

Fundamentfeltet kan fyldes med lidt af hvert. En undersøgelse på tværs af *KorpusDK*, *Hestenettet* og *LANCHART*

Anne Mette Nyvad, Katrine Rosendal Ehlers
& Ken Ramshøj Christensen

Abstract

This study examines the Danish Fundamentfelt (F), the clause-initial position before the finite verb, using data from three corpora: Korpus-DK (formal writing), Hestenettet (informal writing), and LANCHART (spoken language). The findings show that subjects are most frequent in F, but their distribution varies by modality. Full nominal subjects are preferred in formal writing, while spoken language features more pronominal subjects and fronted objects. Adverbials are common across corpora but differ in complexity. The study highlights how discourse structure and grammatical choices in F appear to be shaped by formality and mode of communication.

Nøgleord

formalitetsniveau, modalitet, korpusdata, informationsstruktur, topic

1. Hvad står i Fundamentfeltet?

I dansk optræder hovedsætningens første led på pladsen før det finitte verbum, og i traditionel dansk grammatik har denne initiale plads siden Diderichsen (1936, 1946) gået under betegnelsen Fundamentfeltet (F). F er blevet betragtet som “fundamentet under den sagsfremstilling som sætningen giver” (Jørgensen 2021: 88), og der er tradition for at betragte subjektet som det umarkerede valg i sætningens første position (Mikkelsen 1911: 573ff; Diderichsen 1962: 171; Hansen og Heltoft 2011: 74,

se dog Hansen 1933: 69ff). Mikkelsen (1911: 574) omtaler ordstillingen i subjektsinitiale sætninger som "ligefrem" og den i ikke-subjektsinitiale som "omvendt" og lægger således op til, at den omvendte ordstilling er det markerede valg, da den forekommer "når et ord eller en ordforbindelse sættes først i sætningen enten for eftertrykkets skyld, eller fordi begrebet stiller sig først frem for tanken". Diderichsen (1946) påpegede i tråd hermed, at sætninger ofte starter med entiteter eller situationer, som er blevet nævnt tidligere, og som resten af sætningen således kommenterer på, og han pegede på to betingelser for, at et andet element end subjektet optræder i F: *Hensyn til det foregaaende* og *Udhævelse/Emfase*.

Fremflytning til F i form af topikalisering markerer et syntaktisk elements status som *topic*, og informationsstruktur spiller i det hele taget en åbenlys rolle for, hvordan syntaktiske elementer er organiseret i en sætning. Eftersom teorier om informationsstruktur har en tendens til dels at bruge forskellige betegnelser for det samme begreb og dels at bruge den samme betegnelse for forskellige begreber (se Chafe 1976), holder vi os her til følgende begreber af primær interesse: *topic/tema*, som svarer nogenlunde til 'givet information' (jf. Prag-skolen, Firbas 1992), og *fokus/rema*, der mere eller mindre dækker 'ny information'. Topic er et begreb, hvis definition der ikke er opnået konsensus omkring, men det blev oprindeligt introduceret i den lingvistiske litteratur som "det psykologiske subjekt" (*das Psychologische Subjekt*, von der Gabelentz 1869), og der er muligvis en tværsproglig tendens til at placere givet information sætningsinitialt (Tomlin 1986). I den danske funktionelle tilgang (fx Heltoft 1992; Jørgensen 2000: 81-85) bliver F betragtet som multifunktionel. Hansen & Heltoft (2011) argumenterer for, at en af F's vigtigste funktioner er kodningen af illokutionær ramme (deklarativ vs. interrogativ), snarere end at udgøre sætningens "fundament", da det kan indeholde både *tema* (umarkeret eller markeret) og *rema* (fokus). Derudover skelner de mellem *tema* og *emne* ($\approx$ topic): Uanset hvor subjektet er placeret, er det ifølge Hansen & Heltoft (2011) *tema*, mens F typisk er *emne*.

I tråd hermed skriver Engdahl (1997), at F vil have informationsstruktur-rel funktion som fokus, *continuation* topic eller kontrastivt topic, mens Tøgeby (2003: 157) mener, at topic i dansk optræder først og fokus til sidst, hvilket ifølge Mathesius (1929) er den naturlige rækkefølge tværsprogligt. Lambrecht (1994) argumenterer for, at topic- og fokus-elementer konkurrerer om sætningens privilegerede første-position på tværs af alle sprog. Den funktionelle rolle, som sætningens første-position spiller, interagerer muligvis med V2 (*Verb*

Second, dvs. det forhold, at det finitte verbum som grundregel optræder i sætningens andenposition i hovedsætninger). Ifølge Los (2009) kan V2 tilmed ses som et syntaktisk system, der leverer en plads (den sætningsinitiale) til diskursforankring, og hun foreslår, at V2-fænomenet muligvis oprindeligt opstod som et stilistisk valg, der kunne bruges til at lede opmærksomheden hen på det sætningsinitiale elements særlige informationsstrukturelle status (Los 2009: 100). I tråd med dette har flere syntaktikere noteret, at F i dansk ofte indeholder topic-elementer (Erteschik-Shir 2007; Togeby 2003) og kobler til den forudgående diskurs (Harder & Poulsen 2001).

