

Sætningskompleksitet, frekvens og acceptabilitet

Ken Ramshøj Christensen &
Anne Mette Nyvad

Abstract

Frequency undoubtedly plays a role in language learning and use. Functional approaches have argued that an individual speaker's random experience also determines their acceptability judgments. Formal approaches, on the other hand, have argued that acceptability is primarily determined by structural factors. We present the results from an investigation into the relationship between frequency, construction complexity, and acceptability in a series of different sentence types. We operationalized complexity as a product of embedding, adjunction, movement, number of moved constituents, and length of movement. The results show that acceptability and frequency are both negatively correlated with complexity and, we argue, that complexity predicts both, not vice versa.

Nøgleord

acceptabilitetsbedømmelse, syntaks, arbejdshukommelse, hyppighed

1. Indledning

Det er ukontroversielt at sprogindlæring og sprogbrug afhænger af frekvens. Hvor ofte vi udsættes for forskellige sproglige udtryk, ord og sætninger, spiller en rolle for hvor hurtigt vi lærer dem (hvis overhovedet), og om vi selv bruger dem. Det er dog stadig et stridsspørgsmål om sådanne ord- og konstruktionsfrekvenser er styrende for hvor let det er at processe (dvs. forarbejde og forstå) forskellige slags sætninger, og for hvor acceptable vi synes de er. En acceptabilitetsbedømmelse er en talers intuition (introspektion) om hvorvidt en sætning er grammatisk mulig eller ej, og i givet fald hvor acceptabel eller naturlig den er, fx på en skala fra 1 til 7

eller på en skala fra 'helt uacceptabel', over 'måske OK' til 'helt acceptabel' eller 'OK'. (En klart ugrammatisk sætning er som regel også helt uacceptabel, men grammatiske sætninger kan variere kraftigt i acceptabilitet, som vi vil se nedenfor). Sammen med naturligt forekomne eksempler er sådanne acceptabilitetsbedømmelser en af de primære datakilder indenfor moderne lingvistik (fx Fanselow m.fl. 2006, Francis 2021, Harley 2014: 10, Nyvad & Christensen 2023, Schütze 1996, Sprouse 2013, Togeby 2003: § 198, se også Myers 2017 for overblik og diskussion).

Det er velkendt at højfrekvente ord har tendens til at være kortere end lavfrekvente ord (jf. *Zipfs lov*, Zipf 1935), og de højfrekvente indholdsord genkendes og processeres hurtigere end de lavfrekvente, men det gælder ikke for funktionsord; her spiller frekvens ingen rolle (Bell m.fl. 2009). Nogle studier har også fundet en sammenhæng mellem acceptabilitet og sætnings-/konstruktionsfrekvens i skriftlige korpora (Bybee & Eddington 2006), mens andre studier har vist at der er stor variation i frekvens for sætninger med meget høj eller meget lav acceptabilitet (Featherston 2008, Kempen & Harbusch 2008, Bader & Häussler 2010). Så selvom der ser ud til at være en tendens til at konstruktionsfrekvens og acceptabilitet korrelerer positivt, er det bare ikke helt så enkelt.

Indenfor visse dele af funktionel grammatik er sproglige regler traditionelt blevet set som "emergente" og som et produkt af sprogbrugerens livslange analyse af sproglige input (Ellis 2002, Bybee & Hopper 2001, Bybee 2007)¹. Tomasello (2003: 99) formulerer det således: "Når folk gentagne gange bruger de samme, konkrete sproglige symboler til at danne ytringer til hinanden i 'lignende' situationer, kan der over tid opstå et mønster i sprogborgen, skematiseret mentalt i sprogbrugerne som en eller anden slags sproglig kategori eller konstruktion" (vores oversættelse). Det er det der også kaldes *entrenchment* (Engberg-Pedersen m.fl. 2005: 51–53, Jost & Christiansen 2016, Schmid 2016). Ideen er altså at de sætningstyper som folk finder mest acceptable, er dem de oftest hører eller har hørt. Den enkeltes grammatik afspejler således tilfældigheder og er derfor et unikt statistisk "ensemble" af sproglig erfaring der hele tiden lader sig påvirke af brug. Acceptabilitetsbedømmelser afspejler dermed de specifikke sproglige frekvenser i sprogbrugerens tilfældige erfaring og

1 Det er selvfølgelig sandsynligvis ikke alle funktionalister der mener at frekvens alene kan forklare acceptabilitet eller grammatiske regler, ligesom det ikke er alle formelle lingvister der mener at kompleksitet forklarer det hele, eller underkender frekvensens rolle. Som vi også argumenterer for her i artiklen, er billedet mere kompliceret.

er “ustabile” og “flydende” (Ellis 2002: 161–162). Men disse antagelser er ikke uproblematisk. For det første, som beskrevet ovenfor, er forholdet mellem acceptabilitet og frekvens ikke helt så enkelt. For det andet er der masser af eksperimentel evidens der viser at acceptabilitetsbedømmelser er ret stabile på tværs af forsøg og sprogbrugere (Christensen m.fl. 2013b, Sprouse & Almeida 2017, Nyvad m.fl. 2022).

