: Structural main effect: Number of object

Number of object	
plural	Forældrene henter sine børn . <i>The parents collect REFL's children.</i>
singular	Forældrene henter sit barn . <i>The parents collect REFL's child.</i>

e.g. *sine børn*, and row 2 in table 5.15 gives the estimate for having a singular object, e.g. *sit barn*, rather than a plural one. This effect is statistically significant with a p-value that is much lower than the standard cutoff point of 0.05. The effect is negative with an estimate of -0.21, meaning that a sentence with a singular reflexive complement will be rated 0.21 points lower on average than the same sentence with a plural reflexive complement. The direction of this effect shows the same tendency as the results from KorpusDK discussed in section 4.8.2 on page 198. That is, that *sin* is more acceptable with a plural antecedent if *sin* itself is contained within a plural nominal.

One explanation for this difference is that it may be easier to interpret *sin* as having a distributive reading (i.e. *hver sin*) if it has a plural object, and that distributive readings improve acceptability of *sin* with plural subjects in general (as suggested by e.g. Diderichsen (1939, 68 (fn. 1)), A. Hansen (1965, 115), Lundquist (2014, 530), and Vikner and Ehlers (2017, 188)). This is not obviously the only explanation, and the corpus data does not actually show a particularly high share of distributive examples of plural antecedent *sin*. The boxplots in fig. 5.19 visualize the difference in mean ratings between singular and plural.

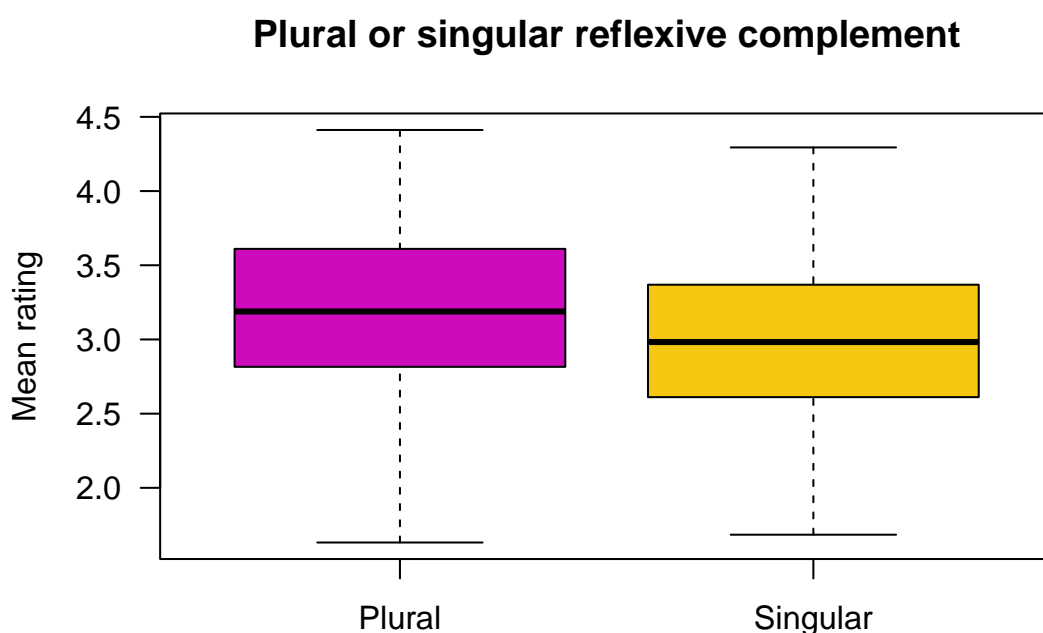


Figure 5.19: Boxplot comparing the rating of sentences with plural objects to the rating of sentences with singular objects. The ratings for plural are higher than those for singular.

Rows 3-5 in table 5.15 are all main effects that are also part of one or more interactions. The examples in table 5.17 to table 5.19 illustrate the basic differences, i.e. how the sentences in the questionnaire vary in terms of only one structural factor. The factors are described in more detail in section 5.2.2 on page 236.

Table 5.17: Structural main effect: Complexity of sentence

Complexity of sentence	
simplex	Forældrene henter sit barn. The parents collect REFL's child.
complex	Forældrene husker at hente sit barn. The parents remember to collect REFL's child.

Table 5.18: Structural main effect: Type of subject

Type of subject	
bare pronoun	De henter sit barn. They collect REFL's child.
full noun	Forældrene henter sit barn. The parents collect REFL's child.

Table 5.19: Structural main effect: Animacy of subject

Animacy	
animate subject	Forældrene henter sit barn. The parents collect REFL's child.
inanimate subject	Træerne mister sin farve. The trees lose REFL's colour.

The examples in table 5.20 show the levels that are the interactions, i.e. the four different sentence types that can be formed by combining two structural categories.

Table 5.21 spells out the interactions in the model that investigates RQ2. The interactions are indirectly available with estimates in table 5.15 but they require a few extra computations to read off directly.

The fact that it makes sense a priori to have these interactions in the model is suggested to some extent by the interaction plots in fig. 5.20 and fig. 5.21. The plot in 5.20 illustrates the interaction between subject type and sentence complexity and is a clear example of two factors interacting. If the two factors did not interact, the two lines would be (more or less) parallel, as the change in rating from one factor level to the other (e.g. from full noun to pronoun subject) would be the same irrespective of the level of the other factor (e.g. simple or complex sentence). In this case, where we do see an interaction, the two lines are not parallel, which means that the amount that the rating changes from one factor level to the other (e.g. from full noun to pronoun

Table 5.20: Interactions in the model

Interaction: Type × Complexity	
bare pronoun × simplex	<i>De henter sit barn.</i>
bare pronoun × complex	<i>De husker at hente sit barn.</i>
full noun × simplex	<i>Forældrene henter sit barn.</i>
full noun × complex	<i>Forældrene husker at hente sit barn.</i>
Interaction: Type × Animacy	
bare pronoun × animate subject	<i>De henter sit barn.</i>
bare pronoun × inanimate subject	<i>De mister sin farve.</i>
full noun × animate subject	<i>Forældrene henter sit barn.</i>
full noun × inanimate subject	<i>Træerne mister sin farve.</i>

Table 5.21: Spelled out and overtly computed estimates for all levels of the interactions from the linear model that addresses RQ2

Interaction	Interacting levels	Estimate
Type×Complexity	Bare pronoun×Simplex	2.56
	Bare pronoun×Complex	$2.56 + 0.09 = 2.65$
	Full noun × Simplex	$2.56 + 0.77 = 3.33$
	Full noun × Complex	$2.56 + 0.09 + 0.77 - 0.49 = 2.93$
Type×Animacy	Bare pronoun × Animate subject	2.56
	Bare pronoun × Inanimate subject	$2.56 + 0.25 = 2.81$
	Full noun × Animate subject	$2.56 + 0.77 = 3.33$
	Full noun × Inanimate subject	$2.56 + 0.25 + 0.77 + 0.13 = 3.71$

subject) depends on the factor level of the other interaction factor (simple or complex sentence).

Based on the graph, the rating is generally low if the sentence has a pronoun subject (1, the blue dotted line), fairly irrespective of whether the sentence is simplex (1) or complex (2). The estimates from the statistical model (as summarized in table 5.21) give the mean estimate for a simple sentence with a pronoun subject as 2.56 and a mean estimate for a complex sentence with a pronoun subject as 2.65, 0.09 points higher. The rating is overall higher for sentences that have a full noun subject (2, the red solid line) but how much higher depends on the sentence complexity. For sentences that are simple (the leftmost datapoint) and have a full noun subject,

the rating is higher than for sentences that are complex (the rightmost datapoint) and have a full noun subject. The mean estimate for a simple sentence with a full noun subject is 3.33, and the mean estimate for a complex sentence with a full noun subject is 2.93, 0.49 point lower. The difference in rating, then, is much greater between *simple* sentences with pronoun vs. full noun subjects (2.56 vs. 3.33, a difference of 0.77 points) than the difference in rating between *complex* sentences with pronoun vs. full noun subjects (2.65 vs. 2.93, a difference of 0.28). This is an example of an interaction where sentences with full noun subjects lose rating in complex sentences compared to simple sentences, and sentences with pronoun subjects conversely seem to even gain a slight amount of rating in complex sentences. This last difference is still statistically significant but numerically very small (and it did turn up as non-significant in a previous run of a similar model on a smaller part of the data set, which likely indicates that the effect is small enough to only be visible as a result of the large amount of respondents).

Figure 5.21 illustrates the interaction between subject type and animacy. Compared to the interaction between subject type and sentence complexity, the effect of this interaction is rather subtle and likely only shows up as significant in the statistical test because the sample size of the experiment is large enough to detect even small differences. Once again, the sentences with pronoun subjects (the blue, dotted line) have lower ratings overall than sentences with full noun subjects (the red, solid line). For both pronoun and full noun subjects, sentences with animate subjects (the two leftmost datapoints on the plot) receive lower ratings than sentences with inanimate subjects (the two rightmost datapoints on the plot). The mean estimate for a sentence with an animate pronoun subject is 2.56, compared to a mean rating of 3.33 for a sentence with an animate full noun subject. A sentence with an inanimate pronoun subject receives an estimated mean rating of 2.81 and one with an inanimate full noun subject receives an estimated mean rating of 3.71. The two factors interact, slightly, in that the ratings of sentences with full noun subjects benefit slightly more from having inanimate subjects (a mean rating of 3.71 for full noun inanimate subjects against 3.33 for full noun animate subjects, a difference of 0.38) than the ratings of sentences with pronoun subjects do (a mean rating of 2.81 for pronoun inanimate subjects against a mean rating of 2.56 for pronoun animate subjects, a difference of 0.25).

Having a plural inanimate subject rather than a plural animate one significantly

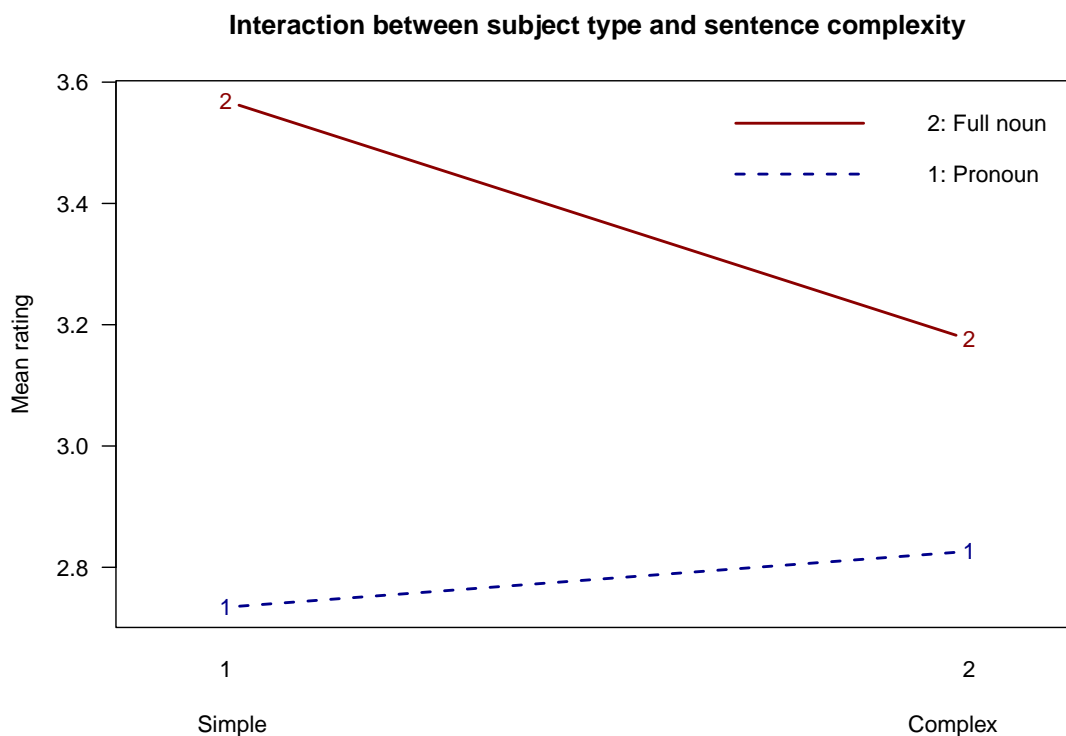


Figure 5.20: Plot of the interaction between subject type and sentence complexity, two of the explanatory variables in the linear model. The y-axis indicates the mean rating. The x-axis has values 1 (simple) and 2 (complex), the two levels of the sentence complexity factor. The blue, dotted line is pronoun type subject and the red, solid line is full noun type subject. The full noun type subject has a higher mean rating overall, but the difference between full noun and pronoun is larger if the sentence is simple.

improves the acceptability of *sin* in the sentence. It does so slightly more when the subject is a full noun compared to when the subject is a pronoun. We see a parallel effect in the corpus data where the inanimate antecedents are greatly overrepresented in the data set with plural antecedent *sin*. One likely influence on the effect within the context of the acceptability judgment study is that the sentences in the questionnaire are presented without context. This means that some of the experimental sentences with inanimate pronoun subjects are fairly outlandish, which no doubt impacts the rating given to them. An actual example is the experimental sentence *De begynder at åbne sine døre* (En. *They start opening their doors*). It is arguably much harder to think up a reasonable context for the sentence with the pronoun subject than for the corresponding sentence with a full noun (*Elevatorerne begynder at åbne sine døre*, En. *The lifts start opening their doors*). The full noun subject likely helps in contextualizing

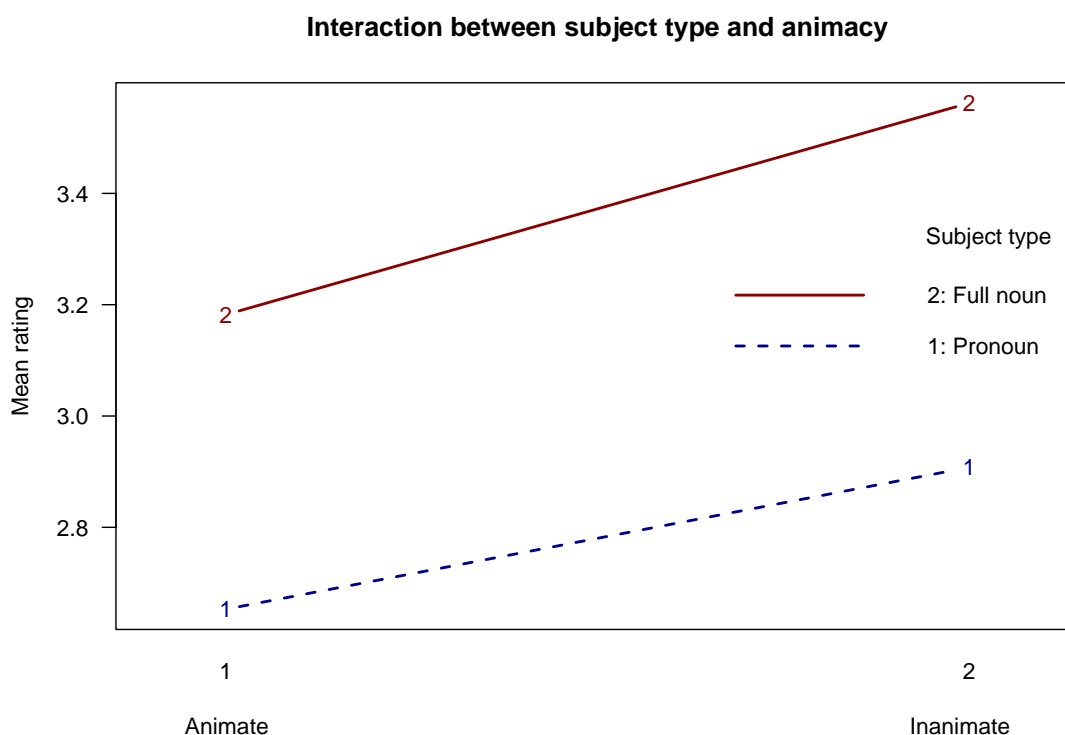


Figure 5.21: Plot of the interaction between subject type and animacy, two of the explanatory variables in the linear model. The y-axis indicates the mean rating. The x-axis has values 1 (animate) and 2 (inanimate), the two levels of the animacy factor. The blue, dotted line is pronoun type subject and the red, solid line is full noun type subject. The full noun type subject has a higher mean rating overall, but the difference between full noun and pronoun is slightly larger if the subject is inanimate.

these outlandish inanimate sentences, which also could go a long way in explaining the interaction effect between animacy and subject type.

The final 15 lines in table 5.15 concern differences between the 16 gymnasiums in the model. These differences are addressed further in section 5.5.3 which discusses the extralinguistic factors that have an impact on ratings of *sin* with plural antecedent in the data set. The way that the data is aggregated in order to build the model for research question 2 has the interesting effect that quite a number of gymnasiums show up as significantly different from the baseline gymnasium, Brønderslev. The statistical analysis that addresses research questions 3 and 4 utilises a data set that is aggregated in a different way. We will see in the sections below that this results in fewer statistically significant differences between gymnasiums. This most likely means that the differences between individual speakers are greater than the differences between

gymnasiums, i.e. that differences in rating largely have more to do with individual grammars than with regional differences.

5.5.3 Research questions 3 and 4

I address research questions 3 and 4 (RQ3 and RQ4) in this subchapter. They are repeated below.

3. To what extent is the acceptability of *sin* with plural antecedent dependent on extralinguistic factors such as region, age, socio-economic status of parents, gender, and place of living (urban or rural)?
4. To what extent is the acceptability of *sin* with plural antecedent dependent on otherwise non-standard use of reflexives?

Compared to research question 2 (RQ2), RQ3 and RQ4 are explored with a different statistical model and a different simplification of the data set. The data set is pooled according to gymnasium and sentence structure for RQ2. The data set is split into the individual students' responses for RQ3 and RQ4. These points are explicated further in section 5.4. A thing to keep in mind throughout this section is that even though the predictor variable *gymnasium* is present in both the statistical model of RQ2 and the statistical model of RQ3 and RQ4, the impact and analysis of this variable will be very different because the data set is divided in very different ways for the two models.

The data set contains a number of extralinguistic variables: *Gymnasium*, *Region*, *Parents' SES*, *Age*, *Gender*, *Danish L1*, and *Urban or rural*. For a more detailed summary of these, see table 5.12 and the full description in section 5.3.1. Only two of these extralinguistic variables turned out to be significant predictors of variation in the final statistical model summarized in table 5.22: Parents' socio-economic status (SES, accessed indirectly through a question about the parents' educational level), and the gymnasium that the participant attends. This means that the variables *region*, *age*, *gender*, and *place of living* did not turn up as statistically significant predictors. There was also no statistically significant effect of the factor *Questionnaire*, meaning that the participants' ratings of the sentences did not differ meaningfully depending on which of the four questionnaires they answered.

The data set additionally contains a number of what I call core-linguistic predictor variables. These are the variables marked with A (Awake), T (Training), F (Filler),

and S (Secondary) in table 5.12. The linguistic variables are described in more detail in section 5.2.2. The variables *Awake* and *Training* did not turn out to be statistically significant predictors of primary rating and are consequently not included in the statistical model.

The subchapter begins with the output from the analysis in R of the statistical model. The output of an ANOVA test on the statistical model can be read in table 5.22. The estimates from the ANOVA test can be read in table 5.23. The subchapter continues into a section with post hoc testing of the predictor variables in the model. The final part of the subchapter is a discussion of the results of the analysis.

Results of the statistical test

Table 5.22: Output of ANOVA on the statistical model that addresses RQ3 and RQ4

Variable	df	Sum Sq	Mean Sq	F	p	
S: Local <i>sin</i> rating	2	28.966	14.483	40.041	$< 2.2 \times 10^{-16}$	***
S: Local <i>hans</i> rating	2	30.052	15.026	41.542	$< 2.2 \times 10^{-16}$	***
F: Filler rating	2	24.939	12.470	34.474	8.994×10^{-15}	***
S: No binder rating	2	48.412	24.206	66.922	$< 2.2 \times 10^{-16}$	***
Parents' SES	6	13.045	2.174	6.011	4.337×10^{-6}	***
S: Non-local <i>sin</i> rating	2	3.200	1.600	4.424	0.012	*
Gymnasium	15	9.263	0.618	1.707	0.046	*
S: DP-spec binder rating	2	8.171	4.086	11.295	1.584×10^{-5}	***
S: <i>sig</i> good rating	2	2.613	1.307	3.613	0.028	*
S: <i>sig</i> bad rating	2	5.511	2.756	7.619	0.0005	***
Residuals	512	185.194	0.3617			

Table 5.23: Estimates in final linear model that addresses RQ3 and RQ4

Variable	Estimate	SE	t	95% CI	p	
Baseline ^a	3.115	0.174	17.907	[2.773;3.456]	$< 2 \times 10^{-16}$	***
S: Local <i>sin</i> : low	-0.167	0.071	-2.342	[-0.307;-0.027]	0.020	*
S: Local <i>sin</i> : high	0.082	0.065	1.258	[-0.046;0.211]	0.209	
S: Local <i>hans</i> : low	-0.195	0.073	-2.684	[-0.338;-0.052]	0.008	**
S: Local <i>hans</i> : high	0.170	0.065	2.641	[0.044;0.297]	0.009	**
F: Filler: low	-0.155	0.069	-2.253	[-0.291;-0.020]	0.025	*
F: Filler: high	0.026	0.070	0.367	[-0.112;0.163]	0.714	
S: No binder: low	-0.385	0.076	-5.066	[-0.534;-0.236]	5.67×10^{-07}	***
S: No binder: high	0.395	0.070	5.610	[0.257;0.533]	3.31×10^{-08}	***
Parents' SES: 2	-0.047	0.149	-0.319	[-0.340;0.245]	0.750	
Parents' SES: 3	-0.019	0.140	-0.135	[-0.294;0.256]	0.893	
Parents' SES: 4	-0.329	0.150	-2.191	[-0.623;-0.034]	0.029	*
Parents' SES: 5	-0.133	0.138	-0.961	[-0.404;0.139]	0.337	
Parents' SES: 6	-0.274	0.144	-1.906	[-0.556;0.008]	0.057	
Parents' SES: 7	0.147	0.164	0.895	[-0.176;0.469]	0.371	
S: Non-local <i>sin</i> : low	-0.162	0.073	-2.214	[-0.305;-0.018]	0.027	*
S: Non-local <i>sin</i> : high	-0.104	0.068	-1.526	[-0.238;0.030]	0.128	
S: DP-spec binder: low	-0.250	0.072	-3.480	[-0.392;-0.109]	0.0005	***
S: DP-spec binder: high	0.164	0.082	1.993	[0.002;0.326]	0.047	*
S: <i>sig</i> good: low	0.119	0.067	1.758	[-0.014;0.251]	0.079	
S: <i>sig</i> good: high	0.180	0.073	2.461	[0.036;0.323]	0.014	*
S: <i>sig</i> bad: low	-0.036	0.075	-0.481	[-0.185;0.112]	0.631	
S: <i>sig</i> bad: high	0.251	0.067	3.741	[0.119;0.384]	0.0002	***
Gymnasium: Esbjerg (HTX)	0.331	0.167	1.976	[0.002;0.660]	0.049	*
Gymnasium: Herning	0.023	0.163	0.144	[-0.296;0.343]	0.886	
Gymnasium: Kolding	-0.003	0.145	-0.022	[-0.288;0.282]	0.983	
Gymnasium: Morsø	0.078	0.148	0.524	[-0.214;0.369]	0.600	
Gymnasium: Odense	0.068	0.144	0.473	[-0.215;0.352]	0.637	
Gymnasium: Randers (HTX)	-0.016	0.198	-0.082	[-0.406;0.373]	0.935	
Gymnasium: Ringkøbing (HHX)	0.141	0.143	0.984	[-0.140;0.422]	0.326	
Gymnasium: Roskilde	0.042	0.146	0.287	[-0.245;0.329]	0.774	
Gymnasium: Rysensteen (CPH)	-0.427	0.195	-2.196	[-0.809;-0.045]	0.029	*
Gymnasium: Rødovre	0.174	0.151	1.150	[-0.123;0.471]	0.251	
Gymnasium: Slagelse	0.236	0.182	1.295	[-0.122;0.593]	0.196	
Gymnasium: Sønderborg (HTX)	0.019	0.164	0.114	[-0.304;0.341]	0.909	
Gymnasium: Aabenraa	0.233	0.147	1.591	[-0.055;0.521]	0.112	
Gymnasium: Aabenraa (HTX)	0.202	0.234	0.860	[-0.259;0.662]	0.390	
Gymnasium: Aarhus	-0.022	0.150	-0.145	[-0.316;0.272]	0.885	

^a The baseline (read as (Intercept) in R) is the mean rating of structure *i* with values
gymnasium = Brønderslev,
All secondary (S) categories = medium,
filler rating = medium,
parents SES = 1 (folkeskole)

Post hoc testing to identify any further differences

The output from the statistical test shown in table 5.23 provides information by comparing the mean for the baseline of the test to the effect on the mean of changing the level of one or more parameters of the baseline. The baseline for the test is the gymnasium Brønderslev, medium rating on all secondary and filler ratings, and a parents' SES of 1 (*folkeskole*). The mean primary rating for the combination of these values is 3.115 (the estimate of mean rating given in the topmost row in table 5.23). If I were to change the gymnasium from Brønderslev to Esbjerg, the mean rating estimate would be 0.331 (the estimate given for the Gymnasium: Esbjerg row in table 5.23) higher than the mean rating for the baseline, $3.115+0.331=3.446$. If I were to also change the value for the secondary category local *sin* from medium to high, the mean rating for this new configuration would be $3.115+0.082+0.331=3.528$ (baseline estimate + Local *sin*: high estimate + Gymnasium: Esbjerg estimate). This means that the estimates and the various test values are specific to this particular baseline and would all look different if given a different baseline. It also means that the table provides no direct information about how the non-baseline groups compare to each other internally. The table specifies that there is a statistically significant difference in rating between Brønderslev and Esbjerg and no statistically significant difference between Brønderslev and Herning but it does not provide information as to whether there is a statistically significant difference between Esbjerg and Herning.

In order to investigate whether there are any more notable differences within the groups, I plot the two large predictor variables *Gymnasium* and *Parents' SES* in figure 5.22 and figure 5.23. This method is a simple visual test for within-group differences that sidesteps the issues associated with multiple testing, particularly for categories with many subgroups. The plots show the mean rating for each sample group within the two categories as a dot with bars extending from it. The bars correspond to $1.96 \times$ the standard error of the group which in itself corresponds to a 95 % confidence interval around the mean for each group (see e.g. Altman and Bland (2005) on the difference between standard deviations and standard errors and their uses). The standard error is defined as the standard deviation of the group divided by the number of observations within the group. Given two groups with similar standard deviations but different numbers of observations, the standard error will be lower in the group with the more observations. A visual example of this is provided by the error bars for Aabenraa (HTX) in figure 5.22, where they span almost the entire plot: This is due

to the fact that Aabenraa (HTX) has a low number of observations and a fairly high standard deviation (i.e. large differences in how the students from the class rate the sentences), making the standard error comparatively large. The fact that the error bars for Aabenraa (HTX) overlap with every other gymnasium means that I cannot assume that the Aabenraa (HTX) ratings vary from any of the other groups. This result might well have been different with a greater number of observations from Aabenraa (HTX).

Figure 5.22 plots means and error bars (based on rating of primary items) for each of the 16 gymnasiums. The great majority of the gymnasiums have overlapping error bars, meaning that we cannot assume that there is a statistically significant difference in primary rating. This is the case for Brønderslev, Herning, Kolding, Morsø, Odense, Randers, Ringkøbing, Roskilde, Rødovre, Slagelse, Sønderborg, Aabenraa (HTX), and Aarhus. Rysensteen is strikingly different from the rest of the data set and only overlaps with Aabenraa (HTX) (which, as discussed above, suffers from a small sampling size and a large standard deviation, resulting in long error bars). Esbjerg (HTX) is sufficiently different from Brønderslev in the statistical test summarized in table 5.23 to show up as significantly different with a p-value of 0.049 in the test, just below the standard cut-off point of 0.05. The same result can be judged visually in the plot here, where the top of Brønderslev's upper error bar and the bottom of Esbjerg's lower error bar seem to just brush each other. Herning and Esbjerg differ in much the same way as Brønderslev and Esbjerg, and Esbjerg and Kolding slightly more so. In addition, Esbjerg and Aarhus have no overlapping error bars and can be assumed to be statistically significantly different. Finally, Aabenraa and Aarhus have non-overlapping error bars and can be assumed to be statistically significantly different.

Table 5.24: Gymnasiums where we can assume a statistically significant difference in mean primary rating.