En række teoretikere (fx Basbøll 1976; Vikner & Jørgensen 2017; Christensen 2005) har påpeget, at Diderichsens sætningsskema og generative træstrukturer i vid udstrækning er compatible med hinanden. Ifølge Mikkelsen (2010: 2) er spørgsmålet om, hvad der står i F ($\approx$ CP-spec) imidlertid mere simpelt i den generative tradition end i den funktionelle: (i) I *hw*-spørgsmål indtager *hw*-elementet F-pladsen, (ii) i *ja/nej*-spørgsmål og bydende hovedsætninger er der en nuloperator i (F), og (iii) i alle andre V2-sætninger er det ikke syntaksen men snarere stilistik og tekstlingvistik, der afgør, hvad der står i F. Til gengæld har generative grammatikere i en lang årrække diskuteret, hvorvidt subjektsinitiale deklarative hovedsætninger indeholder et Complementizer Phrase (CP)-domæne eller ej (se fx Schwartz & Vikner 1989; Travis 1984; Zwart 1991), men konsensus er heller ikke her blevet opnået.¹

Det forhold, at vi kan placere fx et indirekte objekt i F, kan føre til tveetydighed på grund af de manglende kasusbøjninger i dansk, men Hansen (1933: 81-82) skriver om sætninger af typen *Peter gav Poul klø*, at de kun kan misforstås ”på Middelfart Sindssygeanstalt”, og at man i det talte sprog, udover kontekst, kan bruge ”den overdådige anvendelse af pronominer efter A-leddet [$\approx$ F]” til at disambiguere, dvs. *Peter, han gav Poul klø*, et fænomen, der bl.a. også går under betegnelsen ’ekstraposition til venstre’ (Petersen 2023, 2024) eller genoptaget fundament (Diderichsen 1966: 379). Vi kommer ikke nærmere ind på dette fænomen i nærværende artikel. Det gemmer vi til en anden god gang.

I dansk og de andre skandinaviske sprog kan en lang række elementer

1 Muligheden i dansk (og de andre fastlandsskandinaviske sprog) for at ekstrahere fra en række typer af ledsætninger, der egentlig burde være syntaktiske øer (Christensen, Kizach & Nyvad 2013; Christensen & Nyvad 2014; Nyvad, Müller & Christensen 2022; Christensen & Nyvad 2022) hænger muligvis også sammen med tendensen til at fremrykke elementer til F (se fx Engdahl 1997; Poulsen 2005: 45).

optræde i sætningsinitial position. Faktisk er der blevet argumenteret for, at alt i princippet kan stå i F i dansk, med undtagelse af det finitte verbum, negationen *ikke* og en håndfuld andre adverbier såsom *bare, jo, gerne, sgu, skam* (Jensen 2000: 2; Jørgensen 2000: 83; Mikkelsen 2015; se også Christensen 2005: 180-185).

Mikkelsens (2015) kvantitative studie analyserede 750 V2-sætninger fundet i aviser (primært *Weekendavisen*), fiktion (Jakob Ejersbos *Nordkraft*) og talt sprog (*BySoc*). Her fandt hun følgende fordeling af funktioner i F: subjekt 61%, adverbial 22%, objekt 9% og 7% andet (herunder tomt F og byden- de hovedsætninger). Dette mønster stemmer godt overens med Puggaards (2019) analyse af 500 sætninger fra samtaler (*Samtalegrammatik.dk* og *Samtalebank*), hvor han fandt et subjekt i F i 62% (310/500), et adverbial i 24,4% (122/500), objekt i 9,4% (47/500) og andet i 4,2% (21/500). Puggaard (2018: 2) argumenterer for, at den faktiske brug af F i talt, interaktionelt sprog er forskellig fra det mønster, der er blevet beskrevet i litteraturen: (i) Tunge led forekommer langt sjældnere, end tidligere antaget (de optræder typisk i ekstraposition før F i stedet, jf. Brøcker m.fl. 2012; Jørgensen 2016), og (ii) *det* samt *så* er påfaldende højfrekvente i interaktionel sprogbrug. Han betragter det som bemærkelsesværdigt, at F i næsten halvdelen af tilfældene var *det* eller *så*, på hvilken baggrund han konkluderer, at brugen af disse ord ikke støtter antagelsen af, at et ikke-subjekt i F er tekstuel markeret.

2. Undersøgelse på tværs af korpusser

Formålet med nærværende artikel er at udvide de korpusundersøgelser, der tidligere er foretaget, hvad angår den relative fordeling af subjekt, objekt og adverbial i F (eksempelvis Mikkelsen 2015 og Puggaard 2019) til et større empirisk grundlag, samt at undersøge, hvorvidt der er variation i fordelingen på tværs af korpusser med redigeret/uredigeret skrift/tale. Og på et mere detaljeret niveau: Hvad er den relative fordeling af ordklasser og typer?

2.1. Metode

Data til undersøgelsen er udtrukket fra tre forskellige danske korpusser: KorpusDK, Hestenettet og LANCHART. KorpusDK indeholder primært tekster i mere formelt og redigeret skriftsprog og indeholder 56 millioner ord fra tekster skrevet mellem 1983 og 2002. Hestenettet er en underdel af *The Danish Gigaword Corpus* (Strømberg-Derczynski 2021) og indeholder 228,9 millioner ord skrevet mellem 1997 og 2020. Hestenettet var et onlineforum

med en bred vifte af emner (ud over heste), og korpusset indeholder uformelt skriftsprog. LANCHART (Gregersen et al. 2014) er et dansk talesprogs-korpus, der strækker sig fra 1970'erne til helt nutidigt dansk talesprog. De tre korpusser adskiller sig både i type, størrelse og i, hvordan søgningerne kan foregå, hvorfor vi kort beskriver metoden til søgning i hvert korpus herunder.