En række studier har desuden påvist at grammatiske principper interagerer med og ofte annullerer frekvenseffekter, fx ved processing af lokale tvetydigheder eller valgmuligheder (Pickering m.fl. 2000, Bornkessel m.fl. 2002, White & Rawlins 2020). Studier af patienter med Brocas afasi viser at de har ekstra svært ved at forstå og producere sætninger med afledt (ikke-kanonisk) ordstilling, uafhængigt af frekvens og informationsstruktur (Bastiaanse m.fl. 2009). Det er med andre ord den øgede grammatiske kompleksitet der giver problemer, ikke den lave frekvens. Det understøttes også af at Brocas afasi bl.a. er karakteriseret ved fravær eller fejlbrug af verbale bøjningsendelser og grammatiske funktionsord, så som artikler, præpositioner, konjunktioner, etc. – de mest frekvente ord og morfemer (Damasio 1992, Caplan 2015, se også Boye m.fl. 2023).

Selvom sprogtilegnelse i en vis grad afhænger af frekvensen af forskellige ord og vendinger, er børns sprogproduktion kendetegnet ved en konsekvent afvigelse fra frekvensmønstre i det input de hører. De ignorerer diverse talefejl (der ellers forekommer ret hyppigt²) og rettelser (negativ evidens, der forekommer forholdsvist sjældent). De er kreative, og de overgeneraliserer og producerer strukturer eller konstruktioner der ikke findes i inputtet (Pinker 2004, Thornton & Crain 1994, Yang 2004). Allerede i den tidligste sprogproduktion viser børn en præference for strukturelt simple former med lavere frekvens fremfor strukturelt mere komplekse, men højfrekvente former (Anderssen & Westergaard 2010).

2. Komplexitet

Selvom frekvens åbenlyst er en vigtig faktor i sprogbrug såvel som sprog-

2 I et talesprogs-korpus med 5.000 ord (“a collection of spontaneous conversations between educated adult native speakers of British English”) fandt Garnham m.fl. (1982) ikke færre end 191 talefejl (fx substitution, anticipation, ombytning, udeladelse og sammenblanding af segmenter, stavelser eller ord). Ifølge Mehl m.fl. (2007) siger vi i gennemsnit ca. 16.000 ord i løbet af en dag på 17 timer. Hvis det er muligt at generalisere fra disse to studier, vil det svare til 611 fejl om dagen, 36 fejl i timen eller ca. en fejl hvert halve minut. Det er dog nok lidt i overkant. Om det så ’kun’ er 1-2 fejl for hvert 1.000 ord, svarende til 7-22 fejl om dagen, som Pincott (2012) hævder, er det stadigvæk ret hyppigt.

forståelse, så forudsiger frekvens ikke rigtigt hvad børn der er ved at lære sprog, og folk der har afasi som følge af en hjerneskade, rent faktisk siger og *ikke* siger. Ud fra den betragtning at kompleksitet ser ud til at spille en vigtigere rolle, forventer vi også at kompleksitet har større betydning end frekvens har, i taleres sproglige acceptabilitetsbedømmelser.

Vi vil her diskutere nogle få af resultaterne fra en større undersøgelse (Christensen & Nyvad 2023) og fokusere på forholdet mellem **konstruktionsfrekvens**, **kompleksitet** og **acceptabilitet**. Vi har operationaliseret kompleksitet som summen af de fem nedenstående faktorer:

Faktor 1: Indlejring. Når en sætnings indlejres i en anden, øger det naturligvis mængden af syntaktisk struktur (Chomsky 2001), hvilket kræver mere forarbejdning, specielt indlejrede finite sætninger (Hawkins 1994, 2004, Pallier m.fl. 2011). Folk med Brocas afasi har særligt svært ved at producere indlejrede sætninger generelt, specielt indlejrede finite sætninger (Friedmann 2001).

Faktor 2: Adjunktion eller modifikation. Når en frase, fx en nominalfrase (NP) modificeres af en anden, fx en adjektivfrase (AdjP), øges antallet af fraser (kaldet XPer) med 1 ud over de to der indgår i modifikationen: [_{NP} [_{AdjP} *elektrisk*] [_{NP} *guitar*]]. Denne forøgelse af XPer forøger den syntaktiske kompleksitet, hvilket kræver mere arbejdshukommelse, hvilket også forklarer hvorfor folk med Brocas afasi har tendens til at undgå adjunktion (Lee & Thompson 2011, Meltzer-Asscher & Thompson 2014). Fremflytning af eller fra et adverbialt led (dvs. fra en adjungeret position) er syntaktisk særligt problematisk (jf. *the Empty Category Principle (ECP)*, Haegeman 1994: 442, og *the Condition on Extraction Domain (CED)*, Huang 1982: 505), og studier har også vist at det som regel fører til reduceret acceptabilitet (Nyvad m.fl. 2022).