Gymnasium to compare	Other gymnasium	Higher or lower
Esbjerg (HTX)	Brønderslev	Lower
	Herning	Lower
	Kolding	Lower
	Aarhus	Lower
Rysensteen	All gymnasiums (except Aabenraa (HTX))	Higher
Aabenraa	Aarhus	Lower

Figure 5.23 plots means and error bars (± 1.96 standard errors of the mean) for the seven parents' SES categories. The mean in SES group 6 (students with one or

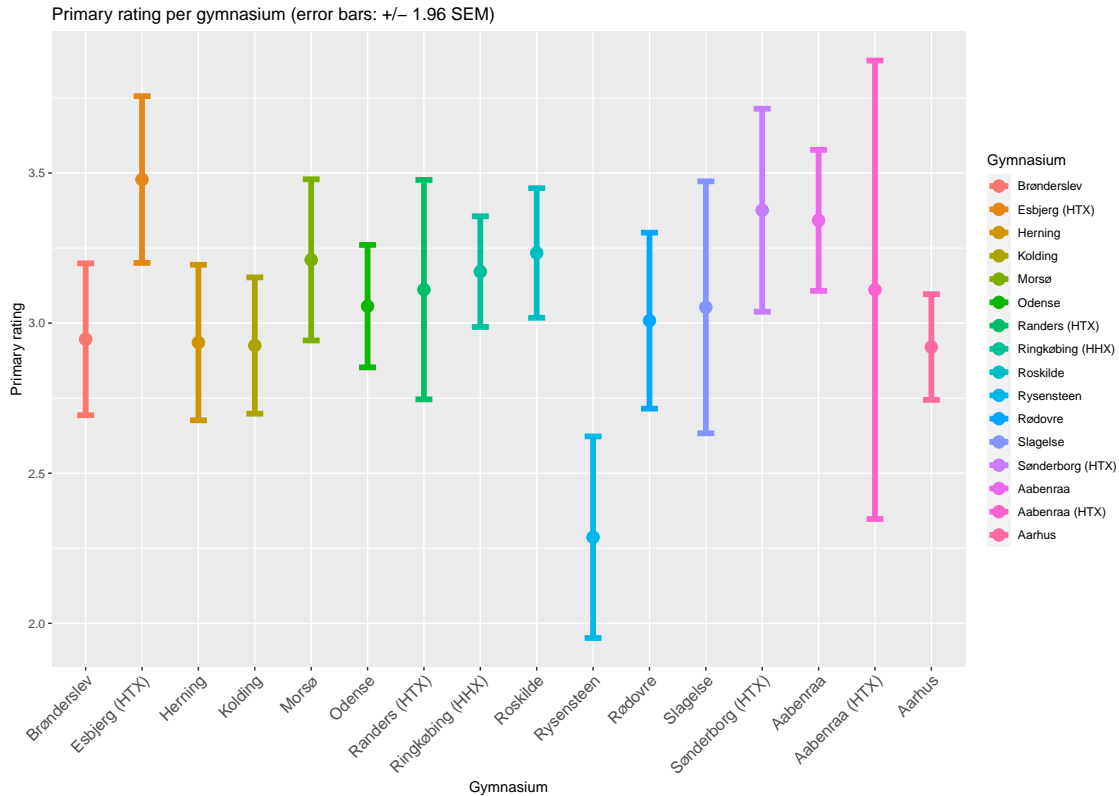


Figure 5.22: Plot of mean primary rating for each of the 16 gymnasiums with error bars marking +/- 1.96 standard errors of the mean. Allows visual judgment of differences in means across gymnasiums. Data is taken from the data set used to model research questions 3 and 4.

more parents with a Master's degree) is significantly lower than groups 2 and 3 (students with parents whose highest education is a gymnasium or vocational degree) and 7 (students who gave unclear or no description of their parents' educational level). SES group 7, in turn, is significantly higher than groups 4 to 6. Groups 1 and 7 both contain fairly few observations compared to the other groups (26 and 34, respectively) which explains the larger error bars (as the seven groups all have fairly similar standard deviations).

All levels of the other predictor variables in the model (i.e. the categories *Filler* and every *Secondary* category) have statistically significantly different means. Every group was tested with pairwise t-tests with the standard Bonferroni correction. Given the definition of the categories (Low, Medium, High defined as the observations in the lowest quartile of the group, the span between the lowest and the highest quartile, and the observations in the highest quartile in the group) this is not a surprising result. The p-values for the pairwise t-tests are reported in table 5.25.

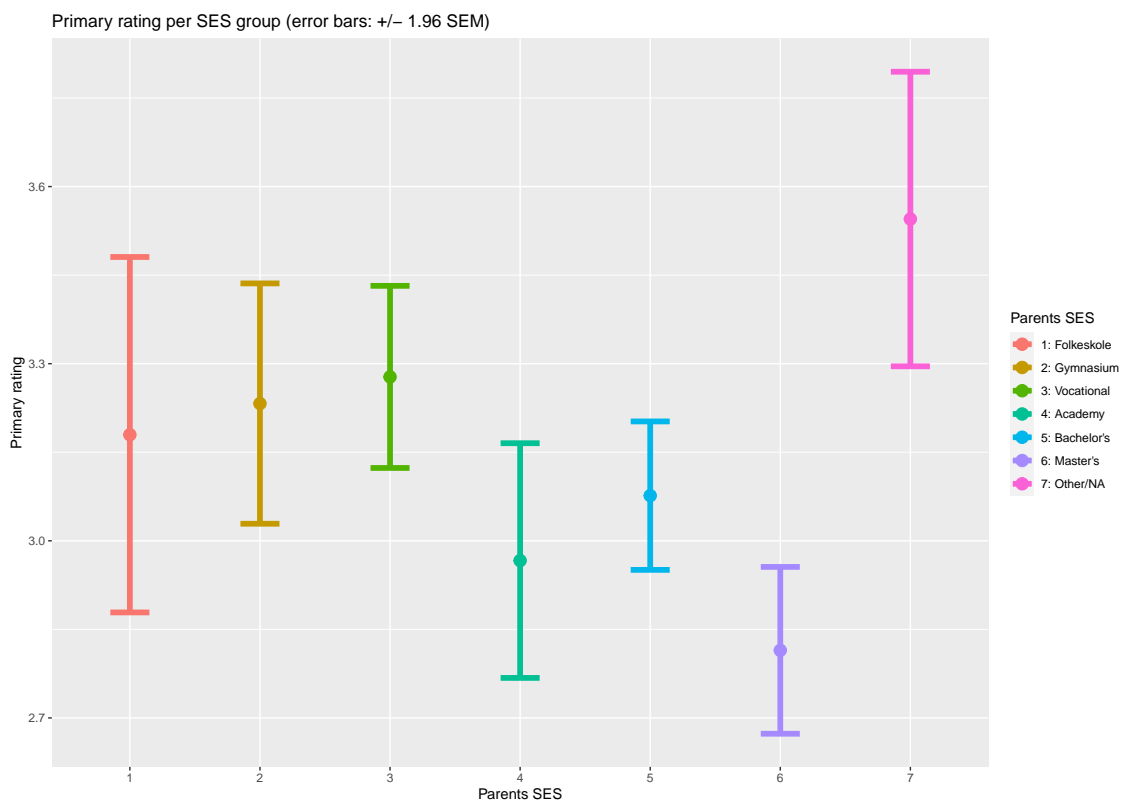


Figure 5.23: Plot of mean primary rating for each of the 7 parent SES groups with error bars marking +/- 1.96 standard errors of the mean. Allows visual judgment of differences in means across SES groups.

Discussion

The outcome variable, i.e. the factor that I actually want to investigate, is the participants' rating of sentences that contain a *sin* bound by a plural antecedent. The plot in fig. 5.24 shows every participants' mean rating of these primary sentences as a dot on the plot. The participants are grouped according to their gymnasium. The ratings spread out across the entire scale, from 1 to 5, for almost all locations. This means that there is a relatively large group across all the participants who largely accepts plural antecedent *sin*, and another relatively large group who absolutely do not find these sentences acceptable. This could point to a language variation (and perhaps change) situation where some speakers do have plural antecedent *sin* as part of their grammar (those who rate the primary sentences the highest), and some speakers do not have it as part of their grammar (those who rate the primary sentences the lowest). There is also a large middle group where some speakers perhaps accept the "good" primary sentences (see section 5.5.2) but not the "less good" primary sentences, which could

Table 5.25: Results from post-hoc pairwise t-tests for filler and secondary categories for RQ3 and RQ4 (Bonferroni correction for multiple testing). P-values: $0.05 > *$, $0.01 > **$, $0.001 > ***$.

Category		Low	High
F: Filler rating	Low	-	***
	Medium	***	***
S: Local <i>sin</i> rating	Low	-	***
	Medium	***	**
S: Local <i>hans</i> rating	Low	-	***
	Medium	***	***
S: No binder rating	Low	-	***
	Medium	***	***
S: Non-local <i>sin</i> rating	Low	-	***
	Medium	***	*
S: DP-spec binder rating	Low	-	***
	Medium	***	***
S: <i>sig</i> good rating	Low	-	***
	Medium	**	***
S: <i>sig</i> bad rating	Low	-	***
	Medium	***	***

lead to mid-scale ratings overall.

The discussion in the next two sections will address the various sociolinguistic and core-linguistic factors that are built into the data and how they may impact participants' primary ratings.

RQ3: Sociolinguistic factors

3. To what extent is the acceptability of *sin* with plural antecedent dependent on extralinguistic factors such as region, age, socio-economic status of parents, gender, and place of living (urban or rural)?

The participants in the study were asked to answer a number of demographic questions, as described in more detail in section 5.3.1. Only two of these factors turned out to be statistically significant in the analysis of the data: The participants' parents' socio-economic status and the gymnasium that the participant attends. This also means that the factors *region*, *age*, *gender*, *Danish L1*, and *urban or rural* place

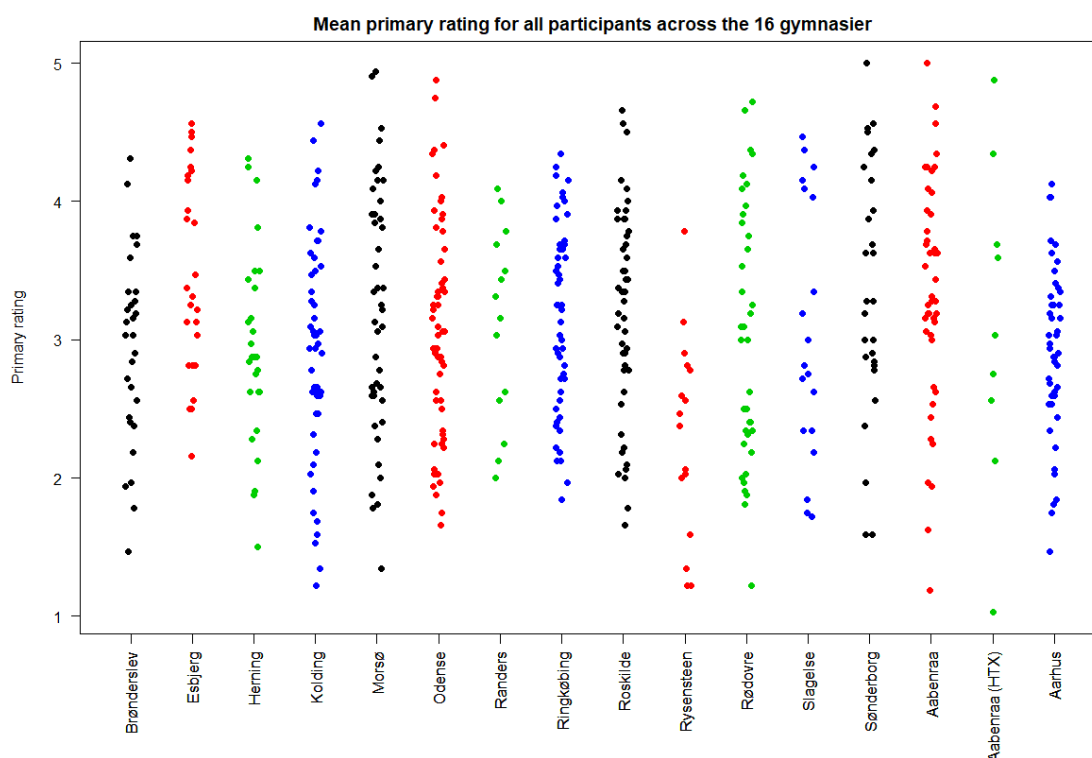


Figure 5.24: Strip chart of all participants' individual mean rating of the primary sentences. Grouped according to gymnasium.

of living did not turn up as statistically significant. There is a group of 83 participants who rate *sin* with plural antecedents highly overall (above 4 on average) and I comment on how they are distributed sociolinguistically throughout the sections.

It is very possible that the reason that the gymnasium turns up as a statistically significant factor (rather than some of the other background factors) is that it combines some of the aspects of other factors. Attendees of the same gymnasium overwhelmingly come from the same region, and a gymnasium in the middle of a larger city has more urban students than a gymnasium placed in a smaller town in the countryside. From a statistical modelling viewpoint *region* was a significant predictor in the statistical model only if *gymnasium* was not included as a predictor. Western and Southern Jutland (region 1 and region 3 on the map in fig. 5.25) are the regions where the traditional dialects have a use of *sin* that is qualitatively different from the *sin* in standard Danish. One hypothesis in the study was that speakers from Western and Southern Jutland would be more positive towards *sin* with plural antecedents than people from other regions, perhaps due to hyper-correction. There are actually more participants from Western and Southern Jutland who rate *sin* with plural antecedents above 4 on

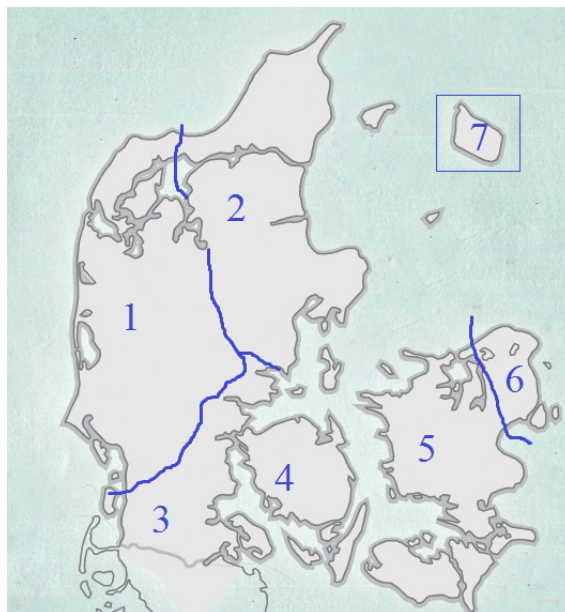


Figure 5.25: Map of Denmark with numbered linguistic regions.

average than in the sample overall: 50 (60.2 %) out of the 83 high-rating participants are from region 1 or 3 compared to 277 (50.4 %) out of 550 in the full sample.

The age span is very small in the sample population and most participants are between 16 and 19 years old. It would consequently be rather surprising if there had been a notable effect of age in the data. I had originally intended to collect data from gymnasium students and their parents (or another cohort with a similar age gap). It would be more likely to find an effect of age in a data set with a larger age gap between participants but the practicalities of data collection unfortunately overruled my original intention of doing a comparative apparent time study.

Women tend to lead linguistic change (see e.g. Labov 2001) and it is conceivable that an ongoing linguistic change could be signalled by a gender difference in the acceptability of plural antecedent *sin*. There was no statistically significant effect of gender in the data, however. This does not necessarily mean that gender has no role to play in the acceptability of *sin* with plural antecedents. It does mean, however, that the variation within the gender groups is larger than the variation between the gender groups. There are 219 (39.8 %) male participants and 331 (60.2 %) female participants in the sample overall. In the group of people who rate *sin* with plural antecedents the highest (above 4 on average) there are 40 (48.2 %) male participants and 43 (51.8 %) female participants. The proportion of male participants who rate *sin* with plural antecedents high is close to 10 percentage points larger than in the overall sample. This

does look like a gender effect but in the other direction, i.e. that the male participants actually seem to accept *sin* with plural antecedents more than the female participants. This difference is even greater when I only consider regions 1 and 3 where the acceptability of *sin* with plural antecedents is predicted to be higher. Boys from regions 1 and 3 make up 21.3 % of the full data set (117 out of 550 participants). Boys from regions 1 and 3 make up 33.7 % of the data set with the highest raters (28 out of 83 participants). These effects, to reiterate, are not large enough to show up as statistically significant in the analysis of the full sample. This is likely because the variation found in the data is more influenced by other factors, such as the differences in participants' individual grammars described in section 5.5.3 from page 295 and onwards.

The great majority of the sample are Danish L1 speakers (534 out of 550) and there was no statistically significant effect of having a different first language. This may have looked different if the group of Danish L2 speakers in the sample were larger.

There was no statistically significant effect of place of living in the sample, i.e. whether participants were brought up in (and/or live in) a more urban or more rural setting did not have an impact on their rating of *sin* with plural antecedents.

The factor *gymnasium* is a statistically significant predictor in the model of both research question 2 (section 5.5.2) and research questions 3 and 4 (this section). Because of the different ways of aggregating the data, the two statistical models diverge quite a bit in regard to which gymnasiums that show up as statistically significantly different. I focus on the model for research questions 3 and 4 for this issue as the data set for research question 2 obscures too much of the individual variation. The approximate positions of the 16 gymnasiums in the study are shown on the map of Denmark in fig. 5.26. The estimated effects of gymnasium in the statistical model of RQ3 and RQ4 are shown in table 5.23. The baseline gymnasium in the table is Brønderslev. The table reports that participants from Esbjerg (an HTX gymnasium in Western Jutland) rate the primary sentences higher than Brønderslev and that participants from Rysensteen (an STX gymnasium in Copenhagen) rate the primary sentences lower than Brønderslev. There are slightly more differences to be found with post hoc testing of the data and the model (reported on page 286 and onwards). Participants from Esbjerg (HTX) rate the primary sentences lower than participants from not only Brønderslev, but also participants from Herning, Kolding, Aarhus, and Rysensteen. Participants from Rysensteen, in turn, rate the primary sentences significantly lower than participants from all gymnasiums in the sample apart from Aabenraa (HTX). Partici-



Figure 5.26: Map of Denmark with numbered linguistic regions and the approximate location of the 16 gymnasiums.

participants from Aabenraa (STX) rate the primary sentences significantly lower than participants from Aarhus. Esbjerg (HTX) and Aabenraa (STX) are in region 1 and region 3, respectively, which are the regions predicted to have the highest acceptability scores, and this prediction is indeed (to some extent) borne out by the data.

The second sociolinguistic factor that turned up as statistically significant in the analysis overall was the socio-economic status (SES) of the participants' parents. Participants were asked to input the educational level of one or more of their parents. This is an alternative to e.g. asking participants about their parents' income which, presumably, not all teenagers would know about. It is necessarily a simplified view of SES since a person's (child's perception of) level of education does not translate directly into the various factors that make up the full picture of socio-economic status⁵. Plot 5.23 on page 289 works as a post hoc analysis of the differences between the seven SES groups in the data. Participants with parents from SES group 6 (parents with a Master's degree or higher) rate the primary sentences significantly lower

⁵According to the APA Dictionary of Psychology, socio-economic status is the "combination of social and economic factors such as income, amount and kind of education, type and prestige of occupation, place of residence, and—in some societies or parts of society—ethnic origin or religious background".

than participants with parents from SES group 2 (parents with a gymnasium degree or similar), 3 (parents with a vocational degree), and 7 (others/NA). The error bars for SES groups 5 (parents with a Bachelor's degree) and 6 overlap slightly but enough so that we cannot say that there is a statistically significant difference between them. SES group 1 (parents with a *folkeskole* education and no continuing education) is very small compared to the other groups which translates into long error bars that overlap with every other group, even if the mean value of group 1 is quite a bit higher than groups 4 to 6. Group 7 stands out from the other groups with a much higher mean value and also relatively long error bars, which can also be ascribed to the low number of participants in the group. The tendency, even if not every group is significantly different from each other, is that participants who have parents with longer educations rate the primary sentences lower than participants who have parents with shorter educations. A standard result from the sociolinguistic research tradition is that speakers from higher SES groups generally tend to use the higher-prestige standard forms of a language (e.g. Kristiansen and Pedersen 2006, 234) while a speaker from a lower SES group could tend to hyper-correct more in order to get closer to the higher prestige of the standard variety. A similar pattern seems to be the case in the data here, most obviously in terms of the participants from SES group 6 who are less happy overall with the non-standard use of *sin* than participants from the other groups.

RQ4: Core-linguistic factors

4. To what extent is the acceptability of *sin* with plural antecedent dependent on otherwise non-standard use of reflexives?

Only two of the core-linguistic dependent variables in the survey turned up as non-significant in the statistical model: the Awake and Training sentences. This means that the ratings given to the primary sentences were not found to be meaningfully impacted by how participants rated either the Awake sentences or the Training sentences. The Awake result is not surprising: Due to the exclusion criteria connected to the Awake sentences, any participant who may have diverged too much from the expected responses in his or her responses to the Awake sentences would have been excluded as an outlier prior to the modelling process. Consequently all participants within the final data set used in the model have very similar responses to the Awake sentences.

The purpose of the Training sentences is to familiarize participants with the format of the survey, and there is no reason to expect a systematic connection between this process and the way that participants rate the Primary sentences. In other words, even if the Training sentences had turned out statistically significant based on the p-values, there would be no *a priori* reason to interpret it as anything other than a random occurrence in the data set. Every other core-linguistic dependent variable is found to have a detectable impact on the rating of the primary sentences, i.e. both every Secondary category and the Filler category.

The baseline rating estimate for the model is 3.115. The baseline corresponds to a Medium rating for all Secondary and Filler categories (and gymnasium *Brønderslev* and parents' SES *folkeskole*, both of which are discussed on 282). Having Medium as the baseline category is an active choice that makes it easier to directly read off the differences between the three rating groups (Low, Medium, High) for all Secondary and Filler categories. The automatic alternative in R would be to have High as the baseline, as High is the category level that is the first one alphabetically, which would make it less easy to directly read off a potential difference between Low and Medium.

The first Secondary category in table 5.23 is the rating of Local *sin*. This is a class of grammatical sentences with singular subjects that either contain a locally bound *sin* or a locally free non-reflexive *hendes* (or *hans*). These are expected to receive high ratings overall. This expectation is supported across the data set where the sentences within this group are rated between 4.12727 and 4.71272, i.e. very close to the ceiling of 5. There is no statistically significant difference between those who rate these sentences High and Medium. The difference between groups Medium and Low is statistically significant with an estimate of -0.167 (95% CI [-0.307;-0.027]), i.e. those participants who rate these sentences the lowest also tend to give a lower rating to the primary sentences. This is a surprising result given the general prediction that a non-standard use of *sin* in one domain could lead to non-standard use of *sin* in another domain (such as with plural antecedents).

The Secondary category Local *hans* is a class of sentences that contain locally bound pronouns *hans* or *hendes*. These are predicted to show a great deal of regional variation. It should be noted that the overall rating in the data set is quite high for both sentences in the category (4.07818 for the sentence with locally bound *hans* and 4.12181 for the sentence with locally bound *hendes*). This in itself indicates that this hypothesized regional variation may well not be very pronounced, and rather that

sentences like these are quite acceptable for most members of the group of speakers included in the survey. Both the Low and the High groups are statistically significantly different from the Medium group with estimated differences in primary rating of -0.195 (95 % CI [-0.338;-0.052]) for the Low group and 0.170 (95 % CI [0.044;0.297]) for the High group. These results say nothing about regional variation but they do indicate that those respondents who rate these supposedly non-standard sentences high also rate the primary sentences higher, while those who rate these sentences low also rate the primary sentences lower.

The Secondary category No binder consists of sentences where *sin* has no proper binder within the binding domain. The sentences are rated low overall (between 1.64545 and 2.45272), as expected. Both the Low and the High groups are significantly different from the Medium baseline group. The estimated differences are the largest ones found for all the categories in the model: -0.385 for the Low group (95 % [-0.534;-0.236]) and 0.395 for the High group (95 % CL [0.257;0.533]). Those respondents who rate these (intended to be ungrammatical) sentences low also rate the primary sentences quite a bit lower, while those who rate these sentences high also rate the primary sentences quite a bit higher.

The Secondary category Non-local *sin* contains two sentences with non-locally bound pronouns, one sentence with non-locally bound *sin* and one sentence with non-locally bound *hendes*. Both of these configurations are grammatical and expected to be reasonably acceptable. The mean ratings of the two sentences are quite different from each other: The sentence with non-locally bound *sin* (Hun bad mig passe sin kat. En. gloss 'She asked me look-after REFL's cat.') receives a mean rating of 3.56545 while the sentence with non-locally bound *hendes* (Hun bad mig passe hendes kat. En. 'She asked me look-after her cat.') receives a mean rating of 4.49818. Viewed as a group, both the High and Low groups have a negative impact on the Primary rating estimate. The rating estimate for the High group is -0.104 (95 % CI [-0.283;0.030]) and the rating estimate for the Low group is -0.162 (95 % CI [-0.305;-0.018]). The Low group is statistically significantly different from the Medium baseline group and the High group is not. It should be noted that even though the estimated rating difference for the Low group is statistically significantly different from the baseline, the difference is not very large and the higher range of the 95 % CI is very close to 0. The participants who diverge from the Medium group, in other words, overall rate the Primary sentences lower. The effect is fairly small, even in the Low group which tests as statistically sig-

nificantly different.

The Secondary category DP-spec binder contains three sentences where the only possible third person singular binder for the bound pronoun is a proper name that constitutes the specifier in a possessive DP. If we assume that DPs with possessors are regular binding domains, this would predict higher ratings for sentences where a *sin* is bound by the specifier of the DP that contains *sin*. It would also predict lower ratings for sentences where a pronoun is bound locally in the same configuration. The mean ratings for the three individual sentences are quite varied and they do not conform to the predictions outlined above. The sentence *Jeg prøvede at genskabe Julies tegning af sine heste* (En. gloss 'I tried to recreate Julie's drawing of REFL's horses') received a mean rating of 3.37454. The sentence *Vi elsker Magnus' tegning af sin kat* (En. gloss 'We love Magnus' drawing of REFL's cat') receives a mean rating of 3.89454. The sentence *Jeg så Peters fotografi af hans forældre og var meget imponeret* (En. gloss 'I saw Peter's photograph of his parents and was very impressed') receives a mean rating of 4.23272. The last and highest-rated sentence could in principle have two readings: One (DP-spec bound) reading where the parents are Peter's own and one (unbound) reading where the parents are some other male person's parents. The questionnaire does not specify that the reading with coreference between *Peter* and *hans* (En. 'his') is the intended one and it is possible that some participants have rated according to the non-coreferent reading. This could bring about the effect that the (intended to be) locally bound *hans* sentence is overall rated higher than locally bound *sin*. The estimated rating change for the Low group is -0.250 (95 % CI [-0.392;-0.109]), statistically significantly different from the Medium baseline. The estimated rating change for the High group is 0.164 (95 % CI [0.002;0.326]), which is statistically significantly different from the Medium baseline but only just: The lower limit of the 95 % confidence interval is 0.002, very close to 0. Once again, those participants who rate this (perhaps) non-standard form lower, also rate the primary sentences lower, and those participants who rate this (perhaps) non-standard form higher, also rate the primary sentences higher.

The last two secondary categories, *sig* good and *sig* bad, contain sentences which cover a variety of non-possessive reflexives and non-reflexives. The 'good' category covers those sentences which are supposed to be grammatical (and presumably acceptable) according to the reflexive system outlined in section 2.3 on page 69 and onwards. The 'bad' category covers those sentences which are supposed to be ungram-

matical (and presumably unacceptable) according to the same system. Interestingly, one sentence from the 'good' group scores lower than four out of six sentences from the 'bad' group: The grammatically well-formed *Hun bad mig hjælpe sig med lektierne* (En. gloss 'She asked me help REFL with homework.PL.DEF') has a mean rating of just 2.29636, compared e.g. to the mean rating of 3.11272 of the grammatically ill-formed *Vi så ham ramme ham selv med malingen* (En. gloss 'We saw him hit himself with paint.DEF') which is the highest-rated sentence from the 'bad' group. The grammatically ill-formed and minimally different *Hun bad mig hjælpe sig selv med lektierne* (En. gloss 'She asked me help REFL self with homework.PL.DEF') receives a rating of 1.77090, approximately 0.5 points lower than its grammatical counterpart. What seems to be the case is that non-local *sig* is grammatical but somewhat unacceptable, perhaps especially in writing and when seen without much context. The results from Ehlers and Vikner (2016) could point to much the same: Non-local *sig* does appear in the texts in KorpusDK but it is fairly infrequent and mostly appears in embedded sentences with a present participle main verb. It is possible that a test sentence with *sig* placed in an embedded sentence with a present participle main verb could have elicited a larger difference in acceptability score between the presumably well-formed and ill-formed versions of the sentence.