Vi har trukket 1060 tilfældige sætninger ud af offlineudgaven af KorpusDK ved hjælp af et søgescript skrevet til formålet i Python (se Ehlers 2024: 70-71 for mere om offlineversionen af KorpusDK). Hestenettet har vi tilgået gennem *SketchEngine* (<https://www.sketchengine.eu/>) (Kilgarriff et al. 2004, 2014). Her har vi søgt på <s>, som markerer adskillelse mellem sætninger, og derefter taget et tilfældigt udsnit på 6000 sætninger af dette resultat. Vi har fået adgang til LANCHART gennem det online søgeværktøj *KORP* (<https://alf.hum.ku.dk/korp/>) (Borin et al. 2012) med søgetilladelse givet fra Københavns Universitet. Vi har fravalgt de tre underkorpusser, som indeholder dansk fra Amerika, Canada og Argentina, men vi har inkluderet de 18 resterende underkorpusser. Dette giver et samlet korpus på 10,79 millioner *tokens*. Heri har vi søgt på strengen <sentence> [redpos = ”.*”] som en CQP-søgning. Kommandoen <sentence> markerer adskillelse mellem sætninger, og [redpos = ”.*”] søger på et tilfældigt ord. Herefter har vi downloadet alle 907333 sætninger i resultatet og taget et tilfældigt udsnit af disse på 2500 sætninger ved hjælp af Excels indbyggede randomiseringsfunktion, RAND().

Vi ekskluderede resultater, som ikke indeholdt en hovedsætning (fx *hvis de ellers lige husker det ja*, fra LANCHART, eller udtryk såsom *Busstop lige udenfor rideklubben*, fra Hestenettet). Vi talte ikke *og*, *men*, *visse for* og lignende side- og samordnende konjunktioner med som elementer foran det finitte verbum, og heller ikke fx tøvende *oh*'er i LANCHART. Vi inkluderede dagbogssætninger som et nulusbjekt (*Tænker at I godt ved...*). Engelske sætninger (og sætninger der udelukkende er på andre sprog end dansk) blev udeladt. Sætninger, hvor det er uklart om elementet i F er subjekt eller ej, udelades (fx *Karina Kan sagtens hjælpe dig i gang, har en allround hafflinger som...* fra Hestenettet, hvor *Karina* enten er subjekt eller en vokativ). Det gav i alt 1000 eksempler i KorpusDK, 5106 i Hestenettet og 1984 i LANCHART.

Alle tre datasæt er annoteret efter de samme overordnede kategorier: Subjekt (herunder foreløbige og formelle subjekter og passiverede objekter af typen *Barylerne blev døbt af ...*), Objekt (herunder subjektsprædikater og præpositionelle objekter af typen *Lars tænkte jeg på*), Adverbial og Nul. Vi har gennemlæst sætningerne i de tre datasæt, markeret det finitte verbum i ho-

vedsætningen og derefter annoteret elementet foran det finitte verbum ud fra de fire kategorier. Adverbialerne inkluderer alt fra enkeltordsadverbialer til sætningsformede adverbialer (Petersen 2023). Kategorien Nul inkluderer blandt andet *ja/nej*-spørgsmål, men også fx dagbogsagtige sætninger med underforstået subjekt, som dog primært optræder i Hestenettet. I (1)–(4) viser vi eksempler fra hver overordnet kategori. Vi har markeret det relevante finitte verbum med **fed** og elementet i F med understregning.

Subjekter

- (1) a. Der **er** intet til hinder for at begynde med Trains And Boats And Planes. (KorpusDK)
- b. Den 2. ridehjem jeg har, **er** sådan en med ørebøffer på... (Hestenettet)
- c. ja ... øh vi **snakkede** jo lige om de der sociale medier ikke (LANCHART)

Objekter

- (2) a. Den tyske kaffe **har** han ikke rørt, han sidder tilbagelænet med armene over kors. (KorpusDK)
- b. Dem **solgte** jeg og købte tre nye, (...) (Hestenettet)
- c. ja hende **har** jeg snakket nok med.. . så vil jeg hh trods alt hellere ingen snakke med.. (LANCHART)

Adverbialer

- (3) a. I deres fem-fingrede kløer **havde** de køller, der var lavet af kraftige knogler. (KorpusDK)
- b. Alternativt **kan** du blande [sic] høet med med halm og blandede gulerødder i. (Hestenettet)
- c. hvis jeg er sammen med ældre mennesker **siger** jeg nok ikke så meget sgu . vel men eller om du sådan øh ja el- ja præcis i forskellige situationer .. øh mm (LANCHART)

Nul

- (4) a. ”___ **Er** det whisky ”, spurgte hun. (KorpusDK)
- b. ___ **Glæder** mig til at se billeder, og vil også meget gerne høre om fødslen, (...) (Hestenettet)
- c. Men ___ **havde** du bil ja plus en arbejdsdag på otte timer (LANCHART)

2.2. Resultater

Der er interessante ligheder og forskelle i brugen af de elementer, som kan relateres til den form for sprog, der primært findes i KorpusDK, Hestenettet og LANCHART (henholdsvis redigeret skrift, uformel skrift og talesprog), især i fordelingen af subjekter og objekter i F. Fordelingen af de fire kategorier i vores korpusundersøgelse minder i nogen grad om det mønster, man har fundet tidligere (fx Mikkelsen 2015 og Puggaard 2019), om end vi kan konstatere en interessant variation på tværs af de tre korpusser. Generelt ser vi, at subjekter er den hyppigst forekommende kategori i F i alle tre korpusser, se Tabel 1. I KorpusDK er andelen af subjekter (67,3%) i F højere end i Hestenettet (hvor de udgør 58,9%), hvor de er mere frekvente end i LANCHART (53,4%). Det er værd at bemærke, at andelen af subjekter i F i Hestenettet er på samme niveau som i KorpusDK, hvis man lægger nulusbjekterne i 1., 2. og 3. person (396, se Tabel 5) til subjektsskategorien, nemlig 66,6%.