Faktor 3: Afstand. Flytning, navnlig ekstraktion fra en indlejret sætning (dvs. 'sætningsknuder'), koster i form af arbejdshukommelse (jf. problemerne med afledte sætninger hos folk med Brocas afasi, se ovenfor). En sådan lang flytning kræver at man holder det fremflyttede led i arbejdshukommelsen indtil det kan blive 'integreret' i basispositionen hvor dets semantiske rolle og syntaktiske funktion kan blive afklaret. Afledte (dvs. ikke-kanoniske) sætninger har som regel lavere acceptabilitet, og jo længere afstanden er mellem udgangspunkt (basisposition) og endestation (fx forfeltet), jo værre (Clifton & Frazier 1989). (På engelsk kaldes det flyttede element *the filler* og den tomme plads i udgangspunktet *the gap*.) Vi følger Hawkins (1999: 248–249, se også Christensen m.fl. 2013a) og beregner

afstanden som antallet af XPer mellem udgangspunkt og endestation, og vi medregner kun dem der er 'synlige', dvs. dem der faktisk kan ses eller høres (og ikke dem der teoretisk også må være der, såsom underforståede subjekter eller dem der kun indeholder spør efter flytning) (jf. *Path* hos Collins 1994: 56; jf. også princippet *Economy of Derivation*, der siger at den syntaktiske struktur skal holdes så simpel som mulig, og flytninger så korte som muligt, Chomsky 1995: 91).

Faktor 4: Fillers. Antallet af led der flyttes (*fillers*) er også vigtigt. Presset på arbejdshukommelsen stiger med antallet af fremflyttede led, uafhængigt af hvor langt de flytter (afstand). Hvis to led hver flytter, fx det ene to og det andet tre skridt, giver det en samlet afstand på 7, det samme som hvis et enkelt led flytter syv skridt, men det er mere krævende at holde styr på to led der flytter, end kun ét (jf. *Hvorfor₁ spurgte du -₁ hvor₂ de kommer fra -₂?* vs. *Hvorfor₁ sagde du -₁ de kommer fra Fyn?*).

Faktor 5: Udflytning. Udover prisen for selve flytningen (afstanden og antallet af fillers), koster ekstraktion, dvs. flytning ud af en indlejret sætning, specielt mange resurser. Selve ekstraktionen koster ekstra (Christensen m.fl. 2013a, 2013b, Nyvad m.fl. 2022). (Jf. også *the Phase Impenetrability Condition*, Chomsky 2001: 13, der kræver at ekstraktion fra finite sætninger "mellemlander" ved venstrekanten af den indlejrede sætning.)

Det overordnede forskningsspørgsmål var om der er en sammenhæng mellem acceptabilitet, kompleksitet og frekvens. Mere specifikt forventede vi at finde en negativ korrelation mellem acceptabilitet og kompleksitet (acceptabiliteten falder når kompleksiteten stiger). Selvom vi også forventede en vis sammenhæng mellem konstruktionsfrekvens og acceptabilitet, forventede vi dog at konstruktionsfrekvensen omvendt ville følge af kompleksiteten (frekvensen falder når kompleksiteten stiger), hvilket så igen ville forklare sammenhængen mellem frekvens og acceptabilitet.

3. En undersøgelse af konstruktionsfrekvens, kompleksitet og acceptabilitet

3.1 Stimuli

I vores undersøgelse havde vi sætninger med stigende kompleksitet, både simple sætninger (A) og en række komplekse sætninger: sætninger med hhv. en komplementsætning (B), en adverbial sætning (C) og en relativ ledsætning (D), alle \pm objektsekstraktion [\pm Eks] (type 1 vs. 2). Der var desuden to slags kontrolsætninger (*fillers*), grammatiske og ugrammatiske

(hhv. E1 og E2). Alle sætningerne blev så vidt muligt konstrueret ud fra det samme leksikalske materiale kontrolleret for leksikalisk frekvens i Korpus-DK, fra lav til høj frekvens, og kontrolleret for koherens.

(1) Kontekst: *Peter påstår at han er sur på alt og alle. Selv hunden og katten synes han også er sygt irriterende.*

A1.	Men han elsker jo alligevel hunden.	Simpel
A2.	Men <u>hunden</u> elsker han jo alligevel __.	[+Eks]
B1.	Men jeg mener jo alligevel [at han elsker hunden].	Kompl.
B2.	Men <u>hunden</u> mener jeg jo alligevel [at han elsker __].	[+Eks]
C1.	Men han elsker jo alligevel hunden [uden at ville passe den].	Adverbiel
C2.	Men <u>hunden</u> ₁ elsker han jo alligevel __ ₁ [OP ₂ uden at ville passe __ ₂].	[+Eks]
D1.	Men han elsker jo alligevel alle [OP der __ vil passe hunden].	Relativ
D2.	Men <u>hunden</u> ₁ elsker han jo alligevel alle [OP ₂ der __ ₂ vil passe __ ₁].	[+Eks]

Typerne (C)-(D) gjorde det muligt at undersøge konstruktionsfrekvens, kompleksitet og acceptabilitet hvor den ene eller den anden faktor var i bund. Adverbielle ledsætninger er ret almindelige (C1).³ Sætninger som (C2), som vi vil kalde for 'snyltehuller' (engelsk *parasitic gaps*), er grammatiske og meget komplekse, men selvom de er meget sjældne eller ikke-eksisterende i korpora, er de alligevel acceptable (Engdahl 1983, Culicover & Postal 2001). I (C2) er der et underforstået element (en såkaldt 'tom operator', OP), her er det objektet, der er fremflyttet, og derfor er der en tom plads, et 'hul', i objektpositionen. Dette 'hul' er kun muligt hvis der også er et 'hul' efter en fremflytning i matrixsætningen, og derfor 'snylter' det ene 'hul' (__₂) på det andet (__₁), jf. **Men jeg elsker jo alligevel hunden uden at ville passe __.*