Another interesting case from the *sig* bad group is the sentence *De skammer dem over deres gæld* (En. gloss 'They shame them over their debt', intended Danish meaning and only possible reading translates to *They are ashamed of their debt*). The mean rating for all participants in the data set is 2.77818, a score in the lower end of the middle of the scale. Behind this mean lies the observation that almost 100 participants (out of 550) give this sentence the maximum rating of 5. This is particularly interesting in light of the topic of chapter 3 which discusses the loss of reflexive plural *dem* in Danish. The relatively high number of participants who do rate this sentence as fully acceptable could point to a situation where locally bound plural reflexive *dem* is still a grammatical option for a number of speakers.

Overall, the acceptability ratings for the *sig* good category are higher than the acceptability ratings for the *sig* bad category. The mean rating for the *sig* good sentences overall is 3.83 (range: 2.29636-4.83818) and the mean rating for the *sig* bad sentences overall is 2.40 (range: 1.52727-3.11272).

Both *sig* good and *sig* bad contribute enough to explaining variation in the final linear model to be kept in the model. The estimates for the Low and High rat-

ing groups are given in table 5.23. The Low and Medium rating groups for *sig* good cannot be assumed to be significantly different from each other. The Primary rating difference estimate for the Low group is 0.119 (95 % CI [-0.014;0.251]) with a p-value of 0.079, i.e. above the cut-off point of 0.05. The High rating group can be assumed to be significantly different from the Medium group with a Primary rating difference estimate of 0.180 (95 % CI [0.036;0.323]) and a p-value of 0.014. The 25 % of participants who rate the *sig* good sentences the highest, in other words, also rate the Primary sentences slightly higher. The 25 % of participants who rate the *sig* good sentences the lowest also rate the primary sentences higher but this difference is not large enough to be statistically significant.

The participants who fall within the *sig* bad Low group cannot be assumed to be significantly different from those in the Medium group: The Primary rating difference estimate for the Low group is -0.036 (95 % CI [-0.185;0.112]) with a p-value of 0.631. Conversely, the *sig* bad High group has a rating estimate of 0.251 (95 % [0.119;0.384]) with a p-value of 0.0002. This means that the 25 % of participants who rate the ill-formed *sig* sentences the highest, on average rate the Primary sentences 0.251 higher than the rest of the group of participants. This is a point in favour of the hypothesis that non-standard reflexive use in one domain (higher acceptance of ill-formed *sig* sentences) seems to go together with non-standard reflexive use in another domain (*sin* with plural antecedents).

The final core-linguistic set of estimates in table 5.23 is the set of Filler ratings. I selected five sets of six sentences that the Danish speakers in the Nordic Syntax Database rate predominantly as 1, 2, 3, 4, and 5 respectively, which then together make up the 30 filler sentences. This selection is elaborated further in section 5.2.2 on page 242. The selection seems to have worked as intended across the data set where the overall mean rating for the filler sentences is 3.13 (Q1: 2.8, Q3: 3.4), which precisely indicates a reasonably even spread of ratings along the whole spectrum across the 30 filler sentences. A closer look at the individual ratings indicates that this is not exactly what has taken place: The rating 1 makes up 25.8 % of the filler ratings, the rating 2 makes up 13.9 %, the rating 3 makes up 13.7 % of the filler ratings, the rating 4 makes up 15.6 % of the filler ratings, and the rating 5 makes up 31.0 % of the filler ratings. There is a clear preference for the two extreme values on the acceptability spectrum. Participants do still use values from the middle of the spectrum, however, even if the usage frequency (between 13 and 15 % per rating category) is somewhat

lower than the approximately 20 % per rating category that would have been the ideal even spread. A part of an explanation for the discrepancy could be that the participant cohort in the Nordic Syntax Database is rather different demographically (and much smaller) from the participant cohort in the current study, and that the rating scale may have been explained somewhat differently to the Nordic Syntax Database participants.

The Filler rating category is a significant enough predictor of variation in Primary rating to be retained in the final statistical model. Looking at the estimates in table 5.23, there is a significant difference in Primary rating between the Low group and the Medium group: The estimate is -0.155 (95 % CI [-0.291;-0.020]) with a p-value of 0.025. There is no significant difference in Primary rating between the Medium group and the High group (the Primary mean rating estimated difference here is 0.026 (95 % CI [-0.112;0.163]) with a p-value of 0.714). These results show that the 25 % of participants who rate the Fillers the lowest also tend to rate the Primary sentences lower than the rest of the group. The reverse is not the case, i.e. there is no indication that the participants who rate the Fillers higher also rate the Primary sentences higher.

Summary The aim of the precedings sections was to answer the two research questions below.

3. To what extent is the acceptability of *sin* with plural antecedent dependent on extralinguistic factors such as region, age, socio-economic status of parents, gender, and place of living (urban or rural)?
4. To what extent is the acceptability of *sin* with plural antecedent dependent on otherwise non-standard use of reflexives?

Two extralinguistic factors were significant predictors in the model: gymnasium and parents' SES. I predicted that students from regions 1 and 3 would rate the primary sentences higher than students from the other regions, perhaps due to hypercorrection. The two gymnasiums that provide statistically significantly higher ratings of the primary sentences are from region 1 (Esbjerg) and region 3 (Aabenraa (STX)) so this prediction is to some extent borne out. However, the other gymnasiums in region 1 and 3 are not generally significantly different from the rest of the gymnasiums in the sample. A closer look at the data also showed that regions 1 and 3 contained a higher relative proportion of students who rated the primary sentences at an average of 4 or

higher than in the general sample. This is a descriptive result and not something that can be extracted from the statistical model. The effect of gymnasium as a predictor in the model is not very large overall.

The other statistically significant extralinguistic factor is parents' SES where students who have parents with longer educations generally rate the primary sentences lower. The post hoc visual test indicated that students with parents with (at least) a Master's degree rate the sentences the lowest, followed by students with parents with an Academy agree. The differences between the groups are not very large, however, and most of the SES groups cannot be said to be significantly different from each other. The extralinguistic factors, in short, do not contribute very much to explaining the variation seen in the acceptability of *sin* with plural antecedents, even if two extralinguistic factors did show up as statistically significant enough to be kept in the model.

It seems warranted to say, on the basis of this data, that people's acceptability of *sin* with plural antecedents is better explained by variation in their individual grammars than by any of the suggested extralinguistic groupings. There is a general tendency in the data that participants who rate other non-standard uses of reflexives higher also tend to rate *sin* with plural antecedents higher. This is the case for the ratings of locally bound pronouns, the ratings of unbound *sin*, the ratings of possessives with DP-spec binders, and the ratings of other types of ungrammatical reflexive sentences (the category *sig* bad). The reverse is also true for the categories with non-standard uses of *sin*: Those who give lower ratings to the sentences with non-standard uses of possessive pronouns and reflexives also rate *sin* with plural antecedents lower. The same effect is not seen for the category with ungrammatical non-possessive reflexive sentences.

The way that participants rate the fillers seems to be less correlated with how they rate the primary sentences. There is a statistically significant negative effect of giving low ratings to the fillers, which presumably means that there are participants who just generally rate the sentences lower. There is no statistically significant effect of medium or higher filler rating. The rating of locally bound *sin* shows a surprising effect in the same direction: A lower rating of locally bound *sin* with singular antecedents also predicts a lower rating of *sin* with plural antecedents. I would have expected the direction of the effect to be the reverse since a lower acceptability of locally bound *sin* must also count as non-standard. One explanation could be that

some participants have a grammar where the expression of gender trumps the expression of reflexiveness (this is described as a general tendency in Danish already by Diderichsen (1939, 91)) which could lead to lower acceptability of sentences with *sin* with gendered antecedents in general. It could be interesting to see if there is a noticeable difference between the acceptability of *sin* with gendered (male, female) and non-gendered antecedents but that is unfortunately not something that I had considered before designing the experiment.

5.5.4 Some remarks on the statistics and design of the study

I have made a number of choices with regards to the make-up of the questionnaire sentences, choice of statistical tests, and simplification of the data set. Some of these I would choose to do differently in a do-over of this acceptability judgment study.

I would not include *animacy* as a factor in the primary sentences in a do-over of this experiment. This would reduce the number of different primary sentences from 16 to 8, allowing for a smaller questionnaire and perhaps less questionnaire fatigue from participants. It would also make the primary sentences more directly comparable since the shift from animate to inanimate changes the sentence much more so than e.g. a shift from a singular *sin* nominal to a plural *sin* nominal. An alternative could be to exchange the animacy factor for a distributivity factor, since distributivity very well could be an underlying relevant factor that is not overtly built into the data. A third alternative would be to include natural gender as a factor on the antecedent.

I would like to do another experiment which specifically targets the acceptability of *sin* with antecedents with and without overt masculine or feminine gender, singular and plural. The data from the study here, as well as from the corpus study in chapter 4, indicates that animacy or natural gender does play a part in the acceptability of *sin* but the design in either study cannot tease apart properly what the effects are.

There are several other factors in the make-up of the experimental items that could be improved in a follow-up experiment. I do not, e.g., control properly for whether the final prepositional phrase that is included in many of the sentences is a modifier or a complement of the verb. I have attempted to choose relatively frequent verbs but that is most likely another factor that could be improved. Some of the verbs used as matrix verbs, in particular, could probably merit closer scrutiny.

I would perhaps choose to analyze the data with a mixed models analysis with random intercepts for participants. This should in principle remove some of the random

noise that is the individual participants' own grammar and own interpretation of the rating scale. This could in principle allow some of the systematic effects of the other predictors in the data to show up more clearly. It is also very possible that it would not make much of a difference, but I cannot know that without having performed the alternative analysis. However, with the way that I have aggregated the data (by calculating one mean per participant and having multiple explanatory factors) the choice of ANOVA (or, more correctly, ANCOVA) as the statistical test seems the right choice. One tweak that I would consider using is to norm participants' responses according to their individual use of the rating scale (following Sprouse, Wagers, and Phillips (2012), where doing the transformation, however, did not change the results compared to analyzing the raw data). A pragmatic counter-point to all of the above is that the sample discussed here is so large that the fine-grained choice of test probably does not matter all that much: Sprouse and Almeida (2017, 17) note that 100 participants is "a likely upper bound for most acceptability judgment experiments" and that good statistical power is still reached with a lower number of participants.

A choice that I would definitely change in an alternative analysis of the data is the choice to mean the data per participant. This removes a large amount of information about individual variation and variation between test items and the statistical analysis would certainly be more informative if performed on un-meanned data.

Another consideration is whether it would have made sense to present participants with a 7-point scale instead of the 5-point scale. The 7-point seems to be the standard choice in the linguistic literature. Song and Oh (2016) tested whether the choice of scale made a significant difference and found that it did not overall. They did find that the 7-point scale detected slightly more variation between participants, however, which could be a point in favour of that particular choice. There are other scale options such as e.g. magnitude estimation or a yes-no design ("Is this sentence acceptable" Y/N) which I have chosen not to use. The yes-no design is less well-suited to a design with multiple explanatory factors according to Sprouse and Almeida (2017, 26). The magnitude estimation design is overall less reliable than Likert scales, according to the results from Langsford et al. (2018).

A very specific point that I would change if I were to redo the statistics, is the structure of the t-test that addresses research question 1. I compare pairs of mean ratings for each gymnasium but I believe that the comparison would have been more valid (if not necessarily with a different result) if I had compared the individual participants'

ratings of the 16 structure pairs. Conversely, the result that I do gain from my chosen comparison is that the specific choice of lexical material overall does not make a difference in rating in the experiment.

As a last point, it is necessary to stress that the conclusions from this study are not applicable at a population level. The sample of study participants were drawn from the population of gymnasium students: A limited age group and a certain social bias in terms of who chooses to study at the gymnasium rather than at some of the other options for secondary education in Denmark. The conclusions based on this sample cannot be said to apply to e.g. middle-aged construction workers, or perhaps even to hairdressing students of the same age as the students in the sample.

5.6 Conclusion

This chapter has addressed the following research questions:

1. Is there a difference in the rating given to the first and second instance of the same grammatical structure, shown with different lexical content?
2. To what extent is the acceptability of *sin* with plural antecedent dependent on morphosyntactic factors such as the number of the object, the type of subject, and the sentence structure?
3. To what extent is the acceptability of *sin* with plural antecedent dependent on extralinguistic factors such as region, age, socio-economic status of parents, gender, and place of living (urban or rural)?
4. To what extent is the acceptability of *sin* with plural antecedent dependent on otherwise non-standard use of reflexives?

The first question is tested with a paired t-test on two groups of data. The groups are the mean ratings of the first and the second instance of the same grammatical structure grouped according to gymnasium. The test could not reject the hypothesis that there is no difference in mean rating between the two groups. We can assume that there is indeed no difference in the ratings of the first instance of a particular grammatical structure compared to the second instance of the same structure.

The second question is addressed by fitting a linear model to the data and testing it with ANOVA. The outcome data consists of mean values, one data point for each

of the 32 primary sentences for each of the 16 gymnasiums, 512 data points in all. The model check showed that the data behaves as it should, meaning that the model assumptions of normally distributed, independent residuals with equal variance are substantiated. The ANOVA test shows that the rating of *sin* with plural subject does vary with the structure of the sentence. It affects the rating positively if the sentence has a plural object, as opposed to a singular one. Sentences with inanimate subjects receive higher ratings than sentences with animate subjects. Sentences with full noun subjects are rated higher than sentences with pronoun subjects. Simple sentences are rated higher than more complex sentences. Animacy and sentence complexity interact slightly: Sentences with inanimate subjects benefit a little more from having a full noun subject than sentences with animate subjects do. Subject type and sentence complexity interact: Full noun subjects have a larger positive impact on the rating of simple sentences than on the rating of complex sentences, where the impact is still, however, positive. Pronoun subjects generally give a lower rating than full noun subjects, no matter the sentence type.

The third and fourth questions are analyzed by fitting a different linear model to the data and analyzing it with an ANOVA test. The data is also aggregated differently: For questions 3 and 4, the outcome data consists of the mean value of each participants' rating of the primary sentences, one mean value per participant which makes for 550 data points in all. The model check showed that the data conforms to the assumptions of variance homogeneity, independent observations, and normally distributed residuals. In answer to research question 3, the analysis showed that two extra-linguistic factors had a statistically significant impact on the rating of *sin* with plural antecedents: *Gymnasium* and *parents' socio-economic status*. The two factors did not explain a lot of the variance in the data but enough to be kept in the statistical model. Two gymnasiums from Western and Southern Jutland were significantly different from other gymnasiums in the sample, Esbjerg (STX) and Aabenraa (STX), in that the participants from there generally rated the primary sentences higher. One gymnasium, Rysensteen in Copenhagen, stood out in rating the primary sentences lower than almost every other gymnasium in the sample. The initial prediction was that students from Western and Southern Jutland would rate the primary sentences higher than students from the rest of the country, e.g. due to hyper-correction. The data shows a slight tendency in that direction but also a lot of students from the rest of the country who are quite happy with *sin* with plural antecedents.

Research question 4 addresses the linguistic factors that have an impact on the rating of *sin* with plural antecedents. There is indeed a general tendency in the data that participants who rate other non-standard uses of reflexives higher also rate *sin* with plural antecedents higher. There is also a tendency that participants who rate non-standard sentences with possessives (reflexive and non-reflexive) lower also rate *sin* with plural antecedents lower. This effect is not visible for sentences with non-possessive reflexives and pronominals. These results support the hypothesis that there is a connection between using *sin* with plural antecedents and a higher use or acceptance of reflexives used in a non-standard way. This does not extend to the ratings of sentences which do not contain reflexive structures. There was no statistically significant effect of a higher rating of filler sentences so the effect seems to be concentrated on reflexive use. The participants who rated the fillers the lowest, however, also rated the primary sentences significantly lower.

The results from the study are limited in scope: The sample is drawn from the population of gymnasium students and thus the conclusions cannot be said to count for the wider population. A follow-up study could productively recruit a sample of older speakers in addition to the younger ones in order to further investigate the possibility of a change in progress signalled by a difference in acceptability between generations.

As an avenue of further research I am very interested in the connection between overtly encoded gender on antecedents and the use of reflexives. It would be interesting to see if we could find the effect suggested by Diderichsen (1939, 91) where he claims that Danish speakers have a preference for expressing a gender when it is overtly present (typically *hans, hendes*) even if a reflexive is standardly required in the position. Another interesting, and apparently unexplored, topic is the phonetics of locally bound pronouns: To my ears, locally bound singular pronouns (e.g. in a sentence like "Han₁ tog hans₁ hat og gik", En. "He took his hat and left") sound markedly different from unbound pronouns in a similar sentence. This observation is purely anecdotal but, if true, it could indicate that even if quite a high number of Danish people accept and use locally bound singular pronouns there is still a reflexive quality to them that comes out through the phonetic system.

Chapter 6

Conclusion

In this thesis, I have investigated the use and acceptability of the possessive reflexive *sin* with plural antecedents in Danish. *Sin* is largely restricted to occurring with singular (or non-plural) antecedents in modern Danish. This makes the Danish *sin* stand out from the possessive reflexives in the other Scandinavian languages where the cognates of *sin* are not similarly number-sensitive. The fact that Danish *sin* mainly occurs with non-plural antecedents also challenges the universality of arguments in the linguistic literature where reflexives are argued to be maximally underspecified or at least not specified for number. This number-restriction may also be changing in Danish, however. Examples of *sin* used with plural antecedents can be heard and read regularly, both in e.g. Danish corpora, in the press, in official communication from the Danish authorities, and on social media. I cannot say for sure that *sin* is undergoing change. I can say, however, that *sin* is used as an alternative to *deres* (the standard form) in specific contexts and in a systematic way. I can say that *sig*, the non-possessive reflexive counterpart to *sin*, has undergone a similar change from being used mainly with singular antecedents to being used with both plural and singular antecedents. *Sig* most likely changed due to external pressure from German and there has been no similar external pressure to change *sin*. Conversely, with *sig* having lost its number restriction, there is less reason for speakers of Danish to acquire a *sin* that is number restricted.

In the first chapter of the thesis, I introduce the theories and technical machinery behind the approach to linguistics that I adopt in this thesis. I situate my thesis within a framework of generative grammar, and within a framework of variation and change with a special focus on microvariation. I sketch the development in the research on

reflexives from the relatively unified Government and Binding approach of the 1980s to the current, rather less unified range of approaches.

In the second chapter of the thesis, I narrow the focus and look specifically at Danish and the Danish reflexive system. The first part of the chapter is a general introduction to Danish grammar, both in terms of clause structure and nominal structure. I also describe the development of the Common Germanic cognates of *sig* and *sin* in the various Germanic languages, from Gothic to the modern languages. The Germanic languages that have prenominal definite articles have all lost the possessive reflexive *sin* or repurposed it as a non-reflexive possessive pronoun. The same thing seems to have happened in the dialects of Danish – Western and Southern Jutlandic – that use prenominal definite articles. I present an analysis of the reflexive systems in both standard Danish and these Jutlandic dialects. The ways that reflexive use in the Jutlandic dialects differ from standard Danish are highly stigmatised and I hypothesize that this could lead to speakers hyper-correcting and using *sin* with plural antecedents more frequently. Finally, I present the use of *sin* in standard Danish that I call number restricted. I present a number of cases where *sin* and *deres* are interchangeable also for speakers who do not generally find plural antecedent *sin* acceptable. I also compare the variation or optionality between *sin* and *deres* with similar variation in predicative adjective agreement.

The third chapter of the thesis is a diachronic study of the use of *sin* and *sig* in Danish over the last millennium. There is textual evidence from runic stones from Denmark that *sin* was used with both plural and singular antecedents before 1000 AD. There is no direct evidence that this was also the case for *sig*. This use is, however, a direct continuation from the Common Germanic stage where both *sig* and *sin* were used with antecedents of all numbers. By the 13th century, where the earliest Danish manuscripts are from, the use of *sin* and *sig* had changed so that both forms are primarily used with singular antecedents and their non-reflexive counterparts (*deres* and *dem*) are used with plural antecedents. This is particularly true for the Western parts of what used to be Denmark, i.e. everything that is west of Scania (in present-day Sweden). Texts in the following 6-700 years shows different developmental trajectories for *sig* and *sin*. *Sig* becomes the predominant form in the written language with both singular and plural antecedents after the Reformation (with a great deal of variation). This is probably due to influence from German, and locally bound *dem* stays in use in the spoken language until the early 20th century (Pedersen 2017). Locally bound

dem must be considered a very marginal form in the modern language outside of the contexts where there is structurally conditioned optionality between *sig* and *dem*. *Sin* stays limited to mainly singular antecedents, although examples of plural antecedent *sin* can be found sporadically all through the period.

The fourth chapter of the thesis is a corpus study. I investigate the occurrence and distribution of plural antecedent *sin* in KorpusDK, which is a written corpus with texts from 1983 to 2002. I found 1218 examples of plural antecedent *sin* in the corpus out of 188,585 instances of *sin* with any kinds of antecedents. I annotated the 1218 plural antecedent *sin* examples with a number of factors, and did the same analysis on a random sample of the standard forms, i.e. *sin* with any kind of antecedent and bound *deres*. I compare the relative distribution of plural antecedent *sin* with the relative distribution of the samples of the standard forms in order to find areas where plural antecedent *sin* is similar or dissimilar to the standard forms. Overall, plural antecedent *sin* occurs relatively more often with inanimate antecedents, with full noun antecedents (rather than pronoun antecedents), with coordinated antecedents with singular coordinands, and in complex clauses (particularly a type of clause where a partitive is modified by a relative clause). The data does not support suggestions from the literature that plural antecedent *sin* is better or more likely to occur within distributive contexts (discounting examples with overtly distributive *hver sin* where plural antecedent *sin* is completely conventionalised). The data also shows that plural antecedents tend to bind plural nominals at a higher rate than singular antecedents. However, this tendency is more pronounced if the bound possessive is *deres* rather than *sin*.

The final chapter is an acceptability judgment experiment with a sample of 550 young Danish speakers from various Danish gymnasiums (upper secondary schools). I used the results from the corpus study to create the experimental items. The experimental setting both allows me to directly control the linguistic factors and to investigate how sociolinguistic variables could impact the acceptability of plural antecedent *sin*. I find in the study that plural antecedent *sin* is more acceptable if it is contained with a plural nominal, if the antecedent is inanimate, and if the antecedent is a full noun. Simple sentences with plural antecedent *sin* are rated higher than complex sentences but that would probably be true for any topic under this type of investigation. I found a few statistically significant sociolinguistic tendencies in the data but it is perhaps more striking that the initially hypothesized regional differences were not

very pronounced. Participants from two gymnasiums in Western and Southern Jutland rated plural antecedent *sin* higher than the rest of the gymnasiums in the sample, and participants from a gymnasium in Copenhagen rated plural antecedent *sin* rather a lot lower than the rest of the gymnasiums in the study. There was also a slight effect of socioeconomic status where participants with parents with longer educations tended to rate plural antecedent *sin* lower. The fact that there were not any more visible effects of sociolinguistic groupings reflects a data set where the variation within the groups (regions, gymnasiums, socio-economic status and more) was greater than between the groups. Participants from all over the country rated plural antecedent *sin* as quite acceptable, just like participants from all over the country rated plural antecedent *sin* as quite unacceptable. I believe that this could show a linguistic state of ongoing change where some speakers have plural antecedent *sin* as part of their grammars (either through acquisition or through hyper-correction), and some speakers do not. Whether or not a speaker accepts plural antecedent *sin* seems more dependent on that speaker's own grammar than on the sociolinguistic factors that I coded the data for. Those speakers who rate other non-standard uses of reflexives (such as using locally bound singular possessive pronouns) higher also tend to rate plural antecedent *sin* higher. This could be an argument in favour of the hyper-correction hypothesis. As a last note, there are a number of decisions about the study design and choice of statistical tests that I would change in a do-over of this study, and I discuss these in the last parts of the chapter.

Some of my initial goals with this thesis have not come to fruition. I had hoped to identify other areas of variation or ongoing change in Danish that could be clearly correlated with a possible change in the use of *sin*. I have not been able to find any that unequivocally do this. I have identified some topics that could be play a part, however, and I will permit myself a bit of speculation: A (perhaps) increasing use of locally bound singular pronouns and a normative push to use *sin* instead could prompt some speakers to over-generalize or hyper-correct and use *sin* with plural antecedents, too. The results of my acceptability judgment study point in this direction, in that speakers who rate non-standard uses of reflexives higher also tend to rate *sin* with plural antecedents higher. Variation resulting from hyper-correction does not reflect a change in the I-language of the hyper-correcting speaker but it could impact the I-language of the next generation(s) of speakers acquiring Danish, pushing in the direction of language change. I had also hoped to be able to use the acceptability judgment study

as an apparent time study and compare the results for the younger speakers with results from older speakers. Juel Jensen (2009b) includes plural antecedent *sin* in his (real time and apparent time) diachronic study but there are so few examples of plural antecedent *sin* in the LANCHART data that the change effect would have had to be rather large to be visible. I hope to be able to do a follow-up study that can detect a smaller change.

Another factor that could push *sin* towards losing its number-restriction is that fact that *sig* has undergone the same process relatively recently. If we assume that the elements in the Danish reflexive system can indeed be said to be a part of the same system, the use of *sig* and *sin* with primarily singular antecedents could have supported the number-restriction for both forms. An external push from German spurred on the change in *sig* from being number-sensitive (between 1200 and until around 1900) to being number-neutral in regards to its antecedent. There has been no similar external push to change *sin* which probably largely explains why only *sig* has changed so far. However, in the other direction, there may be fewer language-internal reasons to acquire a number-sensitive reflexive possessive *sin* now that the reflexive non-reflexive *sig* is not similarly number-sensitive.

Another factor, which is purely speculative but which could certainly be investigated empirically, is whether the degree of optionality in predicative adjective agreement with the impersonal pronoun *man* could have an impact on the use of plural antecedent *sin*. *Man* can only bind *sin* and never *deres*, (240).

- (240) a. **Man**₁ kan fotografere **sin**₁ kat.
 MAN can photograph REFL's cat
 'One can photograph one's cat.'
- b. ***Man**₁ kan fotografere **deres**₁ kat.
 MAN can photograph their cat
 'One can photograph one's cat.'

Man does, however, allow a great deal of variation in the forms of its associated predicative adjectives.

- (241) a. Man er forsigtig-Ø og opfører sig ikke latterligt. (KorpusDK)
 MAN is careful-SG and act REFL not ridiculously
- b. Man er forsigtig-e og har tid til at vente på det rigtige køb.
 MAN are careful-PL and have time to to wait for the right purchase
 (KorpusDK)

This use of *man* contrasts with e.g. collective subjects which can usually bind either *sin* or *deres*, and also allow both singular and plural predicative adjectival inflection. I mention *man* as a particular topic of interest because my grammar seems to permit more wide-spread optionality between singular and plural adjectives than the grammars of speakers from older generations that I have informally quizzed on this. Given *man*'s close connection to *sin*, a change in the former could perhaps push a change in the latter. This topic certainly warrants further study, both in order to see whether the suggested agreement variation is more than my own biased intuitions about change, and in order to see whether there is a connection between agreement variation with adjectives and antecedent agreement variation with *sin*.

Summing up, *sin* can be found and heard with plural antecedents in Danish quite regularly, even if *deres* is still by far the most frequent form. Young speakers from all over the country are happy to accept sentences with plural antecedent *sin* (and another group of speakers from all over the country are quite happy to *not* accept it), particularly so if they also tend to accept other non-standard uses of reflexive pronouns. Going forwards, it would be interesting to compare the data from the gymnasium students with judgments from other age groups in order to see whether we can find evidence of a change in apparent time. It would also be interesting to tease out whether the current number restriction on *sin* is in fact a number restriction (= singular or non-plural antecedents only) or perhaps a gender restriction (= non-male-or-female antecedents only). As previously stated, I cannot say for sure that *sin* will regain the option of being used with plural antecedents by the majority of Danish speakers but I will not be surprised if it does.

Bibliography

- A. Chacón, Dustin. 2021. "Acceptability (and Other) Experiments for Studying Comparative Syntax." In *The Cambridge Handbook of Experimental Syntax*, edited by Grant Goodall, 181–208. Cambridge Handbooks in Language and Linguistics. Cambridge University Press. doi:10.1017/9781108569620.008.
- Aarts, Bas. 2014. "distributive." In *Oxford Dictionary of English Grammar*. Oxford University Press. doi:10.1093/acref/9780199658237.013.0419.
- Abney, Steven Paul. 1987. "The English Noun Phrase in its Sentential Aspect." Thesis.
- Ackema, Peter, and Ad Neeleman. 2013. "Person features and syncretism." *Natural Language Linguistic Theory* 31 (4): 901–950. doi:10.1007/s11049-013-9202-z.
- . 2018. *Features of Person: From the Inventory of Persons to Their Morphological Realization*. The MIT Press. doi:https://doi.org/10.7551/mitpress/11145.001.0001.
- . 2019. "Default person versus default number in agreement." In *Agreement, Case and Locality in the Nominal and Verbal Domains*, edited by Ludovico Franco, Mihaela Marchis Moreno, and Matthew Reeve, 21–54. Open Generative Syntax. Language Science Press. doi:10.5281/zenodo.3458062.
- Adger, David. 2003. *Core syntax: a minimalist approach*. 424. Core linguistics. Oxford: Oxford University Press.
- Adger, David, and Peter Svenonius. 2011. "Features in minimalist syntax." In *The Oxford handbook of linguistic minimalism*, edited by Cedric Boeckx, 27–51. Oxford handbooks in linguistics. Oxford: Oxford University Press.