I skriftsprogskorpusserne KorpusDK og Hestenettet er objektet markant mindre hyppigt i F (henholdsvis 3,2% og 4,8%) end i LANCHART (16,2%). Andelen af adverbialer er nogenlunde konstant på tværs af de tre korpusser (17,3%-27,7%), men ser ud til at dale i frekvens som funktion af faldende formalitetsniveau. I KorpusDK er andelen af tomme F meget lav (1,8%), mens den i Hestenettet og LANCHART er mere end seks gange højere, hvilket stemmer overens med, at der i disse korpusser bliver stillet flere *ja/nej*-spørgsmål og droppet subjekter. Dette mønster kunne indikere, at kohæsion i redigeret skrift primært etableres ved at at placere subjekter i F, mens fremflytning af objekter prioriteres i højere grad i talesprog.

	KorpusDK		Hestenettet		LANCHART		Total	
	n	%	n	%	n	%	n	%
Subjekt	673	67,3	3005	58,9	1059	53,4	4737	58,6
Adverbial	277	27,7	1126	22,0	343	17,3	1745	21,6
Objekt	32	3,2	245	4,8	321	16,2	598	7,4
Nul	18	1,8	730	14,3	261	13,2	1010	12,5
Alle	1000	100,0	5106	100,0	1984	100,0	8090	100,0

Tabel 1: Fordelingen af kategorierne *Subjekt*, *Adverbial*, *Objekt* og *Nul* i F på tværs af KorpusDK, Hestenettet og LANCHART.

Det måske mest bemærkelsesværdige mønster her er givetvis forskellen i fordelingen af subjekt og objekt i KorpusDK og LANCHART. Hvorfor skulle redigeret skriftsprog i højere grad være præget af subjekter i F, mens talesproget i højere grad placerer et objekt i den sætningsposition? For at prøve at få en bedre forståelse af, hvad der kunne ligge til grund for dette mønster, kiggede vi nærmere på, hvilke typer af subjekter og objekter, der rent faktisk står i F.

2.3. *Fordeling inden for kategorierne*

Subjekterne er underinddelt som fulde nominaler, pronominale subjekter, *hv*-subjekter og ekspletive subjekter, som dækker både foreløbige og formelle subjekter, herunder kløvnings-*det*. I vores sammenligning på tværs af de tre korpusser (se Tabel 2) ser det ud til, at placering af subjekter, der er fulde nominaler (fx *Manden med hatten løb*), er mere hyppig i KorpusDK, dvs. i et korpus, der bærer præg af at være mere redigeret skriftsprog. Herimod ser placeringen af pronominale subjekter i F (fx *Jeg løb*) ud til at være hyppigere i Hestenettet og i LANCHART, som i højere grad indeholder uredigeret/uformelt sprog, henholdsvis skrift- og talesprog. Ligesom i Puggaards (2019) undersøgelse er en forholdsvis stor andel af subjekterne et *det* i talesprogskorpusset LANCHART (faktisk hele 30,7%). Et andet bemærkelsesværdigt fund er, at der er meget få subjektsspørgeord i F, på tværs af alle tre korpusser:

	KorpusDK		Hestenettet		LANCHART		Total	
	n	%	n	%	n	%	n	%
Subjekt	381	56,6	750	25,0	102	9,6	1233	26,0
Subjekt-pron	181	26,9	1827	60,8	601	56,8	2609	55,1
Subjekt- <i>hv</i>	3	0,4	15	0,5	5	0,5	23	0,5
Subjekt-ekspl	108	16,0	413	13,7	351	33,1	872	18,4
Alle	673	100,0	3005	100,0	1059	100,0	4737	100,0

Tabel 2: *Fordelingen af fulde nominaler, pronominer, spørgeord og ekspletiver (formelle og foreløbige subjekter) inden for kategorien Subjekt i F*

Objekterne er inddelt overordnet i fulde nominaler, pronominale objekter og *hv*-objekter, se Tabel 3. Objekter, som er fulde nominalled, optræder langt oftere i F i KorpusDK, lidt mindre hyppigt i Hestenettet og for-

holdsvist sjældent i LANCHART. Derimod ser det ud til, at pronominale objekter placeres i F hyppigst i LANCHART, dvs. i talesprogskorpusset. Sidst men ikke mindst er andelen af objekter, som er spørgeord, højere i de to uformelle korpusser, dvs. både Hestenettet og LANCHART, som jo rummer en interaktionsdimension, der givetvis ikke i så høj grad bliver foldet ud i KorpusDK.

	KorpusDK		Hestenettet		LANCHART		Total	
	n	%	n	%	n	%	n	%
Objekt	17	53,1	76	31,0	18	5,6	111	18,6
Objekt-pron	10	31,3	90	36,7	203	63,2	303	50,7
Objekt- <i>hv</i>	5	15,6	79	32,2	100	31,2	184	30,8
Alle	32	100,0	245	100,0	321	100,0	598	100,0

Tabel 3: Fordelingen af fulde nominaler, pronominer og spørgeord inden for kategorien Objekt i F

Som vist i Tabel 4 nedenfor er adverbialerne inddelt i adverbier (typisk enkeltordsadverbialer som *nu* eller *imidlertid* og større adverbialled, som ikke falder under nogen af de andre kategorier), *hv*-adverbier, adverbialle præpositionsforbindelser, adverbielle ledsætninger, adjektiviske adverbialer (herunder frie prædikativer) og nominale adverbialer. Adverbialer, som også udgør en væsentlig del af de elementer, der står i F, optræder i en relativt stabil andel på tværs af de tre korpusser, men vi ser en interessant forskel i typen af adverbialer, der står i F. I LANCHART, som repræsenterer talesprog, fremrykkes enkle adverbier som *faktisk*, *så* og *nu* oftere. I modsætning hertil ser vi i KorpusDK og Hestenettet en højere andel af længere og mere komplekse adverbialer, såsom præpositionsforbindelser (fx *i morgen*) eller betingelsessætninger (fx *hvis det regner*), hvilket muligvis kunne være et resultat af, at skrift – især redigeret skrift – ikke er underlagt de samme arbejdshukommelsesbegrænsninger som talesproget. Dette mønster understøtter ligeledes ideen om, at kompleksiteten af det syntaktiske element påvirker, hvordan det placeres i F, afhængigt af både formalitet og modalitet (talt vs. skrevet sprog). Endelig, som vi har set med de andre kategorier, er der i det hele taget meget få spørgsmål, men flest i LANCHART:

	KorpusDK		Hestenettet		LANCHART		Total	
	n	%	n	%	n	%	n	%
Adverbium [-hw]	91	32,9	508	45,1	244	71,1	843	48,3
Adverbium [+hw]	5	1,8	50	4,4	34	9,9	89	5,1
Adv. præp.forbindelse	92	33,2	204	18,1	16	4,7	312	17,9
Adverbiel ledsætning	75	27,1	327	29,0	40	11,7	442	25,3
Adjektivisk adv.	5	1,8	0	0,0	0	0,0	5	0,3
Nominal adv.	9	3,2	37	3,3	9	2,6	55	3,2
Alle	277	100,0	1126	100,0	343	100,0	1746	100,0

Tabel 4: *Fordelingen af adverbialtyper i F: Ikke-spørgende adverbier ([–hw], fx heldigvis), spørgende adverbier ([+hw], fx hvornår), adverbielle præpositionsforbindelser, adverbielle ledsætninger, adjektiviske adverbialer og nominale adverbialer.*

En anden væsentlig observation er fordelingen af tomme F (Tabel 5), som er markant forskellig mellem de tre korpusser. Nul-elementerne er inddelt i *ja/nej*-spørgsmål (Nul-*ja/nej*), første-, anden- og tredjeper- sons-nul-elementer (hhv., Nul-1P, Nul-2P og Nul-3P) og imperativer (Nul-Imperativ). I KorpusDK er andelen af tomme F meget lav, mens den i Hestenettet og LANCHART er betydeligt højere. Denne forskel peger på, at der i de uformelle korpusser er en højere hyppighed af *ja/ nej*-spørgsmål eller udeladelser af subjektet, hvilket er et karakteristisk træk ved interaktive og spontane samtaler (se fx Puggaard 2018). Dette understøttes af, at talesprogskorpusset LANCHART har den højeste andel af tomme F, mens KorpusDK, der er præget af redigeret skriftsprog, har langt færre:

	KorpusDK		Hestenettet		LANCHART		Total	
	n	%	n	%	n	%	n	%
Nul- <i>ja/nej</i>	11	61,1	191	26,2	214	82,0	416	41,2
Nul-1P	0	0,0	365	50,0	12	4,6	377	37,4
Nul-2P	0	0,0	1	0,1	0	0,0	1	0,1
Nul-3P	3	16,7	30	4,1	5	1,9	38	3,8
Nul-Imperativ	4	22,2	143	19,6	30	11,5	177	17,5
Alle	18	100,0	730	100,0	261	100,0	1009	100,0

Tabel 5: *Fordelingen af tomt F: Ja/nej-spørgsmål, implicitte subjekter i 1., 2. og 3. person og imperativer.*

3. Hvad fortæller korpusdata os om F's rolle?

Ikke uventet er der på tværs af de tre korpusser flere deklarative (91,2%) end interrogative sætninger (8,8%), hvoraf 58,4% er *ja/nej*-spørgsmål, resten *hv*-spørgsmål, se tabel 6. Selvom den relative fordeling af hhv. spørgsmål versus ikke-spørgsmål (deklarativ > interrogativ) og spørgsmålstyper (*ja/nej* > adverbial-*hv* > objekt-*hv* > subjekt-*hv*) er den samme i de tre korpusser, er der også forskel: Procentvist er der mere end dobbelt så mange spørgsmål i LANCHART (17,8%) som i Hestenettet (6,6%), hvor der er næsten tre gange så mange som i KorpusDK (2,4%). Jo mindre formelt sproget er, jo flere spørgsmål.² (Omvendt er der procentvist væsentlig flere subjekt- og adverbielle *hv*-spørgsmål i KorpusDK end i de to andre korpusser, men det er baseret på meget få eksempler.)

	KorpusDK		Hestenettet		LANCHART		Total	
	n	%	n	%	n	%	n	%
Deklarativ	976	97,6%	4771	93,4%	1631	82,2%	7378	91,2%
Interrogativ	24	2,4%	335	6,6%	353	17,8%	712	8,8%
<i>Ja/nej</i> -spørgsmål	11	45,8%	191	57,0%	214	60,6%	416	58,4%
Adverbial- <i>hv</i>	5	20,8%	50	14,9%	34	9,6%	89	12,5%
Subjekt- <i>hv</i>	3	12,5%	15	4,5%	5	1,4%	23	3,2%
Objekt- <i>hv</i>	5	20,8%	79	23,6%	100	28,3%	184	25,8%

Tabel 6: Fordelingen af deklarative og interrogative sætninger, og af de forskellige typer spørgsmål.

Hansen & Heltofts (2011) påstand om, at F's primære funktion er at fungere som illokutionær ramme for sætningen, er ikke en teoretisk betragtning, der kan testes i nærværende undersøgelse. F bruges ganske givet også til at markere illokutionær ramme, men i næsten 92% af tilfældene udfyldes F med en ikke-spørgende konstituent. Konstituenten i F bruges ofte til at "sætte scenen" for sætningsindholdet, og det stemmer overens med den i litteraturen påståede tendens til at vælge konstituer, der er med til at klargøre forholdet mellem den aktuelle sætning og den forudgående diskurs. En af F's væsentligste funktioner ser dermed ud til at være at styre og strukturere diskursen, om end det er uklart, hvordan denne "iscenesættelse" stemmer overens med de informationsstrukturel-

2 For LANCHART's vedkommende er det ret oplagt med spørgsmål, eftersom det i stor stil er baseret på sociolingvistiske interviews hvor en feltarbejder spørger, og informant(er) svarer.

le overvejelser, der er blevet gjort angående sætningens første position (jf. Puggaard 2018).