Ligesom adverbielle ledsætninger, er relativsætninger (D1) komplekse (endda mere komplekse), og tillader ekstraktion (D2) (Müller & Eggers

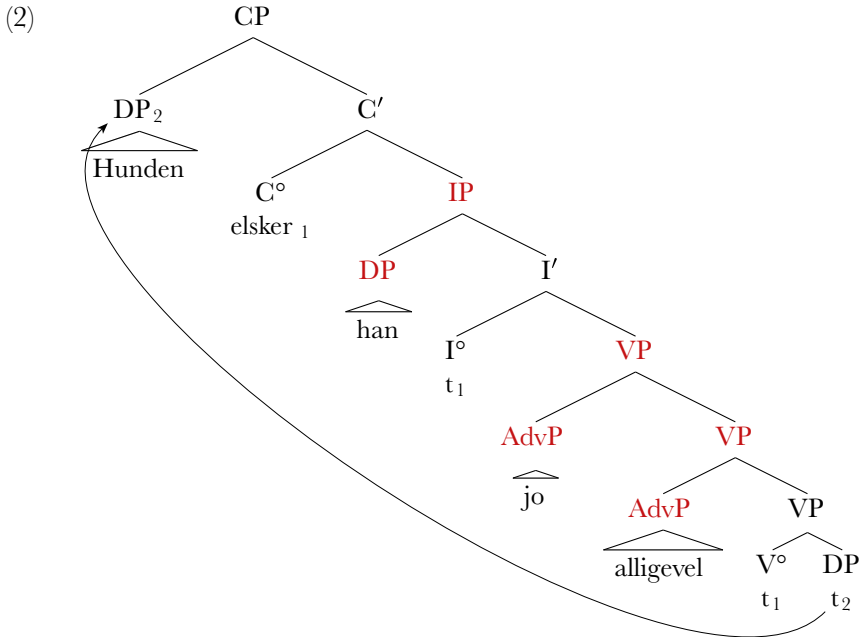
3 I C1 er *uden at ville passe dem* en infinitivsætning, eller 'infiniit ledsætning', med et underforstået subjekt, jf. Mikkelsen (1911: 567) og Kristensen (2015). Det svarer til en "skjult sætning" hos Diderichsens (1946: 56). Den samme analyse findes også i traditionel engelsk grammatik (Greenbaum 1996: 328) og er standard i generativ grammatik, hvor dette subjekt analyseres som et lydlost pronomen *PRO* (Haegeman 1994: 257, Poole 2011: 96).

2022). Ekstraktion fra relativsætninger er dog ikke-eksisterende i vores korpusdata (jf. Tabel 1), og acceptabiliteten spænder fra lav til middel (Christensen & Nyvad 2014).

	A1	A2	B1	B2	C1	C2	D1	D2
Indlejring	0	0	1	1	1	1	1	1
Adjunktion	0	0	0	0	1	1	1	1
Udflytning	0	0	0	1	0	0	0	1
Fillers	1	1	1	1	1	2	2	2
Afstand	(0) -1,0	(6) 0,4	(0) -1,0	(10) 1,3	(0) -1,0	(8) 0,9	(0) -1,0	(10) 1,3
Kompleksitet	0,0	1,4	1,0	4,3	2,0	4,9	3,0	6,3
Frekvens	(1165) 3,1	(120) 2,1	(98) 2,0	(2) 0,3	(157) 2,2	(0) 0,0	(20) 1,3	(0) 0,0

Tabel 1: Oversigt over hvordan kompleksitet beregnes ud fra de fem faktorer for hver af de otte konstruktionstyper. Afstand er z-transformeret (antallet af XPer er angivet i parentes). Kompleksitet er summen af indlejring, adjunktion, udflytning, fillers og (z-transformeret) afstand. Frekvens er $\log(10)$ -transformeret (faktiske værdier i parentes; nulfrekvensen blev sat til 1 ved \log_{10} -transformation, fordi \log_{10} af 0 giver minus uendelig).

Afstand i Tabel 1 beregnes som antallet af leksikaliserede XPer mellem flyttede led og deres basispositioner. For at undgå at afstandsfaktoren totalt styrer og forvrænger billedet (fx B2, der har en lang flytning fra en komplementsætning), har vi z-transformeret værdierne så de er normalfordelt. Det gør at effekten af en flytning og dens afstand er noget mindre og derfor kun toner den overordnede kompleksitet, og at prisen for flytningen af hovedsætningens subjekt til forfelt (dvs. *filler*-effekten i A1, B1, C1 og D1) næsten annulleres. Det præcise antal XPer i afstanden er angivet i parentes i Tabel 1, og eftersom de ikke er centrale (jf. z-transformationen) for diskussionen her, henviser vi til Christensen & Nyvad (2023) for yderligere detaljer. I (2) har vi givet et enkelt eksempel der viser hvordan afstand er beregnet for A2. Mellem det fremflyttede objekt [_{DP} *hunden*]₂ i CP-spec (forfeltet) og basispositionen (t₂ på objektspladsen) er der seks XPer med leksikalsk indhold, markeret med rødt. Fx er de øverste VPer leksikaliserede og tæller (er røde) fordi adverbialerne *jo* og *alligevel* også er det, mens den nederste VP, der kun indeholder spor efter flytninger, ikke tæller.