- Ahn, Byron. 2019. "Features, Identity, and 'Yourself'." In *NELS 49: Proceedings of the 49th Annual Meeting of the North East Linguistics Society*, edited by Maggie Baird and Jonathan Pesetsky, 1:15–24.
- Allen, Cynthia L. 2008. "2 Genitive case and the Germanic languages: Overview." In *Genitives in Early English: Typology and Evidence*, 37–59. Oxford: Oxford University Press. doi:<https://doi.org/10.1093/acprof:oso/9780199216680.003.0002>.
- Altman, Douglas G., and J. Martin Bland. 2005. "Standard deviations and standard errors." *BMJ (Clinical research ed.)* 331 (7521): 903. doi:<https://doi.org/10.1136/bmj.331.7521.903>.
- Arboe, Torben. 2016. "Gender and number peculiarities of uncountable nouns in Jutlandic (Western Danish)." In *Let us have articles betwixt us – Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood*, edited by Sten Vikner, Henrik Jørgensen, and Elly van Gelderen, 99–110. Aarhus: Aarhus Universitet. doi:10.7146/au.119.107.
- Asmussen, Jørg. 2015. *Design of the ePOS tagger*. <http://korpus.dsl.dk/clarin/corpus-doc/pos-design.pdf>.
- . 2016. *Korpusværktøjet CoREST: Manual*. <http://korpus.dsl.dk/corest/manual.pdf>.
- Baayen, R. Harald. 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press. https://www.statsbiblioteket.dk/au/#/search?query=recordID%3A%22sb_3235520%22.
- Bailey, Guy, Tom Wikle, Jan Tillery, and Lori Sand. 1991. "The apparent time construct." *Language Variation and Change* 3 (3): 241–264. doi:10.1017/S0954394500000569.
- Baker, Mark. 2008. "The macroparameter in a microparametric world." In *The limits of syntactic variation*. Linguistik aktuell ; 132. Philadelphia: John Benjamins.
- Bibelen : autoriseret af Hendes Majestæt Dronning Margrethe II*. 1992. København: Det Danske Bibelselskab.

- Biberauer, Theresa, and Ian Roberts. 2017. "Parameter setting." In *The Cambridge Handbook of Historical Syntax*, edited by Adam Ledgeway and Ian Roberts. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press.
- Biberauer, Theresa, and George Walkden. 2015. *Syntax over time : lexical, morphological, and information-structural interactions*. Oxford: Oxford University Press.
- Bjerrum, Marie, and Anders Bjerrum. 1974. *Ordbog over Fjoldemålet*. Akademisk Forlag.
- Bland, J. Martin, and Douglas G. Altman. 1986. "Statistical methods for assessing agreement between two methods of clinical measurement." *The lancet* 327 (8476): 307–310.
- Boskovic, Z. 2005. "On the locality of left branch extraction and the structure of NP." *Studia linguistica* (Oxford, UK) 59 (1): 1–45.
- Brøndum-Nielsen, Johs. 1965. *Gammeldansk Grammatik i sproghistorisk Fremstilling : V*. Copenhagen: J. H. Schultz Forlag og Akademisk Forlag.
- Burzio, Luigi. 1991. "The morphological basis of anaphora." *Journal of linguistics* (Cambridge, UK) 27 (1): 81–105.
- Canac-Marquis, Réjean. 2005. "Phases and Binding of Reflexives and Pronouns in English." In *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, edited by Stefan Müller, 482–502. <http://cslipublications.stanford.edu/HPSG/2005/canac-marquis.pdf>.
- Chomsky, Noam. 1957. *Syntactic structures*. 116. *Janua Linguarum. Studia Memoriae Nicolai van Wijk dedicata ; Nr. 4. 's-Gravenhage*.
- . 1965. *Aspects of the theory of syntax*. 251. Massachusetts Institute of Technology, Research Laboratory of Electronics. Special technical Report ; No. 11. Cambridge, Mass.: The M. I. T. Press.
- . 1975. *Questions of form and interpretation*. Lisse: Peter de Ridder Press.
- . 1981. *Lectures on government and binding*. Foris Publications.

- . 1995. *The minimalist program*. 20th. anniversary edition (2015). Current studies in linguistics ; 28. Cambridge, Mass.: The MIT Press.
- . 2000. “Minimalist inquiries: the framework.” In *Step by step : essays on minimalist syntax in honor of Howard Lasnik*, edited by Roger Martin, David Michaels, and Juan Uriagereka, 143–173. Cambridge, Mass.: MIT Press.
- . 2001. “Derivation by phase.” In *Ken Hale : a life in language*, edited by Michael J. Kenstowicz, 1–52. Current studies in linguistics ; 36. Cambridge, Mass.: MIT.
- . 2008. “On Phases.” In *Current Studies in Linguistics : Foundational Issues in Linguistic Theory : Essays in Honor of Jean-Roger Vergnaud*, edited by Robert Freidin, Carlos P. Otero, and Maria Luisa. Zubizarreta, 133–166. Cambridge, US: The MIT Press.
- Christensen, Ken Ramshøj. 2005. “Interfaces: Negation - Syntax - Brain.” Thesis. https://pure.au.dk/ws/files/41630319/KRC_Dissertation.pdf.
- . 2019. “On the need for experimental syntax.” *A sound approach to language matters. Festschrift in honor of Ocke-Schwen Bohn*: 373–388.
- Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2019. “De her sætninger er jeg vilde med!: Sproglig variation i prædikative adjektivers kongruensmønster.” In *17. Møde om Udforskningen af Dansk Sprog*, edited by Yonatan Goldshtein, Inger Schoonderbeek Hansen, and Tina Thode Hougaard, 165–183. Institut for Kommunikation og Kultur, Aarhus Universitet. <http://projekter.au.dk/muds/program-muds-17/>.
- . 2021. “Vi er (stadig) vild med den her sætning!” In *18. Møde om Udforskningen af Dansk Sprog*, edited by Yonatan Goldshtein, Inger Schoonderbeek Hansen, and Tina Thode Hougaard, 115–132. Institut for Kommunikation og Kultur, Aarhus Universitet.
- Citko, Barbara. 2014. *Phase Theory*. Cambridge: Cambridge University Press.
- Cobley, P. 2005. *The Routledge Companion to Semiotics and Linguistics*. Taylor & Francis.

- Corbett, G. G. 1979. "The agreement hierarchy." *Journal of Linguistics* 15 (2): 203–224. doi:10.1017/S0022226700016352.
- . 2023. "The Agreement Hierarchy and (generalized) semantic agreement." *Glossa: a journal of general linguistics* 8 (1). doi:https://doi.org/10.16995/glossa.9164.
- Dansk sproghistorie*. 2016. Aarhus: Aarhus Universitetsforlag.
- Danske Runeindskrifter - runedatabase*. Online Database. <https://runer.ku.dk/q.php?p=runer/home>.
- Delsing, Lars-Olof. 1993. *The internal structure of noun phrases in the Scandinavian languages: a comparative study*. Dept. of Scandinavian Languages, University of Lund.
- Despić, Miloje. 2015. "Phases, Reflexives, and Definiteness." *Syntax* 18 (3): 201–234. doi:10.1111/synt.12031.
- Diderichsen, Paul. 1939. "Om Pronominerne sig og sin." *Acta philologica Scandinavica* 13:1–95.
- Ehlers, Katrine Rosendal. 2017. "Reflexive pronouns: Danish and English in comparison." Master's thesis, Aarhus University.
- . 2020. "Sin og sig med flertalsantecedent fra runesten til LANCHART." *Danske Studier* 2020:48–84.
- Ehlers, Katrine Rosendal, and Sten Vikner. 2016. "De to slags betingelser i det danske reflexivsystem - hvorfor det er svært både *at skælde sig ud* og *at skille sig selv ud*." Presentation at 16. *Møde om Udforskningen af Dansk Sprog (MUDS 16)*. Aarhus University.
- Falk, Hjalmar, and Alf Torp. 1900. *Dansk-norskens syntax i historisk fremstilling*. 368. Kristiania: Aschehoug.
- Faltz, L.M. 1977. *Reflexivization: A Study in Universal Syntax*. University of California, Berkeley.

- Fanselow, Gisbert, and Stefan Frisch. 2006. "Effects of processing difficulty on judgments of acceptability." *Gradience in grammar: Generative perspectives*: 291–316.
- Fiva, Toril. 1987. *Possessor chains in Norwegian*. Oslo: Novus.
- Francom, Jerid. 2021. "Corpus Studies of Syntax." In *The Cambridge Handbook of Experimental Syntax*, edited by Grant Goodall, 687–713. Cambridge Handbooks in Language and Linguistics. Cambridge University Press. doi:10.1017/9781108569620.026.
- Franks, Steven, and Linda Schwartz. 1994. "Binding and Non-Distinctness: A Reply to Burzio." *Journal of Linguistics* 30 (1): 227–243. <http://www.jstor.org/stable/4176261>.
- Freidin, Robert. 1997. "Binding Theory on Minimalist Assumptions." In *Atomism and Binding*, edited by Hans Bennis, Pierre Pica, and Johan Rooryck, 141–153. Dordrecht: Foris Publications.
- Fukui, Naoki. 1986. "A Theory of Category Projection and Its Applications." Thesis. <https://dspace.mit.edu/handle/1721.1/15105>.
- Goodall, Grant. 2021. "Sentence Acceptability Experiments: What, How, and Why." In *The Cambridge Handbook of Experimental Syntax*, edited by Grant Goodall, 7–38. Cambridge Handbooks in Language and Linguistics. Cambridge University Press. doi:10.1017/9781108569620.002.
- Gregersen, Frans. 2009. "The data and design of the LANCHART study." *Acta Linguistica Hafniensia* 41:3–29.
- Haegeman, Liliane. 1995. *Introduction to government and binding theory*. 2. edition. 701. Blackwell textbooks in linguistics ; 1. Oxford: Blackwell.
- Haegeman, Liliane, and Jacqueline Guéron. 2000. *English grammar : a generative perspective*. 672. Blackwell textbooks in linguistics ; 14. Oxford: Blackwell.
- Hagedorn, Karen Thagaard, and Henrik Jørgensen. 2009. "The status of the pronoun sin in Western Jutland." Ms. University of Aarhus.
- Hansen, Aage. 1965. *Vort vanskelige sprog* [in dan]. 2. udg. København: Grafisk forlag.

- Hansen, Erik. 1993. *Rigtigt dansk*. Kbh: Hans Reitzel.
- Hansen, Erik, and Lars Heltoft. 2011. *Grammatik over det danske sprog*. Copenhagen: Det Danske Sprog- og Litteraturselskab.
- Haugen, Einar. 1976. *The Scandinavian languages : An introduction to their history*. 507. The Great Languages. London: Faber / Faber.
- Hellan, Lars. 1988. *Anaphora in Norwegian and the theory of grammar*. 308. Studies in generative grammar ; 32. Dordrecht: Foris Publications.
- Hinchliffe, Ian, and Philip Holmes. 2018. *Basic Swedish : A Grammar and Workbook*. Grammar Workbooks. Routledge.
- Holmes, Philip, and Hans-Olav Enger. 2018. *Norwegian : a comprehensive grammar*. Routledge Comprehensive Grammars. Milton Park, Abingdon, Oxon ; Routledge.
- Hornstein, Norbert. 2000. *Move! A Minimalist Theory of Construal*. Wiley.
- . 2007. "Pronouns in a minimalist setting." In *The Copy Theory of Movement*, edited by Norbert Corver and Jairo Nunes, 351–385. Amsterdam: John Benjamins Publishing Company.
- Howe, Stephen. 1996. *The personal pronouns in the Germanic languages : a study of personal pronoun morphology and change in the Germanic languages from the first records to the present day*. 390. Studia linguistica Germanica ; 43. Berlin: Gruyter.
- Huang, C.-T. James. 2002. "Distributivity and Reflexivity." In *On the Formal Way to Chinese Linguistics*, edited by Chen-Sheng Tang Sze-Wing Liu. CSLI Publications.
- Jensen, Eva Skafte. 2010. "Selv-studier." *Ny forskning i grammatik* 17:47–65.
- Juel Jensen, Torben. 2009a. "Hvad sker der med *sin* i moderne dansk – og hvorfor sker det?" *mål og mæle* 32 (1): 6–11. https://www.xn--mlogmle-exan.dk/MoM-arkiv/MoM_32/MoM32_1.pdf.
- . 2009b. "Refleksivt anvendte pronominer i moderne dansk." *Ny forskning i grammatik* 16:131–151.
- Jul Nielsen, Bent. 1986. "Om pronominet *sin* i jysk." *Danske folkemaal* 28:41–101.

- Julien, Marit. 2005. "Possessor Licensing, Definiteness and Case in Scandinavian." In *The Function of Function Words and Functional Categories*, edited by Christina M. Dikken Marcel Tortora, 217–249. Amsterdam: John Benjamins.
- Jysk Ordbog*. Online Database. <http://www.jyskordbog.dk/jyskordbog/ordbog.html>.
- Kayne, Richard S. 2002. "Pronouns and their Antecedents." In *Derivation and Explanation in the Minimalist Program*, edited by Samuel David Epstein and Daniel T. Seely, 133–166. Oxford: Blackwell.
- Keller, Frank, and Ash Asudeh. 2001. "Constraints on Linguistic Coreference: Structural vs. Pragmatic Factors." In *Proceedings of the 23rd. Annual Conference of the Cognitive Science Society*, 483–488. Mahwah, NJ: Lawrence Erlbaum Associates.
- Koopman, Hilda, and Dominique Sportiche. 1991. "The position of subjects." *Lingua* 85 (2): 211–258. doi:[https://doi.org/10.1016/0024-3841\(91\)90022-W](https://doi.org/10.1016/0024-3841(91)90022-W).
- Kratzer, Angelika. 1996. "Severing the External Argument from Its Verb." In *Phrase Structure And The Lexicon*, edited by Johan Rooryck and Laurie Zaring, 109–137. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Kristiansen, Tore, and Inge Lise Pedersen. 2006. "Sociolingvistik." *NyS, Nydanske Sprogstudier*.
- Kroman, Erik, and Stig Juul. 1945. *Danmarks gamle Love paa Nutidsdansk*. Gad.
- Kühl, Karoline. 2014. "'Det er easy at tale engelsk også'. Amerikadansk i 1960'erne og 1970'erne." *NyS, Nydanske Sprogstudier* 47 (47): 39–64. doi:[10.7146/nys.v47i47.19916](https://doi.org/10.7146/nys.v47i47.19916).
- Labov, William. 2001. *Principles of linguistic change: Social factors*. Oxford: Wiley-Blackwell.
- Langsford, Steven, Amy Perfors, Andrew T. Hendrickson, Lauren A. Kennedy, and Danielle J. Navarro. 2018. "Design sensitivity and statistical power in acceptability judgment experiments." *Glossa: a journal of general linguistics* 3 (1): 1–34. doi:<https://doi.org/10.5334/gjgl.396>.
- Lasnik, Howard, and Terje Lohndal. 2010. "Government-binding/principles and parameters theory." *Wiley interdisciplinary reviews. Cognitive science* 1 (1): 40–50. doi:[10.1002/wcs.35](https://doi.org/10.1002/wcs.35).

- Lebeaux, David. 1983. "A Distributional Difference between Reciprocals and Reflexives." *Linguistic Inquiry* 14 (4): 723–730.
- Lee-Schoenfeld, Vera. 2008. "Binding, Phases, and Locality." *Syntax* 11 (3): 281–298. doi:10.1111/j.1467-9612.2008.00118.x.
- Lightfoot, David. 2010. "Language acquisition and language change." *Cognitive science* (Hoboken, USA) 1 (5): 677–684.
- Lindstad, Arne Martinus, Anders Nøklestad, Janne Bondi Johannessen, and Øystein Alexander Vangsnes. 2009a. *The Nordic Dialect Database: Mapping Microsyntactic Variation in the Scandinavian Languages*. Edited by Kristiina Jokinen and Eckhard Bick. Online Database. The%20Nordic%20Syntax%20Database : %20https : // tekstlab.uio.no/nsd.
- . 2009b. "The Nordic Dialect Database: Mapping Microsyntactic Variation in the Scandinavian Languages." In *NEALT Proceedings Series*, edited by Kristiina Jokinen and Eckhard Bick, vol. 4. <https://tekstlab.uio.no/nsd>.
- Lockwood, W. B. 1968. *Historical German syntax*. Oxford History of the German Language ; 1. Oxford: Clarendon.
- Lødrup, Helge. 2008. "Objects Binding Reflexives in Norwegian." *Norsk Lingvistisk Tidsskrift* 28:136–160.
- . 2009. "Animacy and long distance binding in Norwegian." *Nordic Journal of Linguistics* 32 (1): 111. doi:10.1017/S0332586509002054.
- Lundquist, Björn. 2014. "Number sensitive anaphors and short distance pronouns." *Nordic Atlas of Language Structures Journal* 1 (1). doi:10.5617/nals.5423.
- McCloskey, Jim. 1997. "Subjecthood and Subject Positions." In *Elements of Grammar: Handbook in Generative Syntax*, edited by Liliane Haegeman, 197–235. Kluwer International Handbooks of Linguistics. Dordrecht: Springer Netherlands. doi:https://doi.org/10.1007/978-94-011-5420-8_5.
- Mey, S. de. 1981. "The dependent plural and the analysis of tense." In *Proceedings of NELS 11*, edited by J. Burke V.A. Pustejovsky. Amherst, GLSA: University of Massachusetts.

- Mikkelsen, Kristian. 1911. *Dansk ordføjningslære med sproghistoriske tillæg: Håndbog for viderekomne og lærere*. Copenhagen: Lehmann & Stage.
- Mitchell, Bruce. 1985. *Old English Syntax: Concord, The Parts of Speech, and the Sentence*. Oxford University Press. doi:10.1093/acprof:oso/9780198119357.003.0002.
- Mustanoja, Tauno F. 1960. *A Middle English Syntax (2016 ed.)* Netherlands: John Benjamins Publishing Company.
- Nevins, Andrew. 2007. "The Representation of Third Person and Its Consequences for Person-Case Effects." *Natural language and linguistic theory* (DORDRECHT) 25 (2): 273–313.
- . 2011. "Multiple agree with clitics: person complementarity vs. omnivorous number." *Natural language and linguistic theory* (Dordrecht) 29 (4): 939–971.
- Nevins, Andrew, and Philipp Weisser. 2019. "Closest conjunct agreement." *Annual Review of Linguistics* 5:219–241.
- Newmeyer, FJ. 1983. *Grammatical Theory: Its Limits and Its Possibilities*. University of Chicago Press.
- Noesgaard, A. 1951. *Danske dialekter og skolebørns skriftlige arbejde*. Copenhagen: Gyldendal.
- Nyvad, Anne Mette, and Ken Ramshøj Christensen. 2023. "'Vi har fundet nogle sætninger på nettet, og dem bliver du vilde med!' Når performansen halter." *MUDS - Møderne om Udforskningen af Dansk Sprog*.
- Nyvad, Anne Mette, Ken Ramshøj Christensen, and Sten Vikner. 2017. "CP-Recursion in Danish: A cP/CP-analysis." *Linguistic Review* 34 (3): 449–477. doi:10.1515/tlr-2017-0008.
- Ordbog over det Danske Sprog*. Online Database. <http://ordnet.dk/ods/forside>.
- Pedersen, Karen Margrethe. 2017. "Refleksivt sig/dem – varianter gennem 800 år." *Danske talesprog* 17:1–43.

- Pesetsky, David, and Esther Torrego. 2007. "The syntax of valuation and the interpretability of features." In *Phrasal and Clausal Architecture : Syntactic derivation and interpretation. In honor of Joseph E. Emonds*, edited by Simin Karimi, Vida Samiian, and Wendy K. Wilkins, 263–294. Linguistic Aktuell/Linguistics Today; 101. Amsterdam: John Benjamins Publishing Company.
- Pinker, Steven. 1995. *The language instinct : the new science of language and mind*. 494. London: Penguin.
- Politzer, Robert L. 1952. "On the Romance Third Person Possessives." *WORD* 8 (1): 65–71. doi:10.1080/00437956.1952.11659421.
- Pollock, Jean-Yves. 1989. "Verb Movement, Universal Grammar, and the Structure of IP." *Linguistic inquiry* (Cambridge) 20 (3): 365–424.
- Postma, Gertjan. 2011. "Language contact and linguistic complexity—the rise of the reflexive pronoun *zich* in a fifteenth-century Netherlands border dialect." In *Grammatical Change*, edited by Dianne Jonas, John Whitman, and Andrew Garrett. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199582624.003.0007.
- Preminger, Omer. 2019. "The Anaphor Agreement Effect: further evidence against binding-as-agreement (ms.)" *LingBuzz* (Tromsø).
- Python Software Foundation. 2016. *Python* (3.5.2). <https://www.python.org/downloads/release/python-352/>.
- Quicoli, Carlos A. 2008. "Anaphora by Phase." *Syntax* 11 (3): 299–329. doi:10.1111/j.1467-9612.2008.00116.x.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Rask, Rasmus. 1820. "Den danske Grammatiks Endelser og Former af det islandske Sprog forklarede. Et Forsög af Prof. R. Rask." *Det Skandinaviske Litteraturselskabs Skrifter* 17:75–150.
- Reinhart, Tanya, and Eric Reuland. 1993. "Reflexivity." *Linguistic Inquiry* 24 (4): 657–720.

- Reuland, Eric. 2001. "Primitives of Binding." *Linguistic Inquiry* 32 (3): 439–492.
- . 2005. "Agreeing to bind." In *Organizing Grammar : Linguistic Studies in Honor of Henk van Riemsdijk*, edited by Mari C. Jones and Edith Esch, 505–513. Hawthorne, NY, USA: Walter de Gruyter.
- . 2011. *Anaphora and language design*. Linguistic inquiry monographs. Cambridge, Mass: The MIT Press.
- . 2014. "Reflexives." In *Oxford Handbooks online*, edited by Sally Thomason. Oxford: Oxford University Press.
- Reuland, Eric, and Peter Zubkov. 2022. "Agreeing to bind: the case of Russian." *Glossa: a journal of general linguistics* 7 (1). doi:<https://doi.org/10.16995/glossa.5730>.
- Roberts, Jane. 2016. "Aldred, glossator and book historian." In *The Old English gloss to the Lindisfarne Gospels: language, author and context*, edited by Julia Fernández Cuesta and Sara María Pons-Sanz. Berlin: De Gruyter Mouton.
- Rooryck, Johan, and Guido Vanden Wyngaerd. 2011. *Dissolving Binding Theory*. Oxford: Oxford University Press. doi:[10.1093/acprof:oso/9780199691326.001.0001](https://doi.org/10.1093/acprof:oso/9780199691326.001.0001).
- Safir, Kenneth J. 1997. "Symmetry and Unity in the Theory of Anaphora." In *Atomism and Binding*, edited by Hans Bennis, Pierre Pica, and Johan Rooryck, 341–379. Dordrecht: Foris Publications.
- . 2004. *The syntax of anaphora*. Oxford studies in comparative syntax. New York: Oxford University Press.
- . 2014. "One True Anaphor." *Linguistic Inquiry* 45 (1): 91–124.
- Schindler, Samuel, and Karen Kiil Brøcker. 2020. "Experiments in syntax and philosophy: The method of choice?" In *Linguistic Intuitions: Evidence and Method*, edited by Samuel Schindler, Anna Drożdżowicz, and Karen Kiil Brøcker, 255–274. Oxford University Press. doi:<https://doi.org/10.1093/oso/9780198840558.003.0014>.

- Scholz, Barbara C., Francis Jeffrey Pelletier, and Geoffrey K. Pullum. 2016. "Philosophy of Linguistics." In *The Stanford Encyclopedia of Philosophy*, Winter 2016, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.
- Skautrup, Peter. 1947. *Det Danske Sprogs Historie*. Vol. II. Copenhagen: Gyldendal.
- . 1953. *Det Danske Sprogs Historie*. Vol. III. Copenhagen: Gyldendal.
- Song, Sanghoun, and Eunjeong Oh. 2016. "5 or 7: Is the Choice So Important in Acceptability Judgment Testing?" *Korean Journal of Linguistics* 41 (3): 449–480. <http://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE07040786>.
- Sørensen, Sara, Katrine Rosendal Ehlers, and Sten Vikner. 2020. "'Override'-refleksiver i dansk og engelsk." *Ny Forskning i Grammatik* 27:125–146. doi:10.7146/nfg.v0i27.122135.
- Sportiche, Dominique. 1988. "A Theory of Floating Quantifiers and Its Corollaries for Constituent Structure." *Linguistic Inquiry* 19 (3): 425–449. <http://www.jstor.org/stable/25164903>.
- Sprouse, Jon. 2015. "Three open questions in experimental syntax." *Linguistics vanguard: multimodal online journal* 1 (1): 89–100.
- Sprouse, Jon, and Diogo Almeida. 2017. "Design sensitivity and statistical power in acceptability judgment experiments." *Glossa: a journal of general linguistics* 2 (1): 1–32. doi:<https://doi.org/10.5334/gjgl.236>.
- Sprouse, Jon, Matt Wagers, and Colin Phillips. 2012. "A Test of the Relation between Working-Memory Capacity Syntactic Island Effects." *Language* 88 (1): 82–123. <http://www.jstor.org/stable/41348884>.
- Strahan, Tania E. 2011. "A typology of non-local reflexives in the Scandinavian languages." *Nordic Journal of Linguistics* 34 (2): 157. doi:10.1017/S0332586511000151.
- Sturt, Patrick. 2022. "Syntactic and semantic mismatches in English number agreement." *Glossa Psycholinguistics* 1 (1). doi:<http://dx.doi.org/10.5070/G6011159>.

- Sundareshan, Sandhya. 2020. "Distinct featural classes of anaphor in an enriched person system." In *Agree to Agree: Agreement in the Minimalist Programme*, edited by Peter W. Smith, Johannes Mursell, and Katharina Hartmann. Open Generative Syntax 6. Berlin: Language Science Press. doi:10.5281/zenodo.3528036.
- Tamm, Ditlev, and Helle Vogt. 2016. *The Danish Medieval Laws: The laws of Scania, Zealand and Jutland*. Routledge medieval translations. London: Routledge.
- Tang, Chih-Chen Jane. 1989. "Chinese Reflexives." *Natural Language Linguistic Theory* 7 (1): 93–121. <https://www.jstor.org/stable/4047781>.
- Tingsell, Sofia. 2013. "Tracking language change. Anaphoric binding in multilingual urban settings in Sweden." In *Young Urban Swedish. Variation and change in multilingual settings*, edited by Roger Källström and Inger Lindberg, 237–250. Gothenburg, Sweden: University of Gothenburg.
- Uriagereka, Juan. 1995. "Aspects of the Syntax of Clitic Placement in Western Romance." *Linguistic Inquiry* 26 (1): 79–123. <http://www.jstor.org/stable/pdf/4178889.pdf>.
- Vikner, Sten. 1985. "Parameters of binder and of binding category in Danish." *Working Papers in Scandinavian Syntax* 23:1–61. <http://www.hum.au.dk/engelsk/engsv/papers/vikn85a.pdf>.
- . 1989. "All Verb Second Clauses are CPs." *Working Papers in Scandinavian Syntax* 43:27–49. <http://www.hum.au.dk/engelsk/engsv/papers/vikn89c.pdf>.
- . 1991. "Relative der and other C0 elements in Danish." *Lingua* 84 (2-3): 109–136. doi:10.1016/0024-3841(91)90067-F.
- . 2011. "Trees and Fields and Negative Polarity." *Hermes – Journal of language and communication in business* 47:39–55. <http://www.hum.au.dk/engelsk/engsv/papers/vikn11a.pdf>.
- . 2014. "Possessorens status inden for nominalgruppen." *Ny forskning i grammatik* 21:193–211.
- . 2021. *Handout 7-8 from Comparative Syntax*, autumn.

- Vikner, Sten, and Katrine Rosendal Ehlers. 2017. "Sin, hendes og KorpusDK." *Ny forskning i grammatik*, no. 24. doi:10.7146/nfg.v25i24.97252.
- Wellejus, Gerd. 1972. *Jysk, sjællandsk eller skånsk? : en undersøgelse af kriterierne for jysk dialekt, baseret på det gammeldanske lovsprog*. Fotografisk optryk af manuskript. 142. Studier ; 2. Universitetsforlaget i København.
- Wurmbrand, Susi. 2003. "Agree: The other VP-internal subject hypothesis." *University of Pennsylvania Working Papers in Linguistics* 8 (1). <http://repository.upenn.edu/pwpl/vol8/iss1/18/>.
- . 2014. "The Merge Condition: A syntactic approach to selection." In *Minimalism and Beyond: Radicalizing the interfaces*, edited by P. Kosta, S.L. Franks, T. Radeva-Bork, and L. Schürcks, 130–166. Amsterdam: John Benjamins Publishing Company.
- Zeijlstra, Hedde. 2008. "On the syntactic flexibility of formal features." In *The Limits of Syntactic Variation*, edited by Theresa Biberauer, 143–173. *Linguistic Aktuell/Linguistics Today*; 132. Amsterdam: John Benjamins Publishing Company.