Subjektet er som nævnt tidligere blevet beskrevet som det umarkerede valg til F (fx Nielsen 1975; Heltoft 1986), og vores korpusundersøgelse ser ud til at understøtte denne antagelse, i og med at det er subjektet, vi finder sætningsinitialt i præcis to tredjedele af tilfældene i KorpusDK. Det er dog værd at bemærke, at der også er en bemærkelsesværdig variation at finde: I LANCHART var F udfyldt af subjektet i lidt mere end halvdelen af tilfældene. En anden betydningsfuld forskel mellem korpusserne er, at subjekterne og objekterne i KorpusDK fortrinsvist er fulde nominaler, mens de i LANCHART primært er pronominer. I Hestenettet er de til gengæld nogenlunde ligeligt fordelt. Dette kunne måske indikere en præference for lette konstituenten i F i talesprog. Det skyldes muligvis kommunikationssituationen, eftersom diskursrummet i samtaler nok er mere indforstået, og man derfor ikke behøver at pensle alting ud i samme grad.³

En vigtig pointe i vores undersøgelse er forskellen mellem redigeret skrift, uformelt skriftsprog og talesprog. KorpusDK repræsenterer et mere formelt og redigeret skriftsprog, Hestenettet uformelt og uredigeret skriftsprog, mens LANCHART indeholder talesprog. Denne forskel er tydelig i vores resultater, og på næsten alle de parametre, vi har undersøgt, er Hestenettet faldet ud et sted mellem KorpusDK og LANCHART. Man kunne på den baggrund være tilbøjelig til at antage, at Hestenettet er et sted midt imellem skriftsprog og talesprog (jf. fx Baron 2009) eller måske indeholder en form for *netspeak* (Crystal 2011). Dette argumenterer Jensen (2014) dog imod, idet hun mener, at der i denne form for konceptualisering af en ny form for hybrid modalitet er tale om en ”kategorisammenblanding”, der ikke tager højde for den fysisk-materielle dimension af skrift og tale på den ene side og formalitets- eller intimitetsgrader på den anden. Denne forveksling mener hun sandsynligvis har basis i, at skriftsproget i flere århundreder har været karakteriseret ved at være ”formelt”, ”distanceret” og ”planlagt”, modsat talesproget (Jørgensen 1999), og at disse karakteristika er under forandring. Ifølge Jensen (2014) er en af de afgørende forskelle mellem de to modaliteter, at der er en forskydning i tid mellem afsendelse og modtagelse af kommunikation. Hun bruger diasystemets kategorier (Koch & Oesterreicher 1990) til at redegøre for, hvordan skriftsprog kan indeholde træk, som vi normalt forbinder med

3 Tak til Eva Skafte Jensen for denne pointe.

talesproget, herunder dimensioner vedrørende kommunikationssituation, socialt hierarki og sociolingvistiske variable. I forhold til begge disse parametre er Hestenettet udpræget uformelt, og det er givetvis denne manglende formalitet, der ser ud til at give os syntaktiske mønstre, der nærmer sig det talesprog, man ser i LANCHART.

Resultaterne af vores korpusundersøgelse understreger vigtigheden af at anvende korpusdata med omtanke, da mønstrene for sprogbrug kan variere betydeligt afhængigt af både form og kontekst. Det er derfor nødvendigt at tage højde for, hvilken type korpus man arbejder med, og hvordan dette kan påvirke de konklusioner, man drager om syntaktisk struktur. Næste trin i vores undersøgelse af F er at analysere sætningerne fra nærværende studie igen med henblik på at få et overblik over den informationsstrukturelle status, som elementet i F har.

4. Konklusion

Ligesom tidligere korpusundersøgelser har vi fundet, at den mest frekvente funktion i F er subjektet, og det vel at mærke på tværs af tre korpusser, der varierer i forhold til modalitet og formalitetsniveau. Dernæst kommer adverbialer, efterfulgt af objekter og tomt F. Der er tydelige forskelle i fordelingen af syntaktiske elementer som subjekt, objekt og adverbial i F afhængigt af korpus. I talesprog ser pronominaler subjekter og objekter samt adverbialer ud til at være mest frekvente i F, mens præpositionsforbindelser og adverbielle ledsætninger optræder hyppigere sætningsinitialt i skrift. Tomt F er forholdsvist sjældent i KorpusDK, men optræder ganske ofte i det uformelle skriftsprogskorpus Hestenettet og talesprogskorpuset LANCHART. Dette understreger, at korpusser ikke blot er generelle repræsentationer af sprogbrug, men at deres indhold er tæt knyttet til den kontekst, hvori de er blevet optaget. Desuden ser modalitet og formalitetsniveau ud til at have indflydelse på længden på det fremrykkede led. I skrift ser der ud til at være en tendens til mere tunge led i F end man finder i talesprog, men det vil fremtidig forskning vise.

Disse forhold tyder på, at pragmatiske og stilistiske krav i de to modaliteter påvirker den grammatiske struktur, og at sprogbrug i tale og skrift kan følge forskellige syntaktiske mønstre, og at det derfor er nødvendigt at tage højde for både formalitetsniveau og modalitet, når man analyserer korpusdata. Samlet set understreger disse fund vigtigheden af at vælge korpusser med omhu, når man arbejder med korpusdata. Korpusfrekvenser bør ikke bruges som den eneste metode til at formulere

teorier om sprog, men snarere som et supplement til mere nuancerede analyser, der også tager hensyn til de kontekstuelle forskelle mellem tale og skrift.