Vi fandt konstruktionsfrekvensen for hver konstruktion ved opslag i KorpusDK. Disse frekvenser er sandsynligvis ret konservative, da det ikke er muligt at søge efter sætningstyper i KorpusDK. Vi søgte på sætninger med pronominalt subjekt og objekt der indledes af en konjunktion som fx *og* eller *men* (alt efter konstruktionen), for at sikre at der var tale om den rette sætningstype.⁴ I Tabel 1 angives det præcise antal eksempler vi fandt, i parentes sammen med de \log_{10} -transformerede værdier, der som standard bruges i analysen. Vi henviser igen til Christensen & Nyvad (2023) for yderligere detaljer om søgestrengene mm.

3.2 Metode

Stimulussættet bestod af 24 sæt (kontekster) af sætninger svarende til (1) (plus fillers). Sætningerne blev fordelt i et Latin-square-design og i randomiseret rækkefølge på 10 lister på Google Forms. Forsøgsdeltagerne, som

4 Fx for A1 brugte vi følgende søgestreng:

[word="men|og"] [word="jeg|du|han|hun|vi|i|de" & pos="PERS"] [pos="V"] [pos="V"]
 [word="mig|dig|ham|hende|os|jer|dem" & pos="PERS"]

For A2 brugte vi:

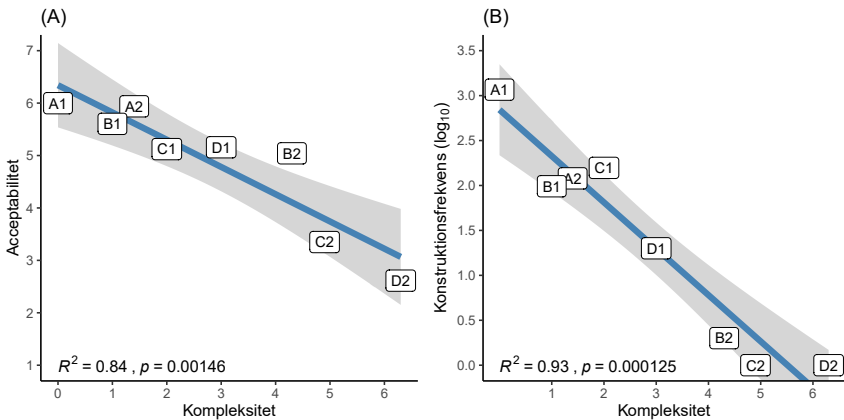
[word="men|og"] [word="mig|dig|ham|hende|os|jer|dem" & pos="PERS"] [pos="V"]
 [word="jeg|du|han|hun|vi|i|de" & pos="PERS"] [pos="V"]

blev rekrutteret på sociale medier, skulle vurdere sætningerne på en skala fra 1 ("helt uacceptabelt") til 7 ("helt acceptabelt"). De blev bedt om at basere deres bedømmelser på deres egen intuition og om at se bort fra tegnsætning.

3.3 Resultater

Som vist i Figur 1 (A), er der en statistisk signifikant negativ korrelation mellem acceptabilitet og kompleksitet ($p < 0,002$). Med andre ord så falder acceptabiliteten lineært i takt med at kompleksiteten stiger. Der er tale om en ret stærk korrelation (Pearsons $R^2 = 0,84$) der viser at kompleksiteten redegør for 84 % af variationen i acceptabilitet. Som vist i Figur 1 (B), er der også en statistisk signifikant negativ korrelation mellem frekvens og kompleksitet ($p < 0,001$): Hyppigheden falder i takt med at kompleksiteten stiger, og denne korrelation er også meget stærk ($R^2 = 0,93$). 93 % af variationen i konstruktionsfrekvenserne i vores data kan beskrives som en funktion af sætningskompleksiteten.

Der var også en signifikant og stærk positiv korrelation mellem acceptabilitet og konstruktionsfrekvens ($R^2 = 0,70$, $p < 0,01$): Jo højere konstruktionsfrekvensen er, jo mere acceptabel er sætningen.



Figur 1: (A) Gennemsnitsacceptabilitet som funktion af sætningskompleksitet. (B) Log₁₀-transformeret frekvens i KorpusDK som funktion af sætningskompleksitet. Grå skravering angiver 95 % konfidensinterval.

4. Diskussion

Indenfor visse grene af funktionel grammatik antages det ofte at en sætnings (eller konstruktions) acceptabilitet er et produkt af dens frekvens, altså hvor hyppig den er, og dermed er grammatikken også i vid udstrækning

et produkt af frekvens (Bybee & Hopper 2001). Fra formel grammatisk side argumenteres der ofte for at en sætnings frekvens primært er et produkt af sætningens strukturelle kompleksitet. I vores forsøg har vi undersøgt forholdet mellem kompleksitet (defineret som summen af indlejring, adjunktion, udflytning, fillers og afstand), acceptabilitet (på en skala fra 1 til 7) og frekvens (\log_{10} -transformeret antal af forekomster i KorpusDK).