Appendix A

Instructions given to participants and teachers about the acceptability study

A.1 Instructions given to participants

A.1.1 Danish original

Dette spørgeskema tager ca. 10 minutter at gennemføre, og dine svar har stor værdi for mig og min forskning. Undersøgelsen er en del af min ph.d.-afhandling på Aarhus Universitet, og hvis du skulle have spørgsmål eller kommentarer, er du meget velkommen til at skrive til mig på mailen karo@cc.au.dk.

Alle spørgsmålene ser nogenlunde ens ud - du kan se et eksempel nederst på siden. Hvert spørgsmål består af en sætning og en skala fra 1 (Helt uacceptabel) til 5 (Helt acceptabel). Denne skala skal du bruge til at vurdere sætningen. Der er ingen rigtige eller forkerte svar, og din første indskydelse er mest interessant for mig.

I din vurdering kan du forestille dig at sætningen bliver sagt af en af dine venner. Får den ham eller hende til at lyde som en der taler dansk? Indholdet er mindre vigtigt.

Jeg er interesseret i det sprog, du og andre på din alder faktisk bruger, og ikke i det, du har fået at vide, er rigtigt. Det er derfor helt optimalt, hvis du bare kører spørgeskemaet hurtigt igennem og vælger din første vurdering for hver sætning. Så er du også hurtigere færdig. Men selvfølgelig uden at det bliver så hurtigt, at dine svar bliver helt meningsløse!

Dine svar er naturligvis anonyme og bliver ikke brugt til andet end min analyse. Som en del af spørgeskemaet spørger jeg bl.a. om din alder og i hvilken del af Danmark, du er vokset op, men det kan ikke bruges til at identificere dig på nogen måde.

Som en del af din deltagelse giver du tilsagn til at jeg må opbevare dine anonyme, indtastede data udelukkende til forskningsmæssig brug. Tak for, at du vil være med!

A.1.2 English translation (only used in this appendix)

This questionnaire will take you approximately 10 minutes to finish, and your responses are very valuable to me and my research. The study is part of my PhD research at Aarhus University, and if you have any questions or comments you can contact me at karo@cc.au.dk.

Every question looks more or less the same – you can find an example in the bottom of this page. Every question consists of a sentence and a scale from 1 (Completely unacceptable) to 5 (Completely acceptable). You will use this scale to judge the sentence. There are no correct or wrong answers and your initial reaction is the most interesting to me.

You can imagine that the sentence is being spoken by one of your friends. Does it make him or her sound like someone who speaks Danish? The contents are less important.

I am interested in the language that you and others of your age actually use and not so much in what you have been taught is correct. For that reason, it is great if you just run quickly through the questionnaire quickly and go with your first gut feeling for each sentence. That also means that you are done quicker. But preferably without it being so fast that your answers become totally meaningless!

Your responses are completely anonymous and they will not be used for anything outside of my analysis. As part of the questionnaire I ask e.g. for your age and the part of Denmark that you grew up in but this information cannot be used to identify you in any way. By taking part in this study, you consent to me storing your anonymous responses solely for research purposes. Thank you for taking part!

A.2 Instructions given to the teachers who run the experiment in the classroom

A.2.1 Danish original

Kære gymnasielærer XYZ

Hermed den udpenslende tekst, samt link til undersøgelsen. Den er skrevet som en

generel introduktion til de lærere, jeg sender spørgeskemaet ud til, og en mere specifik beskrivelse af spørgeskemaet. Skriv meget gerne hvis du har spørgsmål. På forhånd mange tak for hjælpen!

Overordnet beskrivelse:

I min ph.d. undersøger jeg sprogforandringer i brugen af pronominer i dansk, herunder regional variation og variation mellem aldersgrupper. Undersøgelsen handler altså ikke om at vurdere hvad der er "godt" eller "dårligt" dansk, men om hvordan unge i gymnasiealderen faktisk bruger sproget. Til det formål samler jeg sproglige data fra studerende på gymnasier forskellige steder i landet (bl.a. Esbjerg, Sønderborg, Aarhus, Brønderslev, Odense, Roskilde, København, Aasiaat i Grønland). Selve undersøgelsen består af et spørgeskema med en række sætninger, der skal vurderes på en skala fra 1 til 5. Det tager 10-15 minutter at gennemføre, og den kan udfyldes online fra de studerendes egne teknologiske gadgets, f.eks. en hel klasse ad gangen.

I min drømmesituation kan jeg få lov til at få to klasser til at udfylde skemaet, f.eks. som indledning til en time. Med en lille introduktion burde det tage 20-25 minutter i alt når hele klassen sidder samtidig og udfylder skemaet på egne computere.

Nærmere instruktioner:

Her er linket til eleverne (I skal også være meget velkomne til at give det et forsøg selv. Så har I også en fornemmelse for hvad jeg udsætter jeres elever for :) - men skriv gerne en afsluttende kommentar i skemaet om at det er udfyldt af en ikke-gymnasieelev):
[LINK DELETED](#)

Forslag til hvordan spørgeskemaet kan distribueres og forklares:

Eleverne kan få linket tilsendt, og de kan udfylde spørgeskemaet på tablet/computer/telefon/... Det skal gerne foregå samlet i klassen så der er nogenlunde tjek på at de ikke sidder og laver alt muligt andet samtidig. De må gerne f.eks. høre musik mens de udfylder det, hvis det kan hjælpe på koncentrationen. Det burde ikke tage mere end 10 minutter at udfylde, og de skal kun udfylde det én gang.

Der er en ret udførlig vejledning i selve spørgeskemaet (den er copy-pastet ind nedenfor, i sort), der både forklarer hvordan det udfyldes og en lille smule om formålet, uden dog at gå for meget i detaljer. I behøver derfor i virkeligheden ikke sige så meget

om det. Eleverne skal helst gå ind til spørgeskemaet med så lidt forudgående information som muligt, for så vidt muligt at undgå at påvirke deres svar på forhånd.

Et par mulige tvivlsspørgsmål:

- Som det sidste spørgsmål af de demografiske (på første side) beder jeg dem om bare at trykke på den første/øverste valgmulighed og klikke sig videre derefter. Det virker lidt sært, men det er for at sikre at de ser forskellige versioner af spørgeskemaet. Så de skal altså bare vælge den første og klikke sig videre.
- Nogle af sætningerne ér ren ordsalat.
- Indholdet er mindre vigtigt end sætningsopbygningen – og indholdet i nogle af sætningerne hører måske nærmere til i en science fiction-bog end i virkeligheden.
- De må gerne læse sætningerne som talesprog.
- De skal ikke sammenligne sætningerne med hinanden. Bare fordi de har givet én sætning 5, behøver de ikke give den næste 5 bare fordi den ligner.
- Der er fire forskellige versioner af spørgeskemaet, og de udfylder derfor ikke alle det samme.
- Kommateringen betyder ikke noget.

Efter spørgeskemaet:

Hvis eleverne er interesserede i at vide mere om, hvad de har bidraget til, må de meget gerne få det at vide – at jeg undersøger brugen af de refleksive pronominer, specifikt sin med flertalsantecedent, i dansk, og at deres svar kan hjælpe med at afdække hvor sproget måske er på vej hen på det område. Nemlig måske mod større anvendelse af sin i pluralis, på samme måde som sig formentlig har ændret sig fra at være anvendt kun i singularis – Grundtvig skrev f.eks. *de skammer dem* og ikke *de skammer sig* og man hører ofte sætninger a la *Forældrene hentede sine børn i børnehaven* i moderne talesprog. Spørgeskemaet kommer hele vejen rundt i både refleksiverne og deres ikke-refleksive modsvar, så jeg forhåbentlig kan danne mig et solidt billede af hvordan gymnasiealdersgruppen bruger dem (og formentlig, siger i hvert fald min hypotese, bruger dem anderledes end sprogbrugere fra deres forældres generation). Det er velkendt at der også er (eller i hvert fald har været) en god portion dialektal variation i reflektivbrugen, så det bliver også interessant at sammenligne på tværs af gymnasierne til slut.

A.2.2 English translation (only used in this appendix)

Dear gymnasium teacher XYZ

Here follows the promised expanded description and a link to the questionnaire. This piece of text is written as a general introduction to the teachers that I send the questionnaire to, and it is also a more specific description of the questionnaire. Please do write back if you have any questions. I am very grateful for your participation and assistance!

General description:

I study various types of variation in the use of Danish pronouns, among these is regional variation and variation between age groups. The study is not aimed at judging what is "good" or "bad" Danish, but rather at studying how young speakers of gymnasium age actually use the language. For this purpose I am collecting linguistic data from students at gymnasiums in various areas of the country (e.g. Esbjerg, Sønderborg, Aarhus, Brønderslev, Odense, Roskilde, København, Aasiaat in Greenland). The study itself consists of a questionnaire with a number of sentences that must be rated on a scale from 1 to 5. The questionnaire can be completed in 10-15 minutes and it can be fill in online from the students' own devices, e.g. an entire class at a time.

My dream scenario is that you will let two classes fill out the questionnaire, e.g. in the beginning of a lesson. It should take about 20-25 minutes all in all, including time for a little introductory talk, if all students fill in the questionnaire concurrently on their own devices.

Further instructions:

—————
This is the link for the students (you are also very welcome to try out the questionnaire yourself. That also gives you an impression of what it is I am asking your students to do :) - but do please add a final comment in the questionnaire that you have filled it in as a non-gymnasium-student):

LINK DELETED

Suggestions on how the questionnaire can be distributed and explained:

You can send the link to the students and they can fill in the questionnaire on their tablet/computer/phone... This should preferably take place during a lesson in order

to make reasonably sure that the students do not do too many other things while filling in the questionnaire. The students can e.g. listen to music while filling in the questionnaire if that helps their concentration. The questionnaire should not take more than 10 minutes to complete and the students should only fill in the questionnaire once.

There is a fairly elaborate set of instructions in the questionnaire (copy-pasted below, in black) that both explain how the questionnaire should be filled in and a little bit about the purpose of the study without being too detailed. For that reason, you do not have to explain too much about the study beforehand. The students should preferably complete the questionnaire with as little extra information as possible in order to avoid biasing their answers.

A number of possible points for elaboration:

- As the last demographic question (on the first page of the questionnaire) I ask the students to just choose the first/highest option and then click "Next". This seems a little weird but it makes sure that students see different versions of the questionnaire. So they should just choose the first option and move on to the questionnaire.
- Some of the sentences are indeed just pure gibberish.
- The content of the sentence is less important than the structure of the sentence – and the content of some of the sentences may more properly belong in a sci-fi book than in reality.
- The students are welcome to read the sentences as though they were spoken language.
- The students should not compare the sentences to each other. Just because they rate one sentence as a 5, they do not have to rate the next sentence as a 5 just because they seem similar.
- There are four different versions of the questionnaire, which means that not all students will fill in the same questionnaire.
- The specific punctuation choices does not matter.

After filling in the questionnaire:

If the students are interested in knowing more about what they have contributed to, please do let them know. I investigate the use of the reflexive pronouns, specifically *sin* with plural antecedents (subjects), in Danish, and that their answers can help with

uncovering where the language may be heading in that particular linguistic area. Danish may be moving towards a more frequent use of *sin* with plural subjects, just like the other reflexive pronoun *sig* most likely changed from being used mainly with singular subjects – Grundtvig, e.g., wrote *de skammer dem* and not *de skammer sig* and you often hear sentences like *Forældrene henter sine børn i børnehaven* in modern spoken Danish. The questionnaire deals with the reflexive pronouns as well as the non-reflexive pronouns, in the hopes that I will be able to form a comprehensive picture of how speakers in this cohort use the pronouns (and probably, or that is at least my hypothesis, how they use them differently from speakers from their parents' generation). It is well known that there is (or least that there has been) some amount of dialectal variation in the use of reflexives which means that it will also be interesting to compare responses between gymnasiums at the end of the study.

Appendix B

Scripts

B.1 R scripts

This appendix contains various scripts written for the data collection in this project. One script, fig. B.1, is written by Prof. Søren Feodor Nielsen (CBS). Another script, fig. B.2, is my modified version of Feodor's original script. The other scripts are my own work.

Figure B.1, `Katrine.indlæs.data.R`, is a code for extracting data from a single file from KorpusDK. The original code was written by Prof. Søren Feodor Nielsen (CBS) for this project, which I very gratefully acknowledge.

```

1 corpus <- read.table('corpus.txt',fill=T,encoding='UTF-8',stringsAsFactors=F)
2
3 #encoding må du nok rode lidt med; i den fil du har sendt til mig er
4 #æ, ø og å tabt.
5 #fill=T fordi der ikke er data nok i fx linie 1
6 #stringsAsFactors=F er bare 'pynT'; du behøver ikke data indlæst som faktorer,
7 #så jeg slår det fra, så det bare bliver læst ind som tekst
8
9 sentences <- list()
10 sine <- c()
11
12 j <- 1
13 cat('',file='sentences.txt') #åbner en tom fil
14
15 for (i in 1:dim(corpus)[1]){
16     if (corpus[i,1]=='<s') sentence <- i else sentence <- c(sentence,i)
17
18     if (corpus[i,1]=='</s>'){
19         sentences[[j]] <- corpus[sentence,]
20
21         if (any(corpus[sentence,1] %in% c('sin','sit','sine'))){
22             sine <- c(sine,j)
23
24             cat(as.character(corpus[sentence[-c(1,length(sentence))],1]),'\n',
25                 file='sentences.txt',append=T)
26         }
27     }
28     j <- j+1
29 }
30
31 write.table(sentences[[sine[1]]],file='output.txt',row.names=F)
32
33 for (j in sine[-1]) write.table(sentences[[j]],file='output.txt',
34                                 append=T,row.names=F,col.names=F)
35 }
36 }
37
38
39
40
41

```

Figure B.1: R-script that can be used to open a .txt file from the offline version of KorpusDK and provide an output that contains every sentence with *sin*, *sit*, or *sine* in the file.

Figure B.2, `Katrine.indlæs.data-UPDATED.R`, is my own code for extracting data from a large number of files from KorpusDK. It is a modified version of the code in fig. B.1. Notice the working directory, where this particular code has been run in a directory with the first thousand files from KorpusDK. The reason for doing this is that the code would otherwise take several days to run if the input was all 10,000 files from a corpus directory. This is a good indication that this code is not particularly efficient at handling large amounts of data.

```

1 setwd("C:/Users/kcarstensen/Documents/Korpus90-2000-2010/KDK-1990.scrambled.0-1000")
2 corpus <- read.table('0000.txt',fill=T,encoding='UTF-8',stringsAsFactors=F,
3                     quote = "") #NB: tilføjet quote="" for at få al data med
4
5 filelist = list.files(pattern = ".*.txt")
6
7 #assuming tab separated values with a header
8 datalist = lapply(filelist,
9                   function(x)read.table(x, header = F,
10                                       fill = T, quote = "",
11                                       stringsAsFactors=F, encoding="UTF-8"))
12 #assuming the same header/columns for all files
13 corpus = do.call("rbind", datalist)
14 sentences <- list()
15 sine <- c()
16
17 j <- 1
18 cat('',file='sentences.txt') #åbner en tom fil
19
20 for (i in 1:dim(corpus)[1]){
21   if (corpus[i,1]=='<s') sentence <- i else sentence <- c(sentence,i)
22
23   if (corpus[i,1]=='</s>'){
24     sentences[[j]] <- corpus[sentence,]
25
26     if (any(corpus[sentence,1] %in% c('sin', 'sit', 'sine'))){
27       sine <- c(sine,j)
28
29       cat(as.character(corpus[sentence[-c(1,length(sentence))],1]),
30         '\n',file='sentences.txt',append=T)
31     }
32     j <- j+1
33   }
34 }
35 write.table(sentences[[sine[1]]],file='output.txt',row.names=F)
36
37 for (j in sine[-1])
38   write.table(sentences[[j]],file='output.txt',append=T,row.names=F,col.names=F)

```

Figure B.2: R-script that can be used to open any number of .txt file from the offline version of KorpusDK and provide an output that contains every sentence with *sin*, *sit*, or *sine* in the file.

B.2 Python scripts

Figure B.3 is a script which I have called `python-corpus-randomisation.py`. It takes a random selection of 5 % of the sentences in a file and prints them to another file.

```
import random

with open("//uni.au.dk/Users/au280773/Desktop/sin-1990-2000/2010_sin.txt") as f:
    data = f.readlines()

print(len(data))
random.shuffle(data)

wf = open("//uni.au.dk/Users/au280773/Desktop/sin-1990-2000/2010_sin_random-5-procent.txt", "w")
for line in data[:int(0.05*len(data))]:
    wf.writelines(line)
wf.close()
```

Figure B.3: Python script that outputs a random selection of 5 % of the sentences in the given file.

Figure B.4 is a script that takes two files as input and prints the differences between the files to a third file. I used this to compare the different automatic approaches to minimizing the amounts of irrelevant examples in the corpus study.

```
### Nu lidt flottere og nemmere at bruge: finder
### ens data og printer det der er ens
def print_same(infile1,infile2,outfile):
    with open(infile1,'r') as file1:
        with open(infile2, 'r') as file2:
            same = set(file1).intersection(file2)

    same.discard('\n')

    with open(outfile, 'w') as file_out:
        for line in same:
            file_out.write(line)

### Finder forskellig data og printer det der er forskelligt
def print_different(infile1,infile2,outfile):
    with open(infile1,'r') as file1:
        with open(infile2, 'r') as file2:
            different = set(file1).difference(file2)

    different.discard('\n')

    with open(outfile, 'w') as file_out:
        for line in different:
            file_out.write(line)
```

Figure B.4: Python script that outputs a random selection of 5 % of the sentences in the given file.

Appendix C

Acceptability judgment test items

C.1 Pilot study

The following tables contain all the test items from the pilot study. An obvious change between the pilot study and the main study is the fact that there is no initial capitalization and no punctuation in the test items for the pilot study. I have glossed the sentences for the main study.

Table C.1: Pilot: Training items

Code	Sentence
400000	de køber en sofa i genbrugsbutikken
400001	tastaturet larmer når man bruger det
400002	i går har jeg kage med fra arbejde
400003	manden som at der gik på vejen havde en hund med

Table C.2: Pilot: Primary test items

Code	Sentence
0 0 1 1 0 1	de plejer at starte sit prøvesignal klokken 12
0 0 1 1 1 0	ungerne lover at trække sin cykel på fortovet
0 0 0 0 0 1	de sælger sine produkter godt
0 0 0 0 1 1	dronerne gennemfører sine missioner om natten
0 0 0 1 1 0	forfatterne aftaler at udvikle sine hovedpersoner sammen
0 0 1 0 0 0	de finder sit job på nettet
0 0 1 1 1 0	deltagerne spås at beholde sin førsteplads i kagekonkurrencen
0 0 1 0 0 0	de følger sit barn i skole
0 0 1 0 1 1	regionerne holder sit budget
0 0 1 1 0 0	de ventes at hente sit adoptivbarn i Kina
0 0 1 0 0 1	de vender sin blomst mod solen
0 0 0 1 1 1	taxierne kører for at flytte sine kunder fra A til B
0 0 0 1 1 1	skoene ventes at forbedre sine markedsandele hvert år
0 0 0 1 0 1	de begynder at åbne sine døre
0 0 1 1 1 1	bøgerne fortjener at vinde sin konkurrence
0 0 0 0 1 0	gymnasterne taber sine handsker langt over jorden
0 0 1 0 1 1	træerne skjuler sin rod under jorden
0 0 0 1 0 1	de skal til at kaste sine bomber
0 0 0 0 0 0	de har sine egne fester efter forestillingen
0 0 0 0 0 0	de stikker sine sværd i dragen
0 0 0 1 0 0	de undgår at vise sine magiske evner
0 0 0 0 0 1	de beskytter sine brugere
0 0 1 0 0 1	de mister sin farve over tid
0 0 1 1 0 0	de forventer at få sin mobiltelefon i frikvarteret
0 0 0 1 1 0	damerne plejer at begynde sine træninger med mavebøjninger
0 0 0 0 1 1	redskaberne efterlader sine karakteristiske mærker
0 0 1 0 1 0	betjentene lægger sit våben tilbage i bilen
0 0 0 1 0 0	de plejer at fastholde sine forklaringer
0 0 1 0 1 0	fuglene ændrer sin sang i regnvej
0 0 0 0 1 0	medlemmerne sender sine forslag til bestyrelsen
0 0 1 1 0 1	de er brygget til at bevare sin chokoladenuance længe efter åbning
0 0 1 1 1 1	sprinklerne er gode til at sprede sin vandstråle ud over haven

Table C.3: Pilot: Secondary test items

Code	Sentence
100000	det var sin mor Peter mødte
100001	det var sin underbo, der gjorde Peter sur
100002	hun hentede sine pakker på posthuset
100003	hun vidste godt at sine pakker var på vej
100004	Johan fandt sin kat i garagen
100005	jeg tror at Magnus holder af hans kat
100006	Morten ved at Sofie køber hendes grøntsager på torvet
100007	han tror at hun køber sine varer på nettet
100008	hun bad mig passe hendes kat
100009	hun bad mig passe sin kat
100010	hun vidste at jeg passede sin kat
100011	hun vidste at jeg passede hendes kat
100012	vi elsker Magnus' tegning af sin kat
100013	jeg så Peters fotografi af hans forældre og var meget imponeret
100014	jeg prøvede at genskabe Julies tegning af sine heste
100015	vi lovede ham at tage sin plakat med
100016	Mathilde lovede mig at vaske sit hår
100017	hun bad mig hjælpe sig med lektierne
100018	hun bad mig hjælpe sig selv med lektierne
100019	hun slog sig med en grydeske
100020	hun slog sig selv med en grydeske
100021	vi så ham ramme ham selv med malingen
100022	han ramte sig selv med malingen
100023	de skammer dem over deres gæld
100024	de skammer sig over deres gæld
100025	vi fortalte hende om hende selv
100026	jeg fortalte Morten om ham selv
100027	vi fortalte hende om sig selv
100028	hun viste drengen ham selv i spejlet
100029	jeg viste drengen sig selv i spejlet

Table C.4: Pilot: Fillers

Code	Sentence
200000	guldmedaljer har han mange
200001	han har lovet at læse ikke den
200002	Bo havde læst bøgerne aldrig
200003	måske ikke Peter kommer
200004	jeg har en moster bor i København
200005	men det var ikke for det ikke vi gjorde det
200006	fik du filmen set
200007	jeg opdagede at jeg havde ikke læst den her bog før
200008	gæsten blev bagt en kage
200009	se dig lidt for og ikke vær så fjollet
200010	I er nødt til at få det gjort fordi jo tidsfristen udløber den 7.
200011	han spurgte hvorfor ikke hun kom til tiden
200012	hvis at du kommer på besøg, bliver vi meget glade
200013	jeg har kufferterne pakket
200014	når der ikke var flere øller tilbage, gik de hjem
200015	hvis at du kommer på besøg, bliver vi meget glade
200016	I må se at få jer lavet et udhus
200017	når så at du kommer i morgen, kan vi snakke om det der
200018	der kommer et renere miljø hvis også man får gennemført denne her varmeplan
200019	jeg fik min cykel stjålet
200020	måske han ikke kommer
200021	Per gav aldrig Elsa bogen
200022	de som ofte går på cafe behøver ikke at drikke kaffe hjemme
200023	guldmedaljer har han tre af
200024	Bo havde aldrig læst bøgerne
200025	hvis endelig der kommer noget sne, så smelter det nok før vi får fri
200026	Ken fik bilen repareret på værkstedet
200027	når du kommer, skal du tage aviser med
200028	Ib har jo givet Lise bogen
200029	de fik solgt den på loppemarkedet

Table C.5: Pilot: *Awake* items

Code	Sentence
300000	Mia elsker is
300001	børnene spiller bold i frikvarteret
300002	isen på glammer af grønlig
300003	at købte i morges vi

C.2 Main study

This tables in this section contains every item that was included in the acceptability judgment study. The first tables show the Training items, the Awake items, the Secondary test items, and the Fillers. The subsequent tables show the specific Primary sentences that were included in each of the four different questionnaires.

Table C.6: Main study: Training items

Code	Sentence
400000	De køber en sofa i genbrugsbutikken. They buy a couch in charity-shop.DEF
400001	Tastaturet ikke larmer når man bruger det. Keyboard.DEF not make-noise when one uses it
400002	I går har jeg kage med fra arbejde. I yesterday have I cake with from work
400003	Manden som at der gik på vejen havde en hund med. Man.DEF who that who walked on road.DEF had a dog with

Table C.7: Main study: *Awake* items

Code	Sentence
300000	Mia elsker is. Mia loves icecream
300001	Børnene spiller bold i frikvarteret. Kids.DEF play ball in break.DEF
300002	Isen på glammer af grønlig. Ice.DEF on barks(V) of greenish
300003	At købte i morges vi. To bought in morning we

Table C.8: Main study: Secondary test items

Code	Sentence
100000	Det var sin mor Peter mødte. It was REFL's mother Peter met.
100001	Det var sin underbo, der gjorde Peter sur. It was REFL's downstairs-neighbour who made Peter mad.
100002	Hun hentede sine pakker på posthuset. She collected REFL's parcels at post-house.DEF
100003	Hun vidste godt at sine pakker var på vej. She knew well that REFL's parcels were on way
100004	Johan fandt sin kat i garagen. Johan found REFL's cat in garage.DEF
100005	Magnus havde en kat med et fjollet navn. Nu vil han kalde hans nye kat det samme. Magnus had a cat with a silly name. Now will he name his new cat the same.
100006	Morten ved at Sofie fik hendes hudsygdom af stress. Morten knows that Sofie got her skin-disease from stress
100007	Han tror at hun køber sine varer på nettet. He thinks that she buys REFL's groceries on web.DEF
100008	Hun bad mig passe hendes kat. She asked me look-after her cat
100009	Hun bad mig passe sin kat. She asked me look-after REFL's cat
100010	Hun vidste at jeg passede sin kat. She knew that I looked-after REFL's cat
100011	Hun vidste at jeg passede hendes kat. She knew that I looked-after her cat
100012	Vi elsker Magnus' tegning af sin kat. We love Magnus' drawing of REFL's cat
100013	Jeg så Peters fotografi af hans forældre og var meget imponeret. I saw Peter's photograph of his parents and was very impressed
100014	Jeg prøvede at genskabe Julies tegning af sine heste. I tried to recreate Julie's drawing of REFL's horses
100015	Vi lovede ham at tage sin plakat med. We promised him to bring REFL's poster with
100016	Mathilde lovede mig at vaske sit hår. Mathilde promised me to wash REFL's hair
100017	Hun bad mig hjælpe sig med lektierne. She asked me help REFL with homework.DEF.PL
100018	Hun bad mig hjælpe sig selv med lektierne. She asked me help REFL self with homework.DEF.PL

Table C.9: Main study: Secondary test items (contd.)

Code	Sentence
100019	Han hader sig. He hates REFL
100020	Han hader sig selv. He hates REFL self
100021	Vi så ham ramme ham selv med malingen. We saw him hit himself with paint.DEF
100022	Han ramte sig selv med malingen. He hit REFL self with paint.DEF
100023	De skammer dem over deres gæld. They shame them over their debt
100024	De skammer sig over deres gæld. They shame REFL over their debt
100025	Vi fortalte hende om hende selv. We told her about herself
100026	Jeg fortalte Morten om ham selv. I told Morten about himself
100027	Vi fortalte hende om sig selv. We told her about REFL self
100028	Hun viste drengen ham selv i spejlet. She showed boy.DEF himself in mirror.DEF
100029	Jeg viste drengen sig selv i spejlet. I showed boy.DEF REFL self in mirror.DEF

Table C.10: Main study: Fillers (approximate acceptability level from Lindstad et al. 2009a indicated in parentheses)

Code	Sentence
200000	Guldmedaljer har han mange. (1) gold-medals has he many
200001	Han har lovet at læse ikke den. (1) he has promised to read not it
200002	Bo har læst bøgerne aldrig. (1) Bo has read books.DEF never
200003	Måske ikke Peter kommer. (1) maybe not Peter comes
200004	Jeg har en moster bor i København. (1) I have an aunt lives in Copenhagen
200005	Men det var ikke for det ikke vi gjorde det. (1) but it was not for that not we did it
200006	Fik du bogen læst. (2) got you book.DEF read
200007	Jeg opdagede at jeg havde ikke læst den her bog før. (2) I discovered that I had not read this here book before
200008	Gæsten blev bagt en kage. (2) guest.DEF was baked a cake
200009	Se dig lidt for og ikke vær så fjollet. (2) see you a little for and not be so silly
200010	I er nødt til at få det gjort fordi jo tidsfristen udløber den 7. (2) you are required to to get it done because yeah deadline.DEF expires the 7.
200011	Han spurgte hvorfor ikke hun kom til tiden. (2) he asked why not she came to time.DEF
200012	Hvis at du kommer på besøg, bliver vi meget glade.(3) if that you come on visit become we very happy
200013	Jeg har kufferterne pakket. (3) I have suitcases.DEF packed
200014	Når der ikke var flere øller tilbage, gik de hjem. (3) when there not were more beers left went they home
200015	Fordi at du kommer på besøg, bliver vi meget glade. (3) because that you come to visit become we very happy
200016	I må se at få jer lavet et udhus. (3) you must see to get you made an outhouse
200017	Når så at du kommer i morgen, kan vi snakke om det dér. (3) when then that you come in morning.DEF can we talk about that there

Table C.11: Main study: Fillers (contd.)