Om forfatterne

Anne Mette Nyvad, ph.d., lektor, Institut for Kommunikation og Kultur, Afdeling for Engelsk, Aarhus Universitet.

Katrine Rosendal Ehlers, ph.d., studielektor, Institut for Kommunikation og Kultur, Afdeling for Engelsk, Aarhus Universitet.

Ken Ramshøj Christensen, ph.d., lektor, Institut for Kommunikation og Kultur, Afdeling for Engelsk, Aarhus Universitet.

Litteratur

Baron, Naomi (2009): The Myth of Impoverished Signal: Dispelling the Spoken Language Fallacy for Emoticons in Online Communication. I: *Electronic emotion. The mediation of emotion via information and communication technologies*. Jane Vincent & Leopoldina Fortunati (red.). Bern: Peter Lang, 107–135.

Basbøll, Hans (1976): Is an integration of Diderichsen's positional analysis of Danish sentences in a transformational-generative framework feasible? I: *Papers from the Third Scandinavian Conference of Linguistics, Hanaasaari, October 1-3, 1976*. Fred Karlsson (red.). Turku: Academy of Finland, 27–38.

Borin, Lars, Markus Forsberg & Johan Roxendal (2012): Korp – the Corpus Infrastructure of Språkbanken. I: *Proceedings of LREC 2012*. Istanbul: ELRA, 474–78.

Brøcker, Karen Kiil, Magnus Glenvad Tind Hamann, Maria Jørgensen, Simon Bierring Lange, Nicholas Hedegaard Mikkelsen & Jakob Steensig (2012): Samtalesprogets grammatik. Fire fænomener og nogle metodiske overvejelser. *NyS, Nydanske Sprogstudier* 42, 10–40. <https://doi.org/10.7146/nys.v42i42.13672>

Chafe, Wallace L. (1976): Givenness, contrastiveness, definiteness, subjects, topics, and point of view. I: *Subject and topic*. Charles Li (red.). New York: Academic Press, 25–55.

Christensen, Ken Ramshøj (2005): *Interfaces. Negation – Syntax – Brain*. Ph.d.-afhandling, Aarhus Universitet.

- Christensen, Ken Ramshøj, Johannes Kizach & Anne Mette Nyvad (2013): Escape from the island: Grammaticality and (reduced) acceptability of wh-island violations in Danish. *Journal of Psycholinguistic Research* 42, 51–70. <https://doi.org/10.1007/s10936-012-9210-x>
- Christensen, Ken Ramshøj & Anne Mette Nyvad (2014): On the nature of escapable relative islands. *Nordic Journal of Linguistics* 37, 29–45. <https://doi.org/10.1017/S0332586514000055>
- Christensen, Ken Ramshøj & Anne Mette Nyvad (2022): The island is still there: Experimental evidence for the inescapability of relative clauses in English. *Studia Linguistica* 76, 684–708. <https://doi.org/10.1111/stul.12192>
- Crystal, David (2011): *Internet Linguistics – a Student Guide*. London: Routledge.
- Diderichsen, Paul (1936): Prolegomena til en metodisk dansk Syntax. I: *Forhandlinger paa det ottende nordiske Filologmøde i København den 12-14 August 1935*. Ebbe Spang-Hanssen, Viggo Brøndal & Johannes Brøndum-Nielsen (red.). København: J. H. Schultz Forlag, 41–46.
- Diderichsen, Paul (1946): *Elementær dansk grammatik*. København: Gyldendal.
- Diderichsen, Paul (1962): *Elementær dansk grammatik*, 3. udgave. København: Gyldendal.
- Diderichsen, Paul (1966): Sætningsleddene og deres stilling tredive år efter. I: *Helhed og struktur. Udvalgte sprogvidenskabelige afhandlinger*. Anders Bjerrum, Henning Spang-Hansen, Eli Fischer-Jørgensen & Knud Togeby (red.). København: G.E.C. Gads Forlag, 364–386.
- Ehlers, Katrine Rosendal (2024): *Sin* og *deres* med flertalsantecedenter i KorpusDK. *NyS, Nydanske Sprogstudier*, 1(66), 66–100. <https://doi.org/10.7146/nys.v1i66.146748>
- Engdahl, Elisabet (1997): Relative clause extraction in context. *Working Papers in Scandinavian Syntax* 60, 51–79.
- Erteschik-Shir, Nomi (2007): *Information Structure. The Syntax-Discourse Interface*. Oxford: Oxford University Press.
- Firbas, Jan (1985): Thoughts on functional sentence perspective, intonation and emotiveness. *Brno Studies in English* 16, 11–48.
- Gabelentz, Georg von der (1869): Ideen zu einer vergleichenden Syntax. Wort- und Satzstellung. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft* 6, 376–384.