Resultaterne viser som forventet at der er en positiv korrelation mellem kompleksitet og acceptabilitet. Jo mere strukturelt kompleks en sætning er, jo sværere er den at processere, og derfor falder acceptabiliteten også i takt med at kompleksiteten går op. Det passer fint med resultaterne fra en lang række andre undersøgelser (fx Christensen m.fl. 2013a, 2013b, Nyvad m.fl. 2014, 2022). Der er også en positiv korrelation mellem kompleksitet og konstruktionsfrekvens. Jo mere strukturelt kompleks en sætning er, jo sværere er den at processere og producere, og derfor falder frekvensen i takt med at kompleksiteten går op. Det passer også meget godt med andre undersøgelser (se introduktionen). Vi foretrækker det strukturelt simple frem for det mere komplekse. Høj kompleksitet gør at vi er mindre tilbøjelige til at bruge den pågældende sætning eller konstruktion, og det fører til lavere frekvens (Hawkins 1994, 2004, Newmeyer 2005).

Bemærk at korrelationerne mellem kompleksitet og acceptabilitet ($R^2 = 0,84$) og mellem kompleksitet og frekvens ($R^2 = 0,93$) er væsentlig større end korrelationen mellem frekvens og acceptabilitet ($R^2 = 0,70$). Kompleksitet forklarer fx 14 % mere af variation i acceptabilitet end frekvens gør. De stærke negative korrelationer mellem kompleksitet på den ene side og acceptabilitet og konstruktionsfrekvens på den anden gør desuden at det ikke er særligt overraskende at resultaterne også viser at der er en stærk positiv korrelation mellem acceptabilitet og konstruktionsfrekvens. De to falder jo begge to netop lineært i takt med at kompleksiteten stiger. Det er således mere sandsynligt at det er kompleksiteten der forklarer begge effekter.

Vi har som beskrevet, operationaliseret kompleksitet som summen af de fem faktorer, indlejring, adjunktion, udflytning, fillers og (z-transformeret) afstand, men det er sandsynligvis ikke hele historien. Ud over disse fem faktorer, er der mange andre faktorer der kan bidrage til kompleksitet, fx konstituens, argumentstruktur (valens og transitivitet), semantik, pragmatik (såsom relevans og kontekst), tempus og finithed, etc. (Det er højst sandsynligt at i hvert fald nogle af dem nok ville kunne forklare en

del af det der ikke redegøres for med vores kompleksitetsfaktorer.) Det er desuden også muligt at vores faktorer ikke blot er additive, men måske i stedet interagerer så den samlede effekt af (nogle af) disse interaktioner er mere (eller mindre) end blot summen af effekter. Det er også muligt at afstand kan beregnes på en anden måde (end antal XPer og/eller z-transformation) der muligvis ville give et mere præcist mål for effekten af en flytning. I så fald kunne man forestille sig en endnu stærkere korrelation, en der ville forklare mere end de hhv. 84 % og 93 % (værdier der dog må siges at være ret høje i forvejen). Det er også både muligt og sandsynligt at de resterende 7-16 % af acceptabiliteten forklares af noget helt andet som vi ikke har kontrolleret for i vores undersøgelse.

5. Konklusion

Man kunne være fræk og spørge: Hvad skal vi dog med al den grammatisk struktur når vi nu har noget der hedder frekvens? Det korte (frække) svar er at det jo netop er frekvensen der skal forklares, ikke omvendt. Som vi har vist, er det muligt at forklare forskellige sætningstypers acceptabilitet og deres relative frekvenser i KorpusDK vha. kompleksitet, defineret ud fra syntaktiske faktorer og arbejdshukommelse. Vores resultater er således fint kompatible med en formel grammatisk tilgang til sprogbrug og sprogforarbejdning, men ikke særligt godt med en frekvensbaseret tilgang til samme.

Om forfatterne

Anne Mette Nyvad, ph.d., lektor, 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

- Anderssen, M. & M. Westergaard (2010): Frequency and economy in the acquisition of variable word order. *Lingua* 120, 2569–2588. doi:10.1016/j.lingua.2010.06.006.
- Bader, M. & J. Häussler (2010): Toward a model of grammaticality judgments. *Journal of Linguistics* 46, 273–330. doi:10.1017/S0022226709990260.

- Bastiaanse, R., G. Bouma & W. Post (2009): Linguistic complexity and frequency in agrammatic speech production. *Brain and Language* 109, 18–28. doi:10.1016/j.bandl.2008.12.004.
- Bell, A. m.fl. (2009): Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60, 92–111. doi:10.1016/j.jml.2008.06.003.
- Bornkessel, I., M. Schlesewsky & A.D. Friederici (2002): Grammar overrides frequency: Evidence from the online processing of flexible word order. *Cognition* 85, B21–B30. doi:10.1016/S0010-0277(02)00076-8.
- Boye, K. m.fl. (2023): Agrammatism in a usage-based theory of grammatical status: Impaired combinatorics, compensatory prioritization, or both? *Journal of Neurolinguistics* 65, 101108. doi:10.1016/j.jneuroling.2022.101108.
- Bybee, J. (2007): *Frequency of Use and the Organization of Language*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780195301571.001.0001.
- Bybee, J.L. & D. Eddington (2006): A Usage-based Approach to Spanish Verbs of “Becoming”. *Language* 82, 323–355. doi:10.1353/lan.2006.0081.
- Bybee, J.L. & P.J. Hopper (red.). (2001): *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins Publishing Company. doi:10.1075/tsl.45.
- Caplan, D. (2015): Aphasia. I: *International Encyclopedia of the Social & Behavioral Sciences*. J.D. Wright (red.). Oxford: Elsevier, 818–821. doi:10.1016/B978-0-08-097086-8.54001-X.
- Chomsky, N. (1995): *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. (2001): Derivation by phase. I: *Ken Hale: A Life in Language*. M.J. Kenstowicz (red.). Cambridge, MA: MIT Press, 1–52.
- Christensen, K.R., J. Kizach & A.M. Nyvad (2013a): Escape from the Island: Grammaticality and (Reduced) Acceptability of *wh*-island Violations in Danish. *Journal of Psycholinguistic Research* 42, 51–70. doi:10.1007/s10936-012-9210-x.
- Christensen, K.R., J. Kizach & A.M. Nyvad. (2013b): The processing of syntactic islands – An fMRI study. *Journal of Neurolinguistics* 26, 239–251. doi:10.1016/j.jneuroling.2012.08.002.
- Christensen, K.R. & A.M. Nyvad (2014): On the nature of escapable relative islands. *Nordic Journal of Linguistics* 37, 29–45. doi:10.1017/S0332586514000055.