Code	Sentence
200018	Der kommer et renere miljø hvis også man får gennemført denne her varmeplan. (4) there comes a cleaner environment if also one gets completed this here heating-plan
200019	Jeg fik min cykel stjålet. (4) I got my bike stolen
200020	Måske han ikke kommer. (4) maybe he not comes
200021	Per gav aldrig Elsa bogen. (4) Per gave never Elsa book.DEF
200022	De som ofte går på cafe, behøver ikke at drikke kaffe hjemme. (4) they who often to go cafe need not to drink coffee home
200023	Guldmedaljer har han tre af. (4) gold-medals has he three of
200024	Bo havde aldrig læst bøgerne. (5) Bo had never read books.DEF
200025	Hvis endelig der kommer noget sne, så smelter det nok før vi får fri. (5) if finally there comes some snow then melts it probably before we get free
200026	Ken fik bilen repareret på værkstedet. (5) Ken had car.DEF repaired at garage.DEF
200027	Når du kommer, skal du tage aviser med. (5) when you come shall you take newspapers with
200028	Ib har jo givet Lise bogen. (5) Ib has yeah given Lise book.DEF
200029	De fik solgt den på loppemarkedet. (5) they got sold it on flea-market.DEF

Table C.12: Main study: Primary test items (Questionnaire 1)

Code	Sentence
0 0 0 0 0	De skjuler sine formuer i skattely. they hide REFL's assets in tax haven They hide their assets in a tax haven.
0 0 0 0 0	De har sine egne fester efter forestillingen. they have REFL's own parties after show.DEF They have their own parties after the show.
0 0 0 0 1	De beskytter sine brugere mod skadelig software. they protect REFL's users against harmful software They protect their users against harmful software.
0 0 0 0 1	De fortsætter sine svingninger i et vacuum. they continue REFL's oscillations in a vacuum They continue their oscillations in a vacuum.
0 0 0 1 0	Medlemmerne sender sine forslag til bestyrelsen. members.DEF send REFL's suggestions to board.DEF The members send their suggestions to the board.
0 0 0 1 0	Tvillingerne taber sine nøgler i bussen. twins.DEF lose REFL's keys in bus.DEF The twins lose their keys in the bus.
0 0 0 1 1	Træerne strækker sine grene ud over åen. trees.DEF stretch REFL's branches out over stream.DEF The trees stretch their branches over the stream.
0 0 0 1 1	Dronerne gennemfører sine missioner om natten. drones.DEF complete REFL's missions at night.DEF The drones complete their missions at night.
0 0 0 1 0 0	De plejer at fastholde sine forklaringer. they use to maintain REFL's statements They usually maintain their statements.
0 0 0 1 0 0	De prioriterer at folde sine skjorter pænt sammen. they prioritize to fold REFL's shirts nicely together They prioritize folding their shirts nicely.
0 0 0 1 0 1	De begynder at åbne sine døre. they begin to open REFL's doors They start to open their doors.
0 0 0 1 0 1	De programmeres til at kaste sine bomber automatisk. they program.PASS to to throw REFL's bombs automatically They are programmed to throw their bombs automatically.
0 0 0 1 1 0	Babyerne begynder at ligne sine forældre på en prik. babies.DEF begin to look-like REFL's parents on a spot The babies start to look exactly like their parents.
0 0 0 1 1 0	Forfatterne aftaler at udvikle sine hovedpersoner sammen. authors.DEF agree to develop REFL's main-characters together The authors agree to develop their main characters together.
0 0 0 1 1 1	Skoene ventes at forbedre sine markedsandele næste år. shoes.DEF expect.PASS to improve REFL's market-shares next year The shoes are expected to improve their market shares next year.
0 0 0 1 1 1	Motorcyklerne plejer at klare sine 0-100 km/t på 4 sekunder. motor-bikes.DEF use to manage REFL's 0-100 km/hour on 4 seconds The motor bikes usually manage their 0-100 kilometers an hour in 4 seconds.

Table C.13: Main study: Primary test items (Questionnaire 1) (contd.)

Code	Sentence
0 0 1 0 0 0	De finder sit nye job på nettet. they find REFL's new job on web.DEF They find their new job on the internet.
0 0 1 0 0 0	De følger sit barn i skole. they follow REFL's child in school They take their child to school.
0 0 1 0 0 1	De mister sin farve over tid. they lose REFL's colour over time They lose their colour over time.
0 0 1 0 0 1	De kaster sin farlige UV-stråling på planetens få beboere. they throw REFL's dangerous UV-radiation on planet.DEF's few citizens They send their dangerous UV radiation onto the few inhabitants of the planet.
0 0 1 0 1 0	Betjentene lægger sit våben tilbage i bilen. officers.DEF put REFL's weapon back in car.DEF The officers put their weapon back in the car.
0 0 1 0 1 0	Fuglene ændrer sin sang i regnvejr. birds.DEF change REFL's song in rain-weather The birds change their song in rainy weather.
0 0 1 0 1 1	Regionerne holder sit budget. regions.DEF hold REFL's budget The regions stick to their budget.
0 0 1 0 1 1	Mange engelske ord har sin rod i det danske sprog. many English words have REFL's root in the Danish language Many English words have their origin in the Danish language.
0 0 1 1 0 0	De forventer at få sin mobiltelefon i frikvarteret. they expect to get REFL's cellphone in break.DEF They expect to get their cellphone in the break.
0 0 1 1 0 0	De venter på at få sit login med posten. they wait on to get REFL's login with mail.DEF They are waiting to get their login in the mail.
0 0 1 1 0 1	De plejer at starte sit prøvesignal klokken 12. they use to start REFL's test-signal clock.DEF 12 They usually start their test signal at 12 o'clock.
0 0 1 1 0 1	De er brygget til at bevare sin chokoladenuance længe efter åbning. they are brewed to to maintain REFL's chocolate-nuance long after opening They are brewed with the aim of maintaining their chocolate tones long after opening.
0 0 1 1 1 0	Ungerne lover at trække sin cykel på fortovet. kids.DEF promise to wheel REFL's bike on sidewalk.DEF The kids promise to wheel their bike on the sidewalk.
0 0 1 1 1 0	Deltagerne spås at beholde sin førsteplads i kagekonkurrencen. contestants.DEF predict.PASS to keep REFL's first-place in cake-competition.DEF The contestants are predicted to maintain their first place in the cake competition.
0 0 1 1 1 1	Sprinklerne er gode til at sprede sin vandstråle ud over haven. sprinklers.DEF are good to to spread REFL's water-beam out over garden.DEF The sprinklers are good at spreading their water beam across the garden.
0 0 1 1 1 1	Kaffebønnerne kommer til at miste sin naturlige aroma hurtigere ved formaling. coffee-beans.DEF come to to lose REFL's natural aroma quicker by grinding The coffee beans lose their natural aroma quicker when ground.

Table C.14: Main study: Primary test items (Questionnaire 2)

Code	Sentence
0 0 0 0 0	De sender sine forslag til bestyrelsen. they send REFL's suggestions to board.DEF They send their suggestions to the board.
0 0 0 0 0	De får sine adgangsplysninger med posten. they get REFL's login-information.PL with mail.DEF They get their login information (plural) in the mail.
0 0 0 0 1	De åbner sine døre. they open REFL's doors They open their doors.
0 0 0 0 1	De kaster sine farlige UV-stråler på planetens få beboere. they throw REFL's dangerous UV-beams on planet.DEF's few citizens They send their dangerous UV beams onto the few inhabitants of the planet.
0 0 0 1 0	Ungerne trækker sine cykler på fortovet. kids.DEF wheel REFL's bikes on sidewalk.DEF The kids wheel their bikes on the sidewalk.
0 0 0 1 0	Fædrene følger sine børn i skole. fathers.DEF follow REFL's children in school The fathers take their children to school.
0 0 0 1 1	Mange engelske ord har sine rødder i det danske sprog. many English words have REFL's roots in the Danish language Many English words have their origins in the Danish language.
0 0 0 1 1	Kaffebønnerne mister sine naturlige aromaer hurtigere ved formaling. coffee-beans.DEF lose REFL's natural aromas quicker by grinding The coffee beans lose their natural aromas quicker when ground.
0 0 0 1 0 0	De prioriterer at folde sine skjorter pænt sammen. they prioritize to fold REFL's shirts nicely together They prioritize folding their shirts nicely.
0 0 0 1 0 0	De kæmper for at bevare sine tænder i længere tid. they fight for to maintain REFL's teeth in longer time They fight to maintain their teeth longer.
0 0 0 1 0 1	De beskyldes for at ligne sine forgængere for meget. they accuse.PASS for to look REFL's predecessors too much They are accused of looking too much like their predecessors.
0 0 0 1 0 1	De plejer at klare sine 0-100 km/t på 4 sekunder. they use to manage REFL's 0-100 km/hour on 4 seconds They usually manage their 0-100 kilometers an hour in 4 seconds.
0 0 0 1 1 0	Nørderne skal til at kaste sine terninger tre gange. geeks.DEF shall to to throw REFL's dice three times The geeks are about to throw their dice three times.
0 0 0 1 1 0	Håndværkerne ventes at holde sine løfter. tradespersons.DEF expect.PASS to hold REFL's promises The tradespeople are expected to keep their promises.
0 0 0 1 1 1	Sprinklerne er gode til at sprede sine vandstråler ud over haven. sprinklers.DEF are good to to spread REFL's water-beams out over garden.DEF The sprinklers are good at spreading their water beams across the garden.
0 0 0 1 1 1	Træerne får lys ved at strække sine grene ud over åen. trees.DEF get light by to stretch REFL's branches out over stream.DEF The trees get light by stretching their branches over the stream.

Table C.15: Main study: Primary test items (Questionnaire 2) (contd.)

Code	Sentence
0 0 1 0 0 0	De beholder sin førsteplads i kagekonkurrencen. they keep REFL's first-place in cake-competition.DEF They maintain their first place in the cake competition.
0 0 1 0 0 0	De fortsætter sin protest mod den nye præsident. they continue REFL's protest against the new president They continue their protest against the new president
0 0 1 0 0 1	De udvikler sin fine brus gennem gæring. they develop REFL's fine fizz through fermentation They develop their nice fizz through fermentation.
0 0 1 0 0 1	De forbedrer sin markedsandel næste år. they improve REFL's market-share next year They will improve their market share next year.
0 0 1 0 1 0	Klovnene har sin egen fest efter forestillingen. clowns.DEF have REFL's own party after show.DEF The clowns have their own party after the show.
0 0 1 0 1 0	Vidnerne fastholder sin forklaring. witness.DEF maintain REFL's explanation The witnesses maintain their explanation.
0 0 1 0 1 1	Programmerne beskytter sin bruger mod skadelig software. programmes.DEF protect REFL's user against harmful software The programmes protect their user against harmful software.
0 0 1 0 1 1	Skibene ændrer sin lyd i regnvejr. ships.DEF change REFL's sound in rain-weather The ships change their sound in rainy weather.
0 0 1 1 0 0	De prøver at finde sit nye job på nettet. they try to find REFL's new job on web.DEF They try to find their new job on the internet.
0 0 1 1 0 0	De frygter at miste sit syn i mørket. they fear to lose REFL's sight in darkness.DEF They fear losing their sight in the darkness.
0 0 1 1 0 1	De formår at skjule sit sukkerindhold. they manage to hide REFL's sugar-content They manage to hide their sugar contents.
0 0 1 1 0 1	De er i stand til automatisk at lægge sit resultat ud på nettet. they are in able to automatically to put REFL's result out on web.DEF They are capable of putting their result on the internet automatically.
0 0 1 1 1 0	Tvillingerne frygter at tabe sin husnøgle i bussen. twins.DEF fear to lose REFL's house-key in bus.DEF The twins fear to lose their house key in the bus.
0 0 1 1 1 0	Lærerne plejer at starte sin time med sang. teachers.DEF use to start REFL's lesson with song The teachers usually start their lesson with song.
0 0 1 1 1 1	Dronerne forhindres i at gennemføre sin mission om natten. drones.DEF prevent.PASS in to complete REFL's mission at night.DEF The drones are prevented from completing their mission at night.
0 0 1 1 1 1	Digitene ender med at få sin rytme fra Shakespeare. poems.DEF end with to get REFL's rhythm from Shakespeare The poems finally get their rhythm from Shakespeare.

Table C.16: Main study: Primary test items (Questionnaire 3)

Code	Sentence
0 0 0 0 0	De har sine røde næser med i lommen. they have REFL's red noses with in pocket.DEF They bring their red noses with them in their pocket.
0 0 0 0 0	De udvikler sine hovedpersoner sammen. they develop REFL's main-characters together They develop their main characters together.
0 0 0 0 1	De forbedrer sine markedsandele næste år. they improve REFL's market-shares next year They will improve their market shares next year.
0 0 0 0 1	De kaster sine farlige UV-stråler på planetens få beboere. they throw REFL's dangerous UV-beams on planet.DEF's few citizens They send their dangerous UV beams onto the few inhabitants of the planet.
0 0 0 1 0	Vælgerne fortsætter sine protester mod den nye præsident. voters.DEF continue REFL's protests against the new president The voters continue their protests against the new president
0 0 0 1 0	Nørderne kaster sine terninger tre gange. geeks.DEF throw REFL's dice three times The geeks throw their dice three times.
0 0 0 1 1	Atomerne beholder sine elektroner i inderste ring. atoms.DEF keep REFL's electron in inner-most ring The atoms keep their electrons in the inner-most ring.
0 0 0 1 1	Maskinerne trækker sine tandhjul rundt. machines.DEF pull REFL's gears around The machines pull their gears around.
0 0 0 1 0 0	De plejer at ændre sine sange i regnvejr. they use to change REFL's songs in rain-weather They usually change their songs in rainy weather.
0 0 0 1 0 0	De prioriterer at folde sine skjorter pænt sammen. they prioritize to fold REFL's shirts nicely together They prioritize folding their shirts nicely.
0 0 0 1 0 1	De plejer at få sine affaldsstoffer fra forurennet overfladevand. they use to get REFL's waste-products from contaminated surface-water They usually get their waste products from contaminated surface water.
0 0 0 1 0 1	De ender med at få sine rytmer fra Shakespeare. they end with to get REFL's rhythms from Shakespeare They finally get their rhythms from Shakespeare.
0 0 0 1 1 0	Klovnene regner med at have sine egne fester efter forestillingen. clowns.DEF plan with to have REFL's own parties after show.DEF The clowns plan on having their own parties after the show.
0 0 0 1 1 0	Betjentene vælger at lægge sine våben tilbage i bilen. officers.DEF choose to put REFL's weapons back in car.DEF The officers choose to put their weapons back in the car.
0 0 0 1 1 1	Dronerne forhindres i at gennemføre sine missioner om natten. drones.DEF prevent.PASS in to complete REFL's missions at night.DEF The drones are prevented from completing their missions at night.
0 0 0 1 1 1	Sodavandene formår at skjule sine mange gram sukker. sodas.DEF manage to hide REFL's many grams sugar The sodas manage to hide their many grams of sugar.

Table C.17: Main study: Primary test items (Questionnaire 3) (contd.)

Code	Sentence
0 0 1 0 0 0	De bevarer sit tandsæt i længere tid. they maintain REFL's set-of-teeth in longer time They maintain their set of teeth longer.
0 0 1 0 0 0	De finder sit nye job på nettet. they find REFL's new job on web.DEF They find their new job on the internet.
0 0 1 0 0 1	De beskytter sin bruger mod skadelig software. they protect REFL's user against harmful software They protect their user against harmful software.
0 0 1 0 0 1	De ligner sin forgænger ret meget. they look-like REFL's predecessor quite much They look quite a lot like their predecessor.
0 0 1 0 1 0	Fiskerne spreder sit net ud i havet. fishermen.DEF spread REFL's net out in ocean.DEF
0 0 1 0 1 0	Håndværkerne holder sit løfte. tradespersons.DEF hold REFL's promise The tradespeople keep their promise.
0 0 1 0 1 1	Filmene følger sine hovedpersoner gennem flere år. movies.DEF follow REFL's main-characters through several years The movies follow their main characters through several years.
0 0 1 0 1 1	Kaffebønnerne mister sin naturlige aroma hurtigere ved formaling. coffee-beans.DEF lose REFL's natural aroma quicker by grinding The coffee beans lose their natural aroma quicker when ground.
0 0 1 1 0 0	De frygter at miste sit syn i mørket. they fear to lose REFL's sight in darkness.DEF They fear losing their sight in the darkness.
0 0 1 1 0 0	De venter med at sende sit forslag til bestyrelsen. they wait with to send REFL's suggestion to board.DEF They wait with sending their suggestion to the board.
0 0 1 1 0 1	De plejer at tabe sit løv i stormvejret. they use to lose REFL's foliage in storm-weather.DEF They usually lose their foliage in the stormy weather.
0 0 1 1 0 1	De plejer at starte sit prøvesignal klokken 12. they use to start REFL's test-signal clock.DEF 12 They usually start their test signal at 12 o'clock.
0 0 1 1 1 0	Tryllekunstnerne begynder at åbne sin magiske æske. magicians.DEF begin to open REFL's magical box The magicians start opening their magical box.
0 0 1 1 1 0	Forældrene forsøger at klare sin skilsmisse uden skænderier. parents.DEF try to manage REFL's divorce without quarrels The parents try to manage their divorce without quarrels.
0 0 1 1 1 1	Træerne får lys ved at strække sin krone ud over åen. trees.DEF get light by to stretch REFL's crown out over stream.DEF The trees get light by stretching their crown over the stream.
0 0 1 1 1 1	Tegneserierne plejer at fastholde sin læser. comics.DEF use to capture REFL's reader The comics usually capture their reader.

Table C.18: Main study: Primary test items (Questionnaire 4)

Code	Sentence
0 0 0 0 0	De gennemfører sine indbrud om natten. they perform REFL's burglaries at night.DEF They perform their burglaries at night.
0 0 0 0 0	De trækker sine cykler på fortovet. they wheel REFL's bikes on sidewalk.DEF They wheel their bikes on the sidewalk.
0 0 0 0 1	De beskytter sine brugere mod skadelig software. they protect REFL's users against harmful software They protect their users against harmful software.
0 0 0 0 1	De følger sine hovedpersoner gennem flere år. they follow REFL's main-characters through several years They follow their main characters through several years.
0 0 0 1 0	Nørderne kaster sine terninger tre gange. geeks.DEF throw REFL's dice three times The geeks throw their dice three times.
0 0 0 1 0	Børnene får sine mobiltelefoner i frikvarteret. kids.DEF get REFL's cellphones in break.DEF The kids get their cellphones in the break.
0 0 0 1 1	Vandløbene får sine affaldsstoffer fra forurenede overfladevand. streams.DEF get REFL's waste-products from contaminated surface-water The streams get their waste products from contaminated surface water.
0 0 0 1 1	Pendulerne fortsætter sine svingninger i et vacuum. pendulums.DEF continue REFL's oscillations in a vacuum The pendulums continue their oscillations in a vacuum.
0 0 0 1 0 0	De løber ud for at strække sine vinger i solen. they run out for to stretch REFL's wings in sun.DEF They run out to stretch their wings in the sun.
0 0 0 1 0 0	De prøver at finde sine nye jobs på nettet. they try to find REFL's new jobs on web.DEF They try to find their new jobs on the internet.
0 0 0 1 0 1	De holder på duftstoffer ved at folde sine blade sammen om natten. they hold on scents by to fold REFL's leaves together at night.DEF They hold on to their scents by folding their leaves at night.
0 0 0 1 0 1	De er i stand til automatisk at lægge sine resultater ud på nettet. they are in able to automatically to put REFL's results out on web.DEF They are capable of putting their results on the internet automatically.
0 0 0 1 1 0	Håndværkerne ventes at holde sine løfter. tradespersons.DEF expect.PASS to hold REFL's promises The tradespeople are expected to keep their promises.
0 0 0 1 1 0	Pigerne ventes at forbedre sine læsevner hvert år. girls.DEF expect.PASS to improve REFL's reading-skills every year The girls are expected to improve their reading skills every year.
0 0 0 1 1 1	Sprinklerne er gode til at sprede sine vandstråler ud over haven. sprinklers.DEF are good to to spread REFL's water-beams out over garden.DEF The sprinklers are good at spreading their water beams across the garden.
0 0 0 1 1 1	Sodavandene formår at skjule sine mange gram sukker. sodas.DEF manage to hide REFL's many grams sugar The sodas manage to hide their many grams of sugar.

Table C.19: Main study: Primary test items (Questionnaire 4) (contd.)

Code	Sentence
0 0 1 0 0 0	De fastholder sin forklaring. they maintain REFL's explanation They maintain their explanation.
0 0 1 0 0 0	De ligner sin far på en prik. they look-like REFL's father on a spot They look exactly like their father.
0 0 1 0 0 1	De bevarer sin chokoladenuance længe efter åbning. they maintain REFL's chocolate-nuance long after opening They maintain their chocolate tones long after opening.
0 0 1 0 0 1	De har sin rod i det danske sprog. they have REFL's root in the Danish language They have their origin in the Danish language.
0 0 1 0 1 0	Medlemmerne sender sit forslag til bestyrelsen. members.DEF send REFL's suggestion to board.DEF The members send their suggestion to the board.
0 0 1 0 1 0	Klovnene har sin egen fest efter forestillingen. clowns.DEF have REFL's own party after show.DEF The clowns have their own party after the show.
0 0 1 0 1 1	Kaffebønnerne mister sin naturlige aroma hurtigere ved formaling. coffee-beans.DEF lose REFL's natural aroma quicker by grinding The coffee beans lose their natural aroma quicker when ground.
0 0 1 0 1 1	Skibene ændrer sin lyd i regnvejr. ships.DEF change REFL's sound in rain-weather The ships change their sound in rainy weather.
0 0 1 1 0 0	De aftaler at udvikle sin hovedperson sammen. they agree to develop REFL's main-character together The authors agree to develop their main character together.
0 0 1 1 0 0	De forsøger at klare sin skilsmisse uden skænderier. they try to manage REFL's divorce without quarrels They try to manage their divorce without quarrels.
0 0 1 1 0 1	De plejer at starte sit prøvesignal klokken 12. they use to start REFL's test-signal clock.DEF 12 They usually start their test signal at 12 o'clock.
0 0 1 1 0 1	De plejer at tabe sit løv i stormvejret. they use to lose REFL's foliage in storm-weather.DEF They usually lose their foliage in the stormy weather.
0 0 1 1 1 0	Deltagerne spås at beholde sin førsteplads i kagekonkurrencen. contestants.DEF predict.PASS to keep REFL's first-place in cake-competition.DEF The contestants are predicted to maintain their first place in the cake competition.
0 0 1 1 1 0	Nørderne vælger at kaste sin opsparing efter obskure brætspil. geeks.DEF choose to throw REFL's savings.SG after obscure board-games The geeks choose to throw their savings (singular) at obscure board games.
0 0 1 1 1 1	Bilerne ventes at miste sin farve over tid. cars.CARS expect.PASS to lose REFL's colour over time The cars are expected to lose their colour over time.
0 0 1 1 1 1	Elevatorene begynder at åbne sin dør. lifts.DEF begin to open REFL's door The lifts start opening their door.

C.2.1 All 32 primary test items

The tables in this section contain all the 32 primary test items with 16 unique sentences each. The factors that determine the structure of each test item are outlined in table C.20. The binaries in the right-most column is the code that shows the factor level for a given condition. A code of 0000 means a sentence with a plural DP that contains *sin* (0XXX), a simple sentence (00XX), and a pronoun subject (000X) that is also animate (0000). I gloss the 8 first examples fully in each condition. The 8 last examples are the same as the 8 first for most of the examples, except that the DP that contains *sin* is singular. In the cases where the plural DP did not have an obvious singular counterpart, the possessed singular DP is glossed, too.

Table C.20: The four factors that combine to form the 16 different conditions in an experimental item.

Factor	Levels (binary)	
# of DP that contains <i>sin</i>	Plural:	0XXX
	Singular:	1XXX
Complexity of sentence	Simplex:	X0XX
	Complex:	X1XX
Type of subject	Pronoun:	XX0X
	Full noun:	XX1X
Animacy	Animate:	XXX0
	Inanimate:	XXX1

Table C.21: Primary item 1

Code	Sentence
0 0 0 0 0 0	De mister sine sanser i mørket. they lose REFL's senses in darkness.DEF
0 0 0 0 0 1	De mister sine farver over tid. they lose REFL's colours over time
0 0 0 0 1 0	Drengene mister sine sanser i mørket. boys.DEF lose REFL's senses in darkness.DEF
0 0 0 0 1 1	Bilerne mister sine farver over tid. cars.DEF lose REFL's colours over time
0 0 0 1 0 0	De frygter at miste sine sanser i mørket. they fear to lose REFL's senses in darkness.DEF
0 0 0 1 0 1	De ventes at miste sine farver over tid. they expect.PASS to lose REFL's colours over time
0 0 0 1 1 0	Drengene frygter at miste sine sanser i mørket. boys.DEF fear to lose REFL's senses in darkness.DEF
0 0 0 1 1 1	Bilerne ventes at miste sine farver over tid. cars.DEF expect.PASS to lose REFL's colours over time
0 0 1 0 0 0	De mister sit syn i mørket. sit syn = REFL's sight
0 0 1 0 0 1	De mister sin farve over tid.
0 0 1 0 1 0	Drengene mister sit syn i mørket.
0 0 1 0 1 1	Bilerne mister sin farve over tid.
0 0 1 1 0 0	De frygter at miste sit syn i mørket.
0 0 1 1 0 1	De ventes at miste sin farve over tid.
0 0 1 1 1 0	Drengene frygter at miste sit syn i mørket.
0 0 1 1 1 1	Bilerne ventes at miste sin farve over tid.

Table C.22: Primary item 2

Code	Sentence
0 0 0 0 0 0	De får sine mobiltelefoner i frikvarteret. they get REFL's cellphones in break.DEF
0 0 0 0 0 1	De får sine rytmer fra Shakespeare. they get REFL's rhythms from Shakespeare
0 0 0 0 1 0	Børnene får sine mobiltelefoner i frikvarteret. kids.DEF get REFL's cellphones in break.DEF
0 0 0 0 1 1	Digtene får sine rytmer fra Shakespeare. poems.DEF get REFL's rhythms from Shakespeare
0 0 0 1 0 0	De forventer at få sine mobiltelefoner i frikvarteret. they expect to get REFL's cellphones in break.DEF
0 0 0 1 0 1	De ender med at få sine rytmer fra Shakespeare. they end with to get REFL's rhythms from Shakespeare
0 0 0 1 1 0	Børnene forventer at få sine mobiltelefoner i frikvarteret. kids.DEF expect to get REFL's cellphones in break.DEF
0 0 0 1 1 1	Digtene ender med at få sine rytmer fra Shakespeare. poems.DEF end with to get REFL's rhythms from Shakespeare
0 0 1 0 0 0	De får sin mobiltelefon i frikvarteret.
0 0 1 0 0 1	De får sin rytme fra Shakespeare.
0 0 1 0 1 0	Børnene får sin mobiltelefon i frikvarteret.
0 0 1 0 1 1	Digtene får sin rytme fra Shakespeare.
0 0 1 1 0 0	De forventer at få sin mobiltelefon i frikvarteret.
0 0 1 1 0 1	De ender med at få sin rytme fra Shakespeare.
0 0 1 1 1 0	Børnene forventer at få sin mobiltelefon i frikvarteret.
0 0 1 1 1 1	Digtene ender med at få sin rytme fra Shakespeare.