- Gregersen, Frans, Marie Maegaard & Nicolai Pharao (2014): The Lanchart Corpus. I: *The Oxford Handbook of Corpus Phonology*. Jacques Durand, Ulrike Gut & Gjert Kristoffersen (red.). Oxford: Oxford University Press, 534–545. <https://doi.org/10.1093/oxford-hb/9780199571932.013.007>
- Hansen, Aage (1933): *Sætningen og dens led i moderne dansk*. København: Arnold Busck.
- Hansen, Erik & Lars Heltoft (2011): *Grammatik over det danske sprog*. Odense: Syddansk Universitetsforlag.
- Harder, Peter & Signe Poulsen (2001): Editing for speaking: First position, foregrounding and object fronting in Danish and English. *Ikonicitet og struktur*. København: Københavns Universitet, 1–22.
- Heltoft, Lars (1986): Topologi og syntaks. En revision af Paul Diderichsens sætningsskema. *Nydanske studier & Almen kommunikationsteori* 16/17, 105–130.
- Heltoft, Lars (1992): The topology of verb second and SVO languages. A study in the sign functions of word order. I: *Word order. Two studies on central issues in the syntax of Danish and French*. Michael Herslund (red.). København: Handelshøjskolens Forlag, 13–64.
- Jensen, Eva Skafte (2000): *Danske sætningsadverbialer og topologi i diakron belysning*. Ph.d.-afhandling. Københavns Universitet.
- Jensen, Eva Skafte (2014): Tale er tale, skrift er skrift. Om skriftsproget i de nye medier. *NyS, Nydanske Sprogstudier* 46, 11–38. <https://doi.org/10.7146/nys.v46i46.17523>
- Jørgensen, Keld Gall (1999): *Stilistik. Håndbog i tekstanalyse*. København: Gyldendal.
- Jørgensen, Henrik (2016): Doubling left syntactic positions in Danish. I: *Let us have articles betwixt us. Papers in historical and comparative linguistics in honour of Johanna L. Wood*. Sten Vikner, Henrik Jørgensen & Elly van Gelderen (red.). Aarhus: Aarhus Universitet, 281–298.
- Jørgensen, Henrik (2000): *Indføring i dansk syntaks*. Kompendium, Aarhus Universitet.
- Jørgensen, Henrik (2021): *Indføring i dansk syntaks*. Kompendium, Aarhus Universitet.
- Kilgarrieff, Adam, Pavel Rychlý, Pavel Smrž & David Tugwell (2004): The Sketch Engine. *Proceedings of the 11th EURALEX International Congress*, 105–16.

- Kilgariff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel (2014): The Sketch Engine: Ten Years On. *Lexicography* 1, 7–36.
- Koch, Peter & Wulf Oesterreicher (1990): *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Tübingen: Niemeyer.
- Lambrecht, Knud (1994): *Information structure and the sentence form. Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Los, Bettelou (2009): The consequences of the loss of verb-second in English: information structure and syntax in interaction. *English Language & Linguistics* 13(1), 97–125. <https://doi.org/10.1017/S1360674308002876>
- Mathesius, Vilém (1929): Zur Satzperspektive im modernen Englisch. *Archiv für das Studium der neueren Sprachen und Literaturen* 155, 202–210.
- Mikkelsen, Kristian (1911): *Dansk ordføjningslære med sproghistoriske tillæg Håndbog for viderekomne og lærere*. København: Lehmann & Stages Forlag.
- Mikkelsen, Line (2010): On what comes first in a verb-second language. Ms. UC Berkeley.
- Mikkelsen, Line (2015): VP anaphora and verb-second order in Danish. *Journal of Linguistics* 51(3), 595–643. <https://doi.org/10.1017/S0022226715000055>
- Nielsen, Niels Åge (1975): Om anvendelsen af to forskellige modeller på dansk sætningsbygning. I: *At færdes i sproget. Iagttagelser og synspunkter*. København: Dansk Sprognævn, 139–150.
- Nyvad, Anne Mette, Christiane Müller & Ken Ramshøj Christensen (2022): Too true to be good? The non-uniformity of extraction from adjunct clauses in English. *Languages* 7, 244. <https://doi.org/10.3390/languages7040244>
- Petersen, Jan Heegård (2023): En kvantitativ-distributionel syntaktisk analyse af ekstraposition til venstre. *NyS, Ny Forskning i Grammatik* 30, 201–218. <https://doi.org/10.7146/nfg.v1i30.137961>
- Petersen, Jan Heegård (2024): Folk de bliver jo dummere jo mere de får at drikke – Om nominaler som løst fundament. *Danske Talesprog* 23–24, 15–44.
- Poulsen, Mads (2005): *Sentence processing and grammaticality in functional linguistics: extraction in Danish as an example*. Ph.d.-afhandling. Københavns Universitet.

- Puggaard, Rasmus (2018): The function of the foundation field in Danish talk-in-interaction. On the intersection of information structure, interactional organization and syntactic coding. *Speciale*, Aarhus Universitet.
- Puggaard, Rasmus (2019): Flexibility of frequent clause openers in talk-in-interaction: *Det* ‘it, that’ and *så* ‘then’ in the prefield in Danish. *Nordic Journal of Linguistics* 42, 291–327. <https://doi.org/10.1017/S0332586519000088>
- Python Software Foundation (2016): *Python* (3.5.2). <https://www.python.org/downloads/release/python-352/>
- Schwartz, Bonnie D. & Sten Vikner (1989): All Verb Second Clauses are CPs. *Working Papers in Scandinavian Syntax* 43, 27–49.
- Strømberg-Derczynski, Leon m.fl. (2021): The Danish Gigaword Corpus. I: *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*. Simon Dobnik & Lilja Øvrelid (red.). Linköping: Linköping University Electronic Press, 413–421.
- Togeby, Ole (2003): *Fungerer denne sætning? Funktionel dansk sproglære*. Aarhus: Academica.
- Tomlin, Russell S. (1986): *Basic Word Order: Functional Principles*. London & Sydney: Croom Helm.
- Travis, Lisa (1984): *Parameters and Effects of Word Order Variation*. Ph.d.-afhandling, MIT.
- Vikner, Sten & Henrik Jørgensen (2017): En formel vs. en funktionel tilgang til dansk sætningsstruktur. *NyS, Nydanske Sprogstudier* 52-53, 135–168. <https://www.nys.dk/article/view/24954>
- Zwart, Jan-Wouter (1991): Clitics in Dutch: Evidence for the position of INFL. *Groninger Arbeiten zur Germanistischen Linguistik* 33, 71–92.