- Christensen, K.R. & A.M. Nyvad (2023): Frequency, animacy, structural complexity, and acceptability. Ms. indsendt til bedømmelse.
- Clifton, C. & L. Frazier (1989): Comprehending Sentences with Long-Distance Dependencies. I: *Linguistic Structure in Language Processing*. G.N. Carlson & M.K. Tanenhaus (red.). Dordrecht: Springer Netherlands. Bind 7, 273–317. doi:10.1007/978-94-009-2729-2_8.
- Collins, C. (1994): Economy of Derivation and the Generalized Proper Binding Condition. *Linguistic Inquiry* 25, 45–61.
- Culicover, P.W. & P.M. Postal (red.) (2001): *Parasitic gaps*. Cambridge, MA: MIT Press.
- Damasio, A.R. 1992. Aphasia. *The New England Journal of Medicine* 236, 531–539. doi:10.1056/NEJM199202203260806.
- Diderichsen, P. (1946): *Elementær dansk Grammatik*. København: Gyldendal.
- Ellis, N.C. 2002. Frequency effects in language processing. *Studies in second language acquisition* 24, 143–188. doi:10.1017/S0272263102002024.
- Engberg-Pedersen, E. m.fl. (2005): *Dansk funktionel lingvistik: en helhedsforståelse af forholdet mellem sprogstruktur, sprogbrug og kognition*. Københavns & Roskilde: Københavns Universitet, Handelshøjskolen i København, Roskilde Universitetscenter. <https://lingvistikredsen.ku.dk/pdf/df/DFL-bog.pdf>.
- Engdahl, E. (1983): Parasitic gaps. *Linguistics and Philosophy* 6, 5–34. doi:10.1007/BF00868088.
- Fanselow, G. m.fl. (red.) (2006): *Gradience in Grammar: Generative perspectives*. Oxford University Press. doi:10.1093/acprof:oso/9780199274796.001.0001.
- Featherston, S. (2008): The Decathlon Model of Empirical Syntax. I: *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. S. Kepser & M. Reis (red.). Berlin: De Gruyter Mouton, 187–208. doi:10.1515/9783110197549.187.
- Francis, E. (2021): *Gradient acceptability and linguistic theory*. New York: Oxford University Press. doi:10.1093/oso/9780192898944.001.0001.
- Friedmann, N. (2001): Agrammatism and the psychological reality of the syntactic tree. *Journal of Psycholinguistic Research* 30, 71–90. doi:10.1023/A:1005256224207.
- Garnham, A. m.fl. (1982): Slips of the tongue in the London-Lund corpus of spontaneous conversation. I: *Slips of the Tongue and Language Production*. A. Cutler (red.). Berlin: De Gruyter Mouton, 251–264. doi:10.1515/9783110828306.251.

- Greenbaum, S. (1996): *The Oxford English grammar*. Oxford: Oxford University Press.
- Haegeman, L. (1994): *Introduction to government and binding theory*. 2. udg. Oxford: Blackwell.
- Harley, T.A. (2014): *The psychology of language: from data to theory*. 4. udg. Hove: Psychology Press.
- Hawkins, J.A. (1994): *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, J.A. (1999): Processing complexity and filler-gap dependencies across grammars. *Language*. Linguistic Society of America 75, 244–285. doi:10.2307/417261.
- Hawkins, J.A. (2004): *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Huang, C.-T.J. (1982): *Logical Relations in Chinese and the Theory of Grammar*. Cambridge, MA: MIT PhD dissertation. http://pubman.mpdl.mpg.de/pubman/item/escidoc:403106:7/component/escidoc:403105/chinese_huang1982_o.pdf.
- Jost, E. & M.H. Christiansen (2016): Statistical Learning as a Domain-General Mechanism of Entrenchment. I: *Entrenchment and the Psychology of Language Learning*. H.-J. Schmid (red.). Berlin: De Gruyter Mouton, 227–244. DOI: 10.1515/9783110341423-011.
- Kempen, G. & K. Harbusch. (2008): The Relationship between Grammaticality Ratings and Corpus Frequencies: A Case Study into Word Order Variability in the Midfield of German Clauses. I: *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. S. Kepser & M. Reis (red.). Berlin: De Gruyter Mouton, 329–350. doi:10.1515/9783110197549.329.
- Kristensen, S.A. (2015): Infinitte ledsætninger. I: *15. Møde om udforskningen af dansk sprog* (MUDS 15). Y. Hansen, Inger Schoonderbeek & T.T. Hougaard (red.). Aarhus: Institut for Kommunikation og Kultur, Aarhus Universitet, 197–216. https://projekter.au.dk/fileadmin/projekter/Muds.dk/rapporter/MUDS_15.pdf.
- Lee, J. & C.K. Thompson. (2011): Real-time production of arguments and adjuncts in normal and agrammatic speakers. *Language and Cognitive Processes* 26, 985–1021. doi:10.1080/01690965.2010.496237.
- Mehl, M.R. m.fl. (2007): Are Women Really More Talkative Than Men? *Science* 317, 82–82. doi:10.1126/science.1139940.