Table C.23: Primary item 3

Code	Sentence
0 0 0 0 0	De får sine adgangsoplysninger med posten. they get REFL's login-information.PL with mailDEF
0 0 0 0 1	De får sine affaldsstoffer fra forurenede overfladevand. they get REFL's waste-products from contaminated surface-water
0 0 0 1 0	Brugerne får sine adgangsoplysninger med posten. users.DEF get REFL's login-information.PL with mailDEF
0 0 0 1 1	Vandløbene får sine affaldsstoffer fra forurenede overfladevand. streams.DEF get REFL's waste-products from contaminated surface-water
0 0 1 0 0	De venter på at få sine adgangsoplysninger med posten. they wait on to get REFL's login-information.PL with mailDEF
0 0 1 0 1	De plejer at få sine affaldsstoffer fra forurenede overfladevand. they use to get REFL's waste-products from contaminated surface-water
0 0 1 1 0	Brugerne venter på at få sine adgangsoplysninger med posten. users.DEF wait on to get REFL's login-information.PL with mailDEF
0 0 1 1 1	Vandløbene plejer at få sine affaldsstoffer fra forurenede overfladevand. streams.DEF use to get REFL's waste-products from contaminated surface-water
0 0 1 0 0	De får sit login med posten.
0 0 1 0 1	De får sin grumsede farve fra forurenede overfladevand. sin grumsede farve = REFL's muddy colour
0 0 1 0 1	Brugerne får sit login med posten. sit login = REFL's login
0 0 1 0 1	Vandløbene får sin grumsede farve fra forurenede overfladevand.
0 0 1 1 0	De venter på at få sit login med posten.
0 0 1 1 0	De plejer at få sin grumsede farve fra forurenede overfladevand.
0 0 1 1 0	Brugerne venter på at få sit login med posten.
0 0 1 1 1	Vandløbene plejer at få sin grumsede farve fra forurenede overfladevand.

Table C.24: Primary item 4

Code	Sentence
0 0 0 0 0 0	De beholder sine førstepladser i kagekonkurrencen. they maintain REFL's first-places in cake-competition.DEF
0 0 0 0 0 1	De beholder sine elektroner i inderste ring. they keep REFL's electrons in inner-most ring
0 0 0 0 1 0	Deltagerne beholder sine førstepladser i kagekonkurrencen. contestants.DEF maintain REFL's first-places in cake-competition.DEF
0 0 0 0 1 1	Atomerne beholder sine elektroner i inderste ring. atoms.DEF keep REFL's electrons in inner-most ring
0 0 0 1 0 0	De spås at beholde sine førstepladser i kagekonkurrencen. they predict.PASS to maintain REFL's first-places in cake-competition.DEF
0 0 0 1 0 1	De påvirkes til at beholde sine elektroner i inderste ring. atoms.DEF affect.PASS to keep REFL's electrons in inner-most ring
0 0 0 1 1 0	Deltagerne spås at beholde sine førstepladser i kagekonkurrencen. contestants.DEF predict.PASS to maintain REFL's first-places in cake-competition.DEF
0 0 0 1 1 1	Atomerne påvirkes til at beholde sine elektroner i inderste ring. atoms.DEF affect.PASS to keep REFL's electrons in inner-most ring
0 0 1 0 0 0	De beholder sin førsteplads i kagekonkurrencen.
0 0 1 0 0 1	De beholder sin ene elektron i inderste ring.
0 0 1 0 1 0	Deltagerne beholder sin førsteplads i kagekonkurrencen.
0 0 1 0 1 1	Atomerne beholder sin ene elektron i inderste ring.
0 0 1 1 0 0	De spås at beholde sin førsteplads i kagekonkurrencen.
0 0 1 1 0 1	De påvirkes til at beholde sin ene elektron i inderste ring.
0 0 1 1 1 0	Deltagerne spås at beholde sin førsteplads i kagekonkurrencen.
0 0 1 1 1 1	Atomerne påvirkes til at beholde sin ene elektron i inderste ring.

Table C.25: Primary item 5

Code	Sentence
0 0 0 0 0	De mister sine to hovedsponsorer. they lose REFL's two main-sponsors
0 0 0 0 1	De mister sine naturlige aromaer hurtigere ved formaling. they lose REFL's natural aromas quicker by grinding
0 0 0 1 0	Håndboldkvinderne mister sine to hovedsponsorer. handball-ladies.DEF lose REFL's two main-sponsors
0 0 0 1 1	Kaffebønnerne mister sine naturlige aromaer hurtigere ved formaling. coffee-beans.DEF lose REFL's natural aromas quicker by grinding
0 0 0 1 0 0	De forventer at miste sine to hovedsponsorer. they expect to lose REFL's two main-sponsors
0 0 0 1 0 1	De kommer til at miste sine naturlige aromaer hurtigere ved formaling. they come to to lose REFL's natural aromas quicker by grinding
0 0 0 1 1 0	Håndboldkvinderne forventer at miste sine to hovedsponsorer. handball-ladies.DEF expect to lose REFL's two main-sponsors
0 0 0 1 1 1	Kaffebønnerne kommer til at miste sine naturlige aromaer hurtigere ved formaling. coffee-beans.DEF come to to lose REFL's natural aromas quicker by grinding
0 0 1 0 0 0	De mister sin hovedsponsor.
0 0 1 0 0 1	De mister sin naturlige aroma hurtigere ved formaling.
0 0 1 0 1 0	Håndboldkvinderne mister sin hovedsponsor.
0 0 1 0 1 1	Kaffebønnerne mister sin naturlige aroma hurtigere ved formaling.
0 0 1 1 0 0	De forventer at miste sin hovedsponsor.
0 0 1 1 0 1	De kommer til at miste sin naturlige aroma hurtigere ved formaling.
0 0 1 1 1 0	Håndboldkvinderne forventer at miste sin hovedsponsor.
0 0 1 1 1 1	Kaffebønnerne kommer til at miste sin naturlige aroma hurtigere ved formaling.

Table C.26: Primary item 6

Code	Sentence
0 0 0 0 0 0	De bevarer sine tænder i længere tid. they maintain REFL's teeth in longer time
0 0 0 0 0 1	De bevarer sine chokoladenuancer længe efter åbning. they maintain REFL's chocolate-nuances long after opening
0 0 0 0 1 0	Patienterne bevarer sine tænder i længere tid. patients.DEF maintain REFL's teeth in longer time
0 0 0 0 1 1	Øllene bevarer sine chokoladenuancer længe efter åbning. beers.DEF maintain REFL's chocolate-nuances long after opening
0 0 0 1 0 0	De kæmper for at bevare sine tænder i længere tid. they fight for to maintain REFL's teeth in longer time
0 0 0 1 0 1	De er brygget til at bevare sine chokoladenuancer længe efter åbning. they are brewed to to maintain REFL's chocolate-nuances long after opening
0 0 0 1 1 0	Patienterne kæmper for at bevare sine tænder i længere tid. patients.DEF fight for to maintain REFL's teeth in longer time
0 0 0 1 1 1	Øllene er brygget til at bevare sine chokoladenuancer længe efter åbning. beers.DEF are brewed to to maintain REFL's chocolate-nuances long after opening
0 0 1 0 0 0	De bevarer sit tandsæt i længere tid. sit tandsæt = REFL's set of teeth
0 0 1 0 0 1	De bevarer sin chokoladenuance længe efter åbning.
0 0 1 0 1 0	Patienterne bevarer sit tandsæt i længere tid.
0 0 1 0 1 1	Øllene bevarer sin chokoladenuance længe efter åbning.
0 0 1 1 0 0	De kæmper for at bevare sit tandsæt i længere tid.
0 0 1 1 0 1	De er brygget til at bevare sin chokoladenuance længe efter åbning.
0 0 1 1 1 0	Patienterne kæmper for at bevare sit tandsæt i længere tid.
0 0 1 1 1 1	Øllene er brygget til at bevare sin chokoladenuance længe efter åbning.

Table C.27: Primary item 7

Code	Sentence
0 0 0 0 0	De strækker sine vinger i solen. they stretch REFL's wings in sun.DEF
0 0 0 0 1	De strækker sine grene ud over åen. they stretch REFL's branches out over stream.DEF
0 0 0 1 0	Hønsene strækker sine vinger i solen. hens.DEF stretch REFL's wings in sun.DEF
0 0 0 1 1	Træerne strækker sine grene ud over åen. trees.DEF stretch REFL's branches out over stream.DEF
0 0 1 0 0	De løber ud for at strække sine vinger i solen. they run out for to stretch REFL's wings in sun.DEF
0 0 1 0 1	De får lys ved at strække sine grene ud over åen. they get light by to stretch REFL's branches out over stream.DEF
0 0 1 1 0	Hønsene løber ud for at strække sine vinger i solen. hens.DEF run out for to stretch REFL's wings in sun.DEF
0 0 1 1 1	Træerne får lys ved at strække sine grene ud over åen. trees.DEF get light by to stretch REFL's branches out over stream.DEF
0 0 1 0 0	De strækker sit næb ud efter maden.
0 0 1 0 1	De strækker sin krone ud over åen. sin krone = REFL's crown
0 0 1 0 1	Hønsene strækker sit næb ud efter maden. sit næb ud efter maden = REFL's beak out after food.DEF
0 0 1 0 1	Træerne strækker sin krone ud over åen.
0 0 1 1 0	De løber hen for at strække sit næb ud efter maden.
0 0 1 1 0	De får lys ved at strække sin krone ud over åen.
0 0 1 1 1	Hønsene løber hen for at strække sit næb ud efter maden.
0 0 1 1 1	Træerne får lys ved at strække sin krone ud over åen.

Table C.28: Primary item 8

Code	Sentence
0 0 0 0 0 0	De har sine røde næser med i lommen. they have REFL's red noses with in pocket.DEF
0 0 0 0 0 1	De har sine rødder i det danske sprog. they have REFL's roots in the Danish language
0 0 0 0 1 0	Klovnene har sine røde næser med i lommen. clowns.DEF have REFL's red noses with in pocket.DEF
0 0 0 0 1 1	Mange engelske ord har sine rødder i det danske sprog. many English words have REFL's roots in the Danish language
0 0 0 1 0 0	De overrasker ved at have sine røde næser med i lommen. they surprise by to have REFL's red noses with in pocket.DEF
0 0 0 1 0 1	De siges at have sine rødder i det danske sprog. they say.PASS to have REFL's roots in the Danish language
0 0 0 1 1 0	Klovnene overrasker ved at have sine røde næser med i lommen. clowns.DEF surprise by to have REFL's red noses with in pocket.DEF
0 0 0 1 1 1	Mange engelske ord siges at have sine rødder i det danske sprog. many English words say.PASS to have REFL's roots in the Danish language
0 0 1 0 0 0	De har sin røde næse med i lommen.
0 0 1 0 0 1	De har sin rod i det danske sprog.
0 0 1 0 1 0	Klovnene har sin røde næse med i lommen.
0 0 1 0 1 1	Mange engelske ord har sin rod i det danske sprog.
0 0 1 1 0 0	De overrasker ved at have sin røde næse med i lommen.
0 0 1 1 0 1	De siges at have sin rod i det danske sprog.
0 0 1 1 1 0	Klovnene overrasker ved at have sin røde næse med i lommen.
0 0 1 1 1 1	Mange engelske ord siges at have sin rod i det danske sprog.

Table C.29: Primary item 9

Code	Sentence
0 0 0 0 0 0	De folder sine skjorter pænt sammen. boys.DEF fold REFL's shirts nicely together
0 0 0 0 0 1	De folder sine blade sammen om natten. they fold REFL's leaves together at night.DEF
0 0 0 0 1 0	Drengene folder sine skjorter pænt sammen. boys.DEF fold REFL's shirts nicely together
0 0 0 0 1 1	Blomsterne folder sine blade sammen om natten. flowers.DEF fold REFL's leaves together at night.DEF
0 0 0 1 0 0	De prioriterer at folde sine skjorter pænt sammen. they prioritize to fold REFL's shirts nicely together
0 0 0 1 0 1	De holder på duftstoffer ved at folde sine blade sammen om natten. they hold on scents by to fold REFL's leaves together at night.DEF
0 0 0 1 1 0	Drengene prioriterer at folde sine skjorter pænt sammen. boys.DEF prioritize to fold REFL's shirts nicely together
0 0 0 1 1 1	Blomsterne holder på duftstoffer ved at folde sine blade sammen om natten. flowers.DEF hold on scents by to fold REFL's leaves together at night.DEF
0 0 1 0 0 0	De folder sit tøj pænt sammen. sit tøj = REFL's clothes.SG
0 0 1 0 0 1	De folder sit kronblad sammen om natten. sit kronblad = REFL's petal
0 0 1 0 1 0	Drengene folder sit tøj pænt sammen.
0 0 1 0 1 1	Blomsterne folder sit kronblad sammen om natten.
0 0 1 1 0 0	De prioriterer at folde sit tøj pænt sammen.
0 0 1 1 0 1	De holder på duftstoffer ved at folde sit kronblad sammen om natten.
0 0 1 1 1 0	Drengene prioriterer at folde sit tøj pænt sammen.
0 0 1 1 1 1	Blomsterne holder på duftstoffer ved at folde sit kronblad sammen om natten.

Table C.30: Primary item 10

Code	Sentence
0 0 0 0 0 0	De ligner sine forældre på en prik. they look-like REFL's parents on a spot
0 0 0 0 0 1	De ligner sine forgængere ret meget. they look-like REFL's predecessors fairly much
0 0 0 0 1 0	Babyerne ligner sine forældre på en prik. babies.DEF look-like REFL's parents on a spot
0 0 0 0 1 1	Telefonerne ligner sine forgængere ret meget. phones.DEF look-like REFL's predecessors fairly much
0 0 0 1 0 0	De begynder at ligne sine forældre på en prik. they begin to look-like REFL's parents on a spot
0 0 0 1 0 1	De beskyldes for at ligne sine forgængere for meget. they accuse.PASS for to look-like REFL's predecessors too much
0 0 0 1 1 0	Babyerne begynder at ligne sine forældre på en prik. babies.DEF begin to look-like REFL's parents on a spot
0 0 0 1 1 1	Telefonerne beskyldes for at ligne sine forgængere for meget. phones.DEF accuse.PASS for to look-like REFL's predecessors too much
0 0 1 0 0 0	De ligner sin far på en prik. sin far = REFL's father
0 0 1 0 0 1	De ligner sin forgænger ret meget.
0 0 1 0 1 0	Babyerne ligner sin far på en prik.
0 0 1 0 1 1	Telefonerne ligner sin forgænger ret meget.
0 0 1 1 0 0	De begynder at ligne sin far på en prik.
0 0 1 1 0 1	De beskyldes for at ligne sin forgænger for meget.
0 0 1 1 1 0	Babyerne begynder at ligne sin far på en prik.
0 0 1 1 1 1	Telefonerne beskyldes for at ligne sin forgænger for meget.

Table C.31: Primary item 11

Code	Sentence
0 0 0 0 0 0	De lægger sine våben tilbage i bilen. they put REFL's weapons back in car.DEF
0 0 0 0 0 1	De lægger automatisk sine resultater ud på nettet. they automatically put REFL's results out on web.DEF
0 0 0 0 1 0	Betjentene lægger sine våben tilbage i bilen. officers.DEF put REFL's weapons back in car.DEF
0 0 0 0 1 1	Robotterne lægger automatisk sine resultater ud på nettet. robots.DEF automatically put REFL's results out on web.DEF
0 0 0 1 0 0	De vælger at lægge sine våben tilbage i bilen. they choose to put REFL's weapons back in car.DEF
0 0 0 1 0 1	De er i stand til automatisk at lægge sine resultater ud på nettet. they are in able to automatically to put REFL's results out on web.DEF
0 0 0 1 1 0	Betjentene vælger at lægge sine våben tilbage i bilen. officers.DEF choose to put REFL's weapons back in car.DEF
0 0 0 1 1 1	Robotterne er i stand til automatisk at lægge sine resultater ud på nettet. robots.DEF are in able to automatically to put REFL's results out on web.DEF
0 0 1 0 0 0	De lægger sit våben tilbage i bilen.
0 0 1 0 0 1	De lægger automatisk sit resultat ud på nettet.
0 0 1 0 1 0	Betjentene lægger sit våben tilbage i bilen.
0 0 1 0 1 1	Robotterne lægger automatisk sit resultat ud på nettet.
0 0 1 1 0 0	De vælger at lægge sit våben tilbage i bilen.
0 0 1 1 0 1	De er i stand til automatisk at lægge sit resultat ud på nettet.
0 0 1 1 1 0	Betjentene vælger at lægge sine våben tilbage i bilen.
0 0 1 1 1 1	Robotterne er i stand til automatisk at lægge sit resultat ud på nettet.

Table C.32: Primary item 12

Code	Sentence
0 0 0 0 0 0	De sender sine forslag til bestyrelsen. they send REFL's suggestions to board.DEF
0 0 0 0 0 1	De sender sine regndråber mod jorden. they send REFL's rain-drops towards earth.DEF
0 0 0 0 1 0	Medlemmerne sender sine forslag til bestyrelsen. members.DEF send REFL's suggestions to board.DEF
0 0 0 0 1 1	Skyerne sender sine regndråber mod jorden clouds.DEF send REFL's rain-drops towards earth.DEF
0 0 0 1 0 0	De venter med at sende sine forslag til bestyrelsen. they wait with to send REFL's suggestions to board.DEF
0 0 0 1 0 1	De venter med at sende sine regndråber mod jorden. they wait with to send REFL's rain-drops towards earth.DEF
0 0 0 1 1 0	Medlemmerne venter med at sende sine forslag til bestyrelsen. members.DEF wait with to send REFL's suggestions to board.DEF
0 0 0 1 1 1	Skyerne venter med at sende sine regndråber mod jorden. clouds.DEF wait with to send REFL's rain-drops towards earth.DEF
0 0 1 0 0 0	De sender sit forslag til bestyrelsen. sit forslag = REFL's suggestion
0 0 1 0 0 1	De sender sin regn mod jorden. sin regn = REFL's rain
0 0 1 0 1 0	Medlemmerne sender sit forslag til bestyrelsen.
0 0 1 0 1 1	Skyerne sender sin regn mod jorden.
0 0 1 1 0 0	De venter med at sende sit forslag til bestyrelsen.
0 0 1 1 0 1	De venter med at sende sin regn mod jorden.
0 0 1 1 1 0	Medlemmerne venter med at sende sit forslag til bestyrelsen.
0 0 1 1 1 1	Skyerne venter med at sende sin regn mod jorden.

Table C.33: Primary item 13

Code	Sentence
0 0 0 0 0	De trækker sine cykler på fortovet. they wheel REFL's bikes on sidewalk.DEF
0 0 0 0 1	De trækker sine tandhjul rundt. they pull REFL's gears around
0 0 0 1 0	Ungerne trækker sine cykler på fortovet. kids.DEF wheel REFL's bikes on sidewalk.DEF
0 0 0 1 1	Maskinerne trækker sine tandhjul rundt. machines.DEF pull REFL's gears around
0 0 0 1 0 0	De lover at trække sine cykler på fortovet. they promise to wheel REFL's bikes on sidewalk.DEF
0 0 0 1 0 1	De arbejder på at trække sine tandhjul rundt. they work on to pull REFL's gears around
0 0 0 1 1 0	Ungerne lover at trække sine cykler på fortovet. kids.DEF promise to wheel REFL's bikes on sidewalk.DEF
0 0 0 1 1 1	Maskinerne arbejder på at trække sine tandhjul rundt. machines.DEF work on to pull REFL's gears around
0 0 1 0 0 0	De trækker sin cykel på fortovet.
0 0 1 0 0 1	De trækker sit tandhjul rundt.
0 0 1 0 1 0	Ungerne trækker sin cykel på fortovet.
0 0 1 0 1 1	Maskinerne trækker sit tandhjul rundt.
0 0 1 1 0 0	De lover at trække sin cykel på fortovet.
0 0 1 1 0 1	De arbejder på at trække sit tandhjul rundt.
0 0 1 1 1 0	Ungerne lover at trække sin cykel på fortovet.
0 0 1 1 1 1	Maskinerne arbejder på at trække sit tandhjul rundt.

Table C.34: Primary item 14

Code	Sentence
0 0 0 0 0 0	De fortsætter sine protester mod den nye præsident. they continue REFL's protests against the new president
0 0 0 0 0 1	De fortsætter sine svingninger i et vacuum. they continue REFL's oscillations in a vacuum
0 0 0 0 1 0	Vælgerne fortsætter sine protester mod den nye præsident. voters.DEF continue REFL's protests against the new president
0 0 0 0 1 1	Pendulerne fortsætter sine svingninger i et vacuum. pendulums.DEF continue REFL's oscillations in a vacuum
0 0 0 1 0 0	De ønsker at fortsætte sine protester mod den nye præsident. they wish to continue REFL's protests against the new president
0 0 0 1 0 1	De har bedre mulighed for at fortsætte sine svingninger i et vacuum. they have better opportunity for to continue REFL's oscillations in a vacuum
0 0 0 1 1 0	Vælgerne ønsker at fortsætte sine protester mod den nye præsident. voters.DEF wish to continue REFL's protests against the new president
0 0 0 1 1 1	Pendulerne har bedre mulighed for at fortsætte sine svingninger i et vacuum. pendulums.DEF have better opportunity for to continue REFL's oscillations in a vacuum
0 0 1 0 0 0	De fortsætter sin protest mod den nye præsident.
0 0 1 0 0 1	De fortsætter sin svingning i et vacuum.
0 0 1 0 1 0	Vælgerne fortsætter sin protest mod den nye præsident.
0 0 1 0 1 1	Pendulerne fortsætter sin svingning i et vacuum.
0 0 1 1 0 0	De ønsker at fortsætte sin protest mod den nye præsident.
0 0 1 1 0 1	De har bedre mulighed for at fortsætte sin svingning i et vacuum.
0 0 1 1 1 0	Vælgerne ønsker at fortsætte sin protest mod den nye præsident.
0 0 1 1 1 1	Pendulerne har bedre mulighed for at fortsætte sin svingning i et vacuum.

Table C.35: Primary item 15

Code	Sentence
0 0 0 0 0 0	De ændrer sine sange i regnvejr. they change REFL's songs in rain-weather
0 0 0 0 0 1	De ændrer sine lyde i regnvejr. they change REFL's sounds in rain-weather
0 0 0 0 1 0	Fuglene ændrer sine sange i regnvejr. birds.DEF change REFL's songs in rain-weather
0 0 0 0 1 1	Skibene ændrer sine lyde i regnvejr. ships.DEF change REFL's sounds in rain-weather
0 0 0 1 0 0	De plejer at ændre sine sange i regnvejr. they use to change REFL's songs in rain-weather
0 0 0 1 0 1	De plejer at ændre sine lyde i regnvejr. they use to change REFL's sounds in rain-weather
0 0 0 1 1 0	Fuglene plejer at ændre sine sange i regnvejr. birds.DEF use to change REFL's songs in rain-weather
0 0 0 1 1 1	Skibene plejer at ændre sine lyde i regnvejr. ships.DEF use to change REFL's sounds in rain-weather
0 0 1 0 0 0	De ændrer sin sang i regnvejr.
0 0 1 0 0 1	De ændrer sin lyd i regnvejr.
0 0 1 0 1 0	Fuglene ændrer sin sang i regnvejr.
0 0 1 0 1 1	Skibene ændrer sin lyd i regnvejr.
0 0 1 1 0 0	De plejer at ændre sin sang i regnvejr.
0 0 1 1 0 1	De plejer at ændre sin lyd i regnvejr.
0 0 1 1 1 0	Fuglene plejer at ændre sin sang i regnvejr.
0 0 1 1 1 1	Skibene plejer at ændre sin lyd i regnvejr.

Table C.36: Primary item 16

Code	Sentence
0 0 0 0 0 0	De har sine egne fester efter forestillingen. they have REFL's own parties after show.DEF
0 0 0 0 0 1	De har sine faste pladser på reolen. they have REFL's fixed spots on bookshelf.DEF
0 0 0 0 1 0	Klovnene har sine egne fester efter forestillingen. clowns.DEF have REFL's own parties after show.DEF
0 0 0 0 1 1	Vaserne har sine faste pladser på reolen. vases.DEF have REFL's fixed spots on bookshelf.DEF
0 0 0 1 0 0	De regner med at have sine egne fester efter forestillingen. they plan with to have REFL's own parties after show.DEF
0 0 0 1 0 1	De plejer at have sine faste pladser på reolen. they use to have REFL's fixed spots on bookshelf.DEF
0 0 0 1 1 0	Klovnene regner med at have sine egne fester efter forestillingen. clowns.DEF plan with to have REFL's own parties after show.DEF
0 0 0 1 1 1	Vaserne plejer at have sine faste pladser på reolen. vases.DEF use to have REFL's fixed spots on bookshelf.DEF
0 0 1 0 0 0	De har sin egen fest efter forestillingen.
0 0 1 0 0 1	De har sin faste plads på reolen.
0 0 1 0 1 0	Klovnene har sin egen fest efter forestillingen.
0 0 1 0 1 1	Vaserne har sin faste plads på reolen.
0 0 1 1 0 0	De regner med at have sin egen fest efter forestillingen.
0 0 1 1 0 1	De plejer at have sin faste plads på reolen.
0 0 1 1 1 0	Klovnene regner med at have sin egen fest efter forestillingen.
0 0 1 1 1 1	Vaserne plejer at have sin faste plads på reolen.

Table C.37: Primary item 17

Code	Sentence
0 0 0 0 0 0	De skjuler sine formuer i skattely. they hide REFL's assets in tax haven
0 0 0 0 0 1	De skjuler sine mange gram sukker. they hide REFL's many grams sugar
0 0 0 0 1 0	Milliardærerne skjuler sine formuer i skattely. billionaires.DEF hide REFL's assets in tax haven
0 0 0 0 1 1	Sodavandene skjuler sine mange gram sukker. sodas.DEF hide REFL's many grams sugar
0 0 0 1 0 0	De forsøger at skjule sine formuer i skattely. they try to hide REFL's assets in tax haven
0 0 0 1 0 1	De formår at skjule sine mange gram sukker. they manage to hide REFL's many grams sugar
0 0 0 1 1 0	Milliardærerne forsøger at skjule sine formuer i skattely. billionaires.DEF try to hide REFL's assets in tax haven
0 0 0 1 1 1	Sodavandene formår at skjule sine mange gram sukker. sodas.DEF manage to hide REFL's many grams sugar
0 0 1 0 0 0	De skjuler sin formue i skattely.
0 0 1 0 0 1	De skjuler sit høje sukkerindhold. sit høje sukkerindhold = REFL's high sugar-content
0 0 1 0 1 0	Milliardærerne skjuler sin formue i skattely.
0 0 1 0 1 1	Sodavandene skjuler sit høje sukkerindhold.
0 0 1 1 0 0	De forsøger at skjule sin formue i skattely.
0 0 1 1 0 1	De formår at skjule sit sukkerindhold.
0 0 1 1 1 0	Milliardærerne forsøger at skjule sin formue i skattely.
0 0 1 1 1 1	Sodavandene formår at skjule sit sukkerindhold.

Table C.38: Primary item 18

Code	Sentence
0 0 0 0 0 0	De taber sine nøgler i bussen. they lose REFL's house-key in bus.DEF
0 0 0 0 0 1	De taber sine blade i stormvejret. they lose REFL's leaves in storm-weather.DEF
0 0 0 0 1 0	Tvillingerne taber sine nøgler i bussen. twins.DEF lose REFL's house-key in bus.DEF
0 0 0 0 1 1	Planterne taber sine blade i stormvejret. plants.DEF lose REFL's leaves in storm-weather.DEF
0 0 0 1 0 0	De frygter at tabe sine nøgler i bussen. they fear to lose REFL's house-key in bus.DEF
0 0 0 1 0 1	De plejer at tabe sine blade i stormvejret. they use to lose REFL's leaves in storm-weather.DEF
0 0 0 1 1 0	Tvillingerne frygter at tabe sine nøgler i bussen. twins.DEF fear to lose REFL's house-key in bus.DEF
0 0 0 1 1 1	Planterne plejer at tabe sine blade i stormvejret. plants.DEF use to lose REFL's leaves in storm-weather.DEF
0 0 1 0 0 0	De taber sin husnøgle i bussen.
0 0 1 0 0 1	De taber sit løv i stormvejret. sit løv = REFL's foliage
0 0 1 0 1 0	Tvillingerne taber sin husnøgle i bussen.
0 0 1 0 1 1	Planterne taber sit løv i stormvejret.
0 0 1 1 0 0	De frygter at tabe sin husnøgle i bussen.
0 0 1 1 0 1	De plejer at tabe sit løv i stormvejret.
0 0 1 1 1 0	Tvillingerne frygter at tabe sin husnøgle i bussen.
0 0 1 1 1 1	Planterne plejer at tabe sit løv i stormvejret.