- Meltzer-Asscher, A. & C.K. Thompson (2014): The forgotten grammatical category: Adjective use in agrammatic aphasia. *Journal of Neurolinguistics* 30, 48–68. doi:10.1016/j.jneuroling.2014.04.001.
- Mikkelsen, K. (1911): *Dansk ordføjningslære med sproghistoriske tillæg. Håndbog for viderekomne og lærere*. København: Lehman & Stages Forlag.
- Myers, J. (2017): Acceptability Judgments. *Oxford Research Encyclopedia of Linguistics*. doi:10.1093/acrefore/9780199384655.013.333.
- Müller, C. & C.U. Eggers (2022): Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages* 7. 125. doi:10.3390/languages7020125.
- Newmeyer, F.J. (2005): *Possible and probable languages: a generative perspective on linguistic typology*. New York: Oxford University Press.
- Nyvad, A.M. & K.R. Christensen (red.). (2023): *Recent Advances in Research on Island Phenomena*. Basel: MDPI. doi:10.3390/books978-3-0365-6317-6.
- Nyvad, A.M., J. Kizach & K.R. Christensen (2014): (Non-)Arguments in Long-Distance Extractions. *Journal of Psycholinguistic Research* 44, 1–13. doi:10.1007/s10936-014-9300-z.
- Nyvad, A.M., C. Müller & K.R. Christensen (2022): Too True to Be Good? The Non-Uniformity of Extraction from Adjunct Clauses in English. *Languages* 7. 244. doi:10.3390/languages7040244.
- Pallier, C., A.-D. Devauchelle & S. Dehaene (2011): Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences* 108, 2522–2527. doi:10.1073/pnas.1018711108.
- Pickering, M.J., M.J. Traxler & M.W. Crocker (2000): Ambiguity Resolution in Sentence Processing: Evidence against Frequency-Based Accounts. *Journal of Memory and Language* 43, 447–475. doi:10.1006/jmla.2000.2708.
- Pincott, J.E. (2012): Slips of the Tongue. *Psychology Today*. <https://www.psychologytoday.com/intl/articles/201203/slips-the-tongue>. (17 marts, 2023).
- Pinker, S. (2004): Clarifying the logical problem of language acquisition. *Journal of Child Language* 31, 949–953. doi:10.1017/S0305000904006439.
- Poole, G. (2011): *Syntactic theory*. 2. udg. Houndmills: Palgrave Macmillan.
- Schmid, H.-J. (2016): 1. A Framework for Understanding Linguistic Entrenchment and Its Psychological Foundations. I: *Entrenchment and the Psychology of Language Learning*. H.-J. Schmid (red.). Berlin: De Gruyter Mouton, 9–36. doi:10.1515/9783110341423-002.

- Schmid, H.-J. (2016): A Framework for Understanding Linguistic Entrenchment and Its Psychological Foundations: *1. A Framework for Understanding Linguistic Entrenchment and Its Psychological Foundations*, 9–36. De Gruyter Mouton. DOI: 10.1515/9783110341423-002.
- Schütze, C.T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Sprouse, J. (2013): Acceptability Judgments. Oxford University Press. doi:10.1093/obo/9780199772810-0097.
- Sprouse, J. & D. Almeida (2017): Setting the empirical record straight: Acceptability judgments appear to be reliable, robust, and replicable. *Behavioral and Brain Sciences* 40. e311. doi:10.1017/S0140525X17000590.
- Thornton, R. & S. Crain (1994): Successful Cyclic Movement. *Language Acquisition Studies in Generative Grammar*. T. Hoekstra & B.D. Schwartz (red.). Amsterdam: John Benjamins Publishing Company, 215–252. doi:10.1075/lald.8.11tho.
- Togeb, O. (2003): *Fungerer denne sætning?: funktionel dansk sproglære*. København: Gad.
- Tomasello, M. (2003): *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- White, A.S. & K. Rawlins (2020): Frequency, acceptability, and selection: A case study of clause-embedding. *Glossa: a journal of general linguistics* 5. 105. doi:10.5334/gjgl.1001.
- Yang, C.D. (2004): Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8, 451–456. doi:10.1016/j.tics.2004.08.006.
- Zipf, G.K. (1935): *The psycho-biology of language: An introduction to dynamic philology*. Boston: Houghton Mifflin Company.