Table C.39: Primary item 19

Code	Sentence
0 0 0 0 0	De kaster sine penge efter obskure brætspil. they throw REFL's money after obscure board-games
0 0 0 0 1	De kaster sine farlige UV-stråler på planetens få beboere. they throw REFL's dangerous UV-beams on planet.DEF's few citizens
0 0 0 1 0	Nørderne kaster sine penge efter obskure brætspil. geeks.DEF throw REFL's money after obscure board-games
0 0 0 1 1	Solene kaster sine farlige UV-stråler på planetens få beboere. suns.DEF throw REFL's dangerous UV-beams on planet.DEF's few citizens
0 0 0 1 0 0	De vælger at kaste sine penge efter obskure brætspil. they choose to throw REFL's money after obscure board-games
0 0 0 1 0 1	De dræber ved at kaste sine farlige UV-stråler på planetens få beboere. they kill by to throw REFL's dangerous UV-beams on planet.DEF's few citizens
0 0 0 1 1 0	Nørderne vælger at kaste sine penge efter obskure brætspil. geeks.DEF choose to throw REFL's money after obscure board-games
0 0 0 1 1 1	Solene dræber ved at kaste sine farlige UV-stråler på planetens få beboere. suns.DEF kill by to throw REFL's dangerous UV-beams on planet.DEF's few citizens
0 0 1 0 0 0	De kaster sin opsparing efter obskure brætspil. sin opsparing = REFL's savings.SG
0 0 1 0 0 1	De kaster sin farlige UV-stråling på planetens få beboere. sin farlige UV-stråling = REFL's dangerous UV-radiation
0 0 1 0 1 0	Nørderne kaster sin opsparing efter obskure brætspil.
0 0 1 0 1 1	Solene kaster sin farlige UV-stråling på planetens få beboere.
0 0 1 1 0 0	De vælger at kaste sin opsparing efter obskure brætspil.
0 0 1 1 0 1	De dræber ved at kaste sine farlige UV-stråling på planetens få beboere.
0 0 1 1 1 0	Nørderne vælger at kaste sin opsparing efter obskure brætspil.
0 0 1 1 1 1	Solene dræber ved at kaste sine farlige UV-stråling på planetens få beboere.

Table C.40: Primary item 20

Code	Sentence
0 0 0 0 0 0	De åbner sine magiske æsker. they open REFL's magical boxes
0 0 0 0 0 1	De åbner sine døre. they open REFL's doors
0 0 0 0 1 0	Tryllekunsternerne åbner sine magiske æsker. magicians.DEF open REFL's magical boxes
0 0 0 0 1 1	Elevatorene åbner sine døre. lifts.DEF open REFL's doors
0 0 0 1 0 0	De begynder at åbne sine magiske æsker. they begin to open REFL's magical boxes
0 0 0 1 0 1	De begynder at åbne sine døre. they begin to open REFL's doors
0 0 0 1 1 0	Tryllekunsternerne begynder at åbne sine magiske æsker. magicians.DEF begin to open REFL's magical boxes
0 0 0 1 1 1	Elevatorene begynder at åbne sine døre. lifts.DEF begin to open REFL's doors
0 0 1 0 0 0	De åbner sin magiske æske.
0 0 1 0 0 1	De åbner sin dør.
0 0 1 0 1 0	Tryllekunsternerne åbner sin magiske æske.
0 0 1 0 1 1	Elevatorene åbner sin dør.
0 0 1 1 0 0	De begynder at åbne sin magiske æske.
0 0 1 1 0 1	De begynder at åbne sin dør.
0 0 1 1 1 0	Tryllekunsternerne begynder at åbne sin magiske æske.
0 0 1 1 1 1	Elevatorene begynder at åbne sin dør.

Table C.41: Primary item 21

Code	Sentence
0 0 0 0 0	De klarer sine konflikter uden skænderier. they manage REFL's conflicts without quarrels
0 0 0 0 1	De klarer sine 0-100 km/t på 4 sekunder. they manage REFL's 0-100 km/hour on 4 seconds
0 0 0 1 0	Forældrene klarer sine konflikter uden skænderier. parents.DEF manage REFL's conflicts without quarrels
0 0 0 1 1	Motorcyklerne klarer sine 0-100 km/t på 4 sekunder. motor-bikes.DEF manage REFL's 0-100 km/hour on 4 seconds
0 0 0 1 0 0	De forsøger at klare sine konflikter uden skænderier. they try to manage REFL's conflicts without quarrels
0 0 0 1 0 1	De plejer at klare sine 0-100 km/t på 4 sekunder. they use to manage REFL's 0-100 km/hour on 4 seconds
0 0 0 1 1 0	Forældrene forsøger at klare sine konflikter uden skænderier. parents.DEF try to manage REFL's conflicts without quarrels
0 0 0 1 1 1	Motorcyklerne plejer at klare sine 0-100 km/t på 4 sekunder. motor-bikes.DEF use to manage REFL's 0-100 km/hour on 4 seconds
0 0 1 0 0 0	De klarer sin skilsmisse næsten uden skænderier. sin skilsmisse næsten uden skænderier = REFL's divorce almost without quarrals
0 0 1 0 0 1	De klarer sit syn ved mekanikeren uden anmærkninger. they manage REFL's MOT-test by mechanic.DEF without comments
0 0 1 0 1 0	Forældrene klarer sin skilsmisse næsten uden skænderier.
0 0 1 0 1 1	Motorcyklerne klarer sit syn ved mekanikeren uden anmærkninger.
0 0 1 1 0 0	De forsøger at klare sin skilsmisse uden skænderier.
0 0 1 1 0 1	De plejer at klare sit syn ved mekanikeren uden anmærkninger.
0 0 1 1 1 0	Forældrene forsøger at klare sin skilsmisse uden skænderier.
0 0 1 1 1 1	Motorcyklerne plejer at klare sit syn ved mekanikeren uden anmærkninger.

Table C.42: Primary item 22

Code	Sentence
0 0 0 0 0 0	De forbedrer sine læsevner hvert år. they improve REFL's reading-skills every year
0 0 0 0 0 1	De forbedrer sine markedsandele næste år. they improve REFL's market-shares next year
0 0 0 0 1 0	Pigerne forbedrer sine læsevner hvert år. girls.DEF improve REFL's reading-skills every year
0 0 0 0 1 1	Skoene forbedrer sine markedsandele næste år. shoes.DEF improve REFL's market-shares next year
0 0 0 1 0 0	De ventes at forbedre sine læsevner hvert år. they expect.PASS to improve REFL's reading-skills every year
0 0 0 1 0 1	De ventes at forbedre sine markedsandele næste år. they wait.PASS to improve REFL's market-shares next year
0 0 0 1 1 0	Pigerne ventes at forbedre sine læsevner hvert år. girls.DEF expect.PASS to improve REFL's reading-skills every year
0 0 0 1 1 1	Skoene ventes at forbedre sine markedsandele næste år. shoes.DEF wait.PASS to improve REFL's market-shares next year
0 0 1 0 0 0	De forbedrer sin læsevne hvert år.
0 0 1 0 0 1	De forbedrer sin markedandel næste år.
0 0 1 0 1 0	Pigerne forbedrer sin læsevne hvert år.
0 0 1 0 1 1	Skoene forbedrer sin markedandel næste år.
0 0 1 1 0 0	De ventes at forbedre sin læsevne hvert år.
0 0 1 1 0 1	De ventes at forbedre sin markedandel næste år.
0 0 1 1 1 0	Pigerne ventes at forbedre sin læsevne hvert år.
0 0 1 1 1 1	Skoene ventes at forbedre sin markedandel næste år.

Table C.43: Primary item 23

Code	Sentence
0 0 0 0 0	De fastholder sine forklaringer. they maintain REFL's explanations
0 0 0 0 1	De fastholder sine læsere. they capture REFL's readers
0 0 0 1 0	Vidnerne fastholder sine forklaringer. witness.DEF maintain REFL's explanations
0 0 0 1 1	Tegneserierne fastholder sine læsere. comics.DEF capture REFL's readers
0 0 0 1 0 0	De plejer at fastholde sine forklaringer. they use to maintain REFL's explanations
0 0 0 1 0 1	De plejer at fastholde sine læsere. they use to capture REFL's readers
0 0 0 1 1 0	Vidnerne plejer at fastholde sine forklaringer. witness.DEF use to maintain REFL's explanations
0 0 0 1 1 1	Tegneserierne plejer at fastholde sine læsere. comics.DEF use to capture REFL's readers
0 0 1 0 0 0	De fastholder sin forklaring.
0 0 1 0 0 1	De fastholder sin læser.
0 0 1 0 1 0	Vidnerne fastholder sin forklaring.
0 0 1 0 1 1	Tegneserierne fastholder sin læser.
0 0 1 1 0 0	De plejer at fastholde sin forklaring.
0 0 1 1 0 1	De plejer at fastholde sin læser.
0 0 1 1 1 0	Vidnerne plejer at fastholde sin forklaring.
0 0 1 1 1 1	Tegneserierne plejer at fastholde sin læser.

Table C.44: Primary item 24

Code	Sentence
0 0 0 0 0 0	De beskytter sine territorier. they protect REFL's territories
0 0 0 0 0 1	De beskytter sine brugere mod skadelig software. they protect REFL's users against harmful software
0 0 0 0 1 0	Kattene beskytter sine territorier. cats.DEF protect REFL's territories
0 0 0 0 1 1	Programmerne beskytter sine brugere mod skadelig software. programmes.DEF protect REFL's users against harmful software
0 0 0 1 0 0	De kæmper for at beskytte sine territorier. they fight for to protect REFL's territories
0 0 0 1 0 1	De hjælper med at beskytte sine brugere mod skadelig software. they help with to protect REFL's users against harmful software
0 0 0 1 1 0	Kattene kæmper for at beskytte sine territorier. cats.DEF fight for to protect REFL's territories
0 0 0 1 1 1	Programmerne hjælper med at beskytte sine brugere mod skadelig software. programmes.DEF help with to protect REFL's users against harmful software
0 0 1 0 0 0	De beskytter sit territorie.
0 0 1 0 0 1	De beskytter sin bruger mod skadelig software.
0 0 1 0 1 0	Kattene beskytter sit territorie.
0 0 1 0 1 1	Programmerne beskytter sin bruger mod skadelig software.
0 0 1 1 0 0	De kæmper for at beskytte sit territorie.
0 0 1 1 0 1	De hjælper med at beskytte sin bruger mod skadelig software.
0 0 1 1 1 0	Kattene kæmper for at beskytte sit territorie.
0 0 1 1 1 1	Programmerne hjælper med at beskytte sin bruger mod skadelig software.

Table C.45: Primary item 25

Code	Sentence
0 0 0 0 0 0	De starter sine timer med sang. they start REFL's lessons with song
0 0 0 0 0 1	De starter sine prøvesignaler klokken 12. they start REFL's test-signals clock.DEF 12
0 0 0 0 1 0	Lærerne starter sine timer med sang. teachers.DEF start REFL's lessons with song
0 0 0 0 1 1	Sirenerne starter sine prøvesignaler klokken 12. sirens.DEF start REFL's test-signals clock.DEF 12
0 0 0 1 0 0	De plejer at starte sine timer med sang. they use to start REFL's lessons with song
0 0 0 1 0 1	De plejer at starte sine prøvesignaler klokken 12. they use to start REFL's test-signals clock.DEF 12
0 0 0 1 1 0	Lærerne plejer at starte sine timer med sang. teachers.DEF use to start REFL's lessons with song
0 0 0 1 1 1	Sirenerne plejer at starte sine prøvesignaler klokken 12. sirens.DEF use to start REFL's test-signals clock.DEF 12
0 0 1 0 0 0	De starter sin time med sang.
0 0 1 0 0 1	De starter sit prøvesignal klokken 12.
0 0 1 0 1 0	Lærerne starter sin time med sang.
0 0 1 0 1 1	Sirenerne starter sit prøvesignal klokken 12.
0 0 1 1 0 0	De plejer at starte sin time med sang.
0 0 1 1 0 1	De plejer at starte sit prøvesignal klokken 12.
0 0 1 1 1 0	Lærerne plejer at starte sin time med sang.
0 0 1 1 1 1	Sirenerne plejer at starte sit prøvesignal klokken 12.

Table C.46: Primary item 26

Code	Sentence
0 0 0 0 0 0	De gennemfører sine indbrud om natten. they perform REFL's burglaries at night.DEF
0 0 0 0 0 1	De gennemfører sine missioner om natten. they complete REFL's missions at night.DEF
0 0 0 0 1 0	Tyvene gennemfører sine indbrud om natten. burglars.DEF perform REFL's burglaries at night.DEF
0 0 0 0 1 1	Dronerne gennemfører sine missioner om natten. drones.DEF complete REFL's missions at night.DEF
0 0 0 1 0 0	De planlægger at gennemføre sine indbrud om natten. they plan to perform REFL's burglaries at night.DEF
0 0 0 1 0 1	De forhindres i at gennemføre sine missioner om natten. they prevent.PASS in to complete REFL's missions at night.DEF
0 0 0 1 1 0	Tyvene planlægger at gennemføre sine indbrud om natten. burglars.DEF plan to perform REFL's burglaries at night.DEF
0 0 0 1 1 1	Dronerne forhindres i at gennemføre sine missioner om natten. drones.DEF prevent.PASS in to complete REFL's missions at night.DEF
0 0 1 0 0 0	De gennemfører sit indbrud om natten.
0 0 1 0 0 1	De gennemfører sin mission om natten.
0 0 1 0 1 0	Tyvene gennemfører sit indbrud om natten.
0 0 1 0 1 1	Dronerne gennemfører sin mission om natten.
0 0 1 1 0 0	De planlægger at gennemføre sit indbrud om natten.
0 0 1 1 0 1	De forhindres i at gennemføre sin mission om natten.
0 0 1 1 1 0	Tyvene planlægger at gennemføre sit indbrud om natten.
0 0 1 1 1 1	Dronerne forhindres i at gennemføre sin mission om natten.

Table C.47: Primary item 27

Code	Sentence
0 0 0 0 0 0	De kaster sine terninger tre gange. they throw REFL's dice three times
0 0 0 0 0 1	De kaster automatisk sine bomber. they throw REFL's bombs automatically
0 0 0 0 1 0	Nørderne kaster sine terninger tre gange. geeks.DEF throw REFL's dice three times
0 0 0 0 1 1	Flyene kaster automatisk sine bomber. planes.DEF throw REFL's bombs automatically
0 0 0 1 0 0	De skal til at kaste sine terninger tre gange. they shall to to throw REFL's dice three times
0 0 0 1 0 1	De programmeres til at kaste sine bomber automatisk. they program.PASS to to throw REFL's bombs automatically
0 0 0 1 1 0	Nørderne skal til at kaste sine terninger tre gange. geeks.DEF shall to to throw REFL's dice three times
0 0 0 1 1 1	Flyene programmeres til at kaste sine bomber automatisk. planes.DEF program.PASS to to throw REFL's bombs automatically
0 0 1 0 0 0	De kaster sin terning tre gange.
0 0 1 0 0 1	De kaster automatisk sin største bombe. sin største bombe = REFL's biggest bomb
0 0 1 0 1 0	Nørderne kaster sin terning tre gange.
0 0 1 0 1 1	Flyene kaster automatisk sin største bombe.
0 0 1 1 0 0	De skal til at kaste sin terning tre gange.
0 0 1 1 0 1	De programmeres til at kaste sin største bombe automatisk.
0 0 1 1 1 0	Nørderne skal til at kaste sin terning tre gange.
0 0 1 1 1 1	Flyene programmeres til at kaste sin største bombe automatisk.

Table C.48: Primary item 28

Code	Sentence
0 0 0 0 0 0	De finder sine nye jobs på nettet. they find REFL's new jobs on web.DEF
0 0 0 0 0 1	De finder sine inspirationskilder i Korea. they find REFL's inspiration-sources in Korea
0 0 0 0 1 0	Tømrerne finder sine nye jobs på nettet. carpenter.DEF find REFL's new jobs on web.DEF
0 0 0 0 1 1	Musiknumrene finder sine inspirationskilder i Korea. music-numbers.DEF find REFL's inspiration-sources in Korea
0 0 0 1 0 0	De prøver at finde sine nye jobs på nettet. they try to find REFL's new jobs on web.DEF
0 0 0 1 0 1	De anklages for at finde sine inspirationskilder i Korea. they accuse.PASS for to find REFL's inspiration-sources in Korea
0 0 0 1 1 0	Tømrerne prøver at finde sine nye jobs på nettet. carpenter.DEF try to find REFL's new jobs on web.DEF
0 0 0 1 1 1	Musiknumrene anklages for at finde sine inspirationskilder i Korea. music-numbers.DEF accuse.PASS for to find REFL's inspiration-sources in Korea
0 0 1 0 0 0	De finder sit nye job på nettet.
0 0 1 0 0 1	De finder sin inspirationskilde i Korea.
0 0 1 0 1 0	Tømrerne finder sit nye job på nettet.
0 0 1 0 1 1	Musiknumrene finder sin inspirationskilde i Korea.
0 0 1 1 0 0	De prøver at finde sit nye job på nettet.
0 0 1 1 0 1	De anklages for at finde sin inspirationskilde i Korea.
0 0 1 1 1 0	Tømrerne prøver at finde sit nye job på nettet.
0 0 1 1 1 1	Musiknumrene anklages for at finde sin inspirationskilde i Korea.

Table C.49: Primary item 29

Code	Sentence
0 0 0 0 0	De holder sine løfter. they hold REFL's promises
0 0 0 0 1	De holder sine budgetter. they hold REFL's budgets
0 0 0 1 0	Håndværkerne holder sine løfter. tradespersons.DEF hold REFL's promises
0 0 0 1 1	Regionerne holder sine budgetter. regions.DEF hold REFL's budgets
0 0 0 1 0 0	De ventes at holde sine løfter. they expect.PASS to hold REFL's promises
0 0 0 1 0 1	De ventes at holde sine budgetter. they expect.PASS to hold REFL's budgets
0 0 0 1 1 0	Håndværkerne ventes at holde sine løfter. tradespersons.DEF expect.PASS to hold REFL's promises
0 0 0 1 1 1	Regionerne ventes at holde sine budgetter. regions.DEF expect.PASS to hold REFL's budgets
0 0 1 0 0 0	De holder sit løfte.
0 0 1 0 0 1	De holder sit budget.
0 0 1 0 1 0	Håndværkerne holder sit løfte.
0 0 1 0 1 1	Regionerne holder sit budget.
0 0 1 1 0 0	De ventes at holde sit løfte.
0 0 1 1 0 1	De ventes at holde sit budget.
0 0 1 1 1 0	Håndværkerne ventes at holde sit løfte.
0 0 1 1 1 1	Regionerne ventes at holde sit budget.

Table C.50: Primary item 30

Code	Sentence
0 0 0 0 0 0	De spreder sine net ud i havet. they spread REFL's nets out in ocean.DEF
0 0 0 0 0 1	De spreder sine vandstråler ud over haven. they spread REFL's water-beams out over garden.DEF
0 0 0 0 1 0	Fiskerne spreder sine net ud i havet. fishermen.DEF wait on to spread REFL's nets out in ocean.DEF
0 0 0 0 1 1	Sprinklerne spreder sine vandstråler ud over haven. sprinklers.DEF spread REFL's water-beams out over garden.DEF
0 0 0 1 0 0	De venter på at sprede sine net ud i havet. they wait on to spread REFL's nets out in ocean.DEF
0 0 0 1 0 1	De er gode til at sprede sine vandstråler ud over haven. they are good to to spread REFL's water-beams out over garden.DEF
0 0 0 1 1 0	Fiskerne venter på at sprede sine net ud i havet. fishermen.DEF wait on to spread REFL's nets out in ocean.DEF
0 0 0 1 1 1	Sprinklerne er gode til at sprede sine vandstråler ud over haven. sprinklers.DEF are good to to spread REFL's water-beams out over garden.DEF
0 0 1 0 0 0	De spreder sit net ud i havet.
0 0 1 0 0 1	De spreder sin vandstråle ud over haven.
0 0 1 0 1 0	Fiskerne spreder sit net ud i havet.
0 0 1 0 1 1	Sprinklerne spreder sin vandstråle ud over haven.
0 0 1 1 0 0	De venter på at sprede sit net ud i havet.
0 0 1 1 0 1	De er gode til at sprede sin vandstråle ud over haven.
0 0 1 1 1 0	Fiskerne venter på at sprede sit net ud i havet.
0 0 1 1 1 1	Sprinklerne er gode til at sprede sin vandstråle ud over haven.

Table C.51: Primary item 31

Code	Sentence
0 0 0 0 0 0	De udvikler sine hovedpersoner sammen. they develop REFL's main-characters together
0 0 0 0 0 1	De udvikler sine bobler gennem gæring. they develop REFL's bubbles through fermentation
0 0 0 0 1 0	Forfatterne udvikler sine hovedpersoner sammen. authors.DEF develop REFL's main-characters together
0 0 0 0 1 1	Vinene udvikler sine bobler gennem gæring. wines.DEF develop REFL's bubbles through fermentation
0 0 0 1 0 0	De aftaler at udvikle sine hovedpersoner sammen. they agree to develop REFL's main-characters together
0 0 0 1 0 1	De begynder at udvikle sine bobler gennem gæring. they begin to develop REFL's bubbles through fermentation
0 0 0 1 1 0	Forfatterne aftaler at udvikle sine hovedpersoner sammen. authors.DEF agree to develop REFL's main-characters together
0 0 0 1 1 1	Vinene begynder at udvikle sine bobler gennem gæring. wines.DEF begin to develop REFL's bubbles through fermentation
0 0 1 0 0 0	De udvikler sin hovedperson sammen.
0 0 1 0 0 1	De udvikler sin fine brus gennem gæring. sin fine brus = REFL's fine fizz
0 0 1 0 1 0	Forfatterne udvikler sin hovedperson sammen.
0 0 1 0 1 1	Vinene udvikler sin fine brus gennem gæring.
0 0 1 1 0 0	De aftaler at udvikle sin hovedperson sammen.
0 0 1 1 0 1	De begynder at udvikle sin fine brus gennem gæring.
0 0 1 1 1 0	Forfatterne aftaler at udvikle sin hovedperson sammen.
0 0 1 1 1 1	Vinene begynder at udvikle sin fine brus gennem gæring.

Table C.52: Primary item 32

Code	Sentence
0 0 0 0 0 0	De følger sine børn i skole. they follow REFL's children in school
0 0 0 0 0 1	De følger sine hovedpersoner gennem flere år. they follow REFL's main-characters through several years
0 0 0 0 1 0	Fædrene følger sine børn i skole. fathers.DEF follow REFL's children in school
0 0 0 0 1 1	Filmene følger sine hovedpersoner gennem flere år. movies.DEF follow REFL's main-characters through several years
0 0 0 1 0 0	De prioriterer at følge sine børn i skole. they prioritize to follow REFL's children in school
0 0 0 1 0 1	De roses for at følge sine hovedpersoner gennem flere år. they praise.PASS for to follow REFL's main-characters through several years
0 0 0 1 1 0	Fædrene prioriterer at følge sine børn i skole. fathers.DEF prioritize to follow REFL's children in school
0 0 0 1 1 1	Filmene roses for at følge sine hovedpersoner gennem flere år. movies.DEF praise.PASS for to follow REFL's main-characters through several years
0 0 1 0 0 0	De følger sit barn i skole.
0 0 1 0 0 1	De følger sin hovedperson gennem flere år.
0 0 1 0 1 0	Fædrene følger sit barn i skole.
0 0 1 0 1 1	Filmene følger sin hovedperson gennem flere år.
0 0 1 1 0 0	De prioriterer at følge sit barn i skole.
0 0 1 1 0 1	De roses for at følge sin hovedperson gennem flere år.
0 0 1 1 1 0	Fædrene prioriterer at følge sit barn i skole.
0 0 1 1 1 1	Filmene roses for at følge sin hovedperson gennem flere år.

C.3 All 30 secondary sentences with ratings

The tables in this section contain all the 30 secondary sentences. They are divided into categories. The mean rating given to the sentences from the study as a whole is shown next to the sentence.

Table C.53: All secondary sentences divided into subcategories and shown with mean ratings

Subcategory	Sentences used in questionnaire	Mean rating
Local <i>sin</i>	Hun hentede sine pakker på posthuset. She collected REFL's parcels at post-house.DEF ' She ₁ collected her ₁ parcels at the post office.'	4.60909
	Johan fandt sin kat i garagen. Johan found REFL's cat in garage.DEF ' Johan ₁ found his ₁ cat in the garage.'	4.71272
	Han tror at hun køber sine varer på nettet. He thinks that she buys REFL's groceries on web.DEF 'He thinks that she ₁ buys her ₁ groceries online.'	4.20727
	Hun vidste at jeg passede hendes kat. She knew that I looked-after her cat. ' She ₁ knew that I looked after her ₁ cat.'	4.60181
	Mathilde lovede mig at vaske sit hår. Mathilde promised me to wash REFL's hair. ' Mathilde ₁ promised me to wash her ₁ hair.'	4.12727
Local <i>hans</i>	Magnus havde en kat med et fjollet navn. Nu vil han kalde hans nye kat det samme. Magnus had a cat with a silly name. Now will he name his new cat the same. 'Magnus had a cat with a silly name. Now he ₁ wants to name his ₁ new cat the same.'	4.07818
	Morten ved at Sofie fik hendes hudsygdom af stress. Morten knows that Sofie got her skin-disease from stress. 'Morten knows that Sofie ₁ got her ₁ skin disease through stress.'	4.12181
No binder	* Det var sin mor Peter mødte. It was REFL's mother Peter met. 'It was his ₁ mother Peter ₁ met.'	2.45272
	* Det var sin underbo, der gjorde Peter sur. It was REFL's downstairs-neighbour who made Peter mad. 'It was his ₁ downstairs neighbour who made Peter ₁ mad.'	1.88909
	* Hun vidste godt at sine pakker var på vej. She knew well that REFL's parcels were on way. ' She ₁ was well aware that her ₁ parcels were on their way.'	2.46909
	* Hun vidste at jeg passede sin kat. She knew that I looked-after REFL's cat. ' She ₁ knew that I looked after her ₁ cat.'	1.64545
	* Vi lovede ham at tage sin plakat med. We promised him to bring REFL's poster with. 'We promised him ₁ to bring his ₁ poster.'	2.03090

Table C.54: Division of secondary sentences into subcategories (contd.)

Subcategory	Sentences used in questionnaire	Mean rating
Non-local <i>sin</i>	Hun bad mig passe hendes kat. She asked me look-after her cat. ' She ₁ asked me to look after her ₁ cat.'	4.49818
	Hun bad mig passe sin kat. She asked me look-after REFL's cat. ' She ₁ asked me to look after her ₁ cat.'	3.56545
DP-spec binder	Vi elsker Magnus' tegning af sin kat. We love Magnus' drawing of REFL's cat. 'We love Magnus ₁ ' drawing of his ₁ cat.'	3.89454
	Jeg så Peters fotografi af hans forældre og var meget imponeret. I saw Peter's photograph of his parents and was very impressed. 'I saw Peter ₁ 's photograph of his ₁ parents and was very impressed.'	4.23272
sig good	Jeg prøvede at genskabe Julies tegning af sine heste. I tried to recreate Julie's drawing of REFL's horses. 'I tried to recreate Julie ₁ 's drawing of her ₁ horses.'	3.37454
	Hun bad mig hjælpe sig med lektierne. She asked me help REFL with homework.DEFPL ' She ₁ asked me to help her ₁ with her homework.'	2.29636
sig selv	Han hader sig selv. He hates REFL self. ' He ₁ hates himself ₁ .'	4.83818
	Han ramte sig selv med malingen. He hit REFL self with paint.DEF ' He ₁ hit himself ₁ with the paint.'	4.64363
sig selv	De skammer sig over deres gæld. They shame REFL over their debt. ' They ₁ are ashamed of their ₁ debt.'	4.70181
	Vi fortalte hende om hende selv. We told her about herself. 'We told her ₁ about herself ₁ .'	3.46
sig selv	Jeg fortalte Morten om ham selv. I told Morten about himself. 'I told Morten ₁ about himself ₁ .'	3.76727
	Hun viste drengen ham selv i spejlet. She showed boy.DEF himself in mirror.DEF 'She showed the boy ₁ himself ₁ in the mirror.'	3.07636

Table C.55: Division of secondary sentences into subcategories (further contd.)

Subcategory	Sentences used in questionnaire	Mean rating
sig bad	* Hun bad mig hjælpe sig selv med lektierne. She asked me help REFL self with homework.DEF.PL ' She ₁ asked me to help herself ₁ with her homework.'	1.77090
	* Han hader sig. He hates REFL. ' He ₁ hates himself ₁ .'	1.52727
	* Vi så ham ramme ham selv med malingen. We saw him hit himself with paint.DEF 'We saw him hit himself with the paint.'	3.11272
	* De skammer dem over deres gæld. They shame them over their debt. 'They are ashamed of their debt.'	2.77818
	* Vi fortalte hende om sig selv. We told her about REFL self. 'We told her about herself.'	2.71272
	* Jeg viste drengen sig selv i spejlet. I showed boy.DEF REFL self in mirror.DEF 'I showed the boy himself in the mirror.'	2.50363

Appendix D

Examples of plural antecedent *sin* from e.g. news media – a very informal collection

This appendix contains approximately 100 screenshots from various online sources (news media, social media, university communication, a bus fine, a library website, and the Danish health authorities). They are an entirely informal collection from the last four years. I have collected the screenshots myself, which also explains why the topics are rather biased towards e.g. child rearing and the Covid pandemic. They should be read mainly as a curiosity.

This appendix is empty in the online version available at library.au.dk. Contact the author at karo@cc.au.dk to view the full appendix.

