

Henning Bergenholtz

Empiriske metoder i sprogvidenskabelig forskning

Abstract

I presume an interdependent connection between the empiric basis chosen for linguistic research and the possible methods used by the linguist. Considering that the use of self-made (often unnatural) sentences and the reliability of linguistic tests are dubious methods, the analysis of texts will be the best method in all difficult cases.

1. Indledning*

Dette tema er uløseligt forbundet med et andet tema: den empiriske basis i sprogvidenskabelig forskning. Alt efter hvilken basis man går ud fra, har man et større eller mindre antal af muligheder i valget af metode eller metoder. De fleste af mine eksempler vil være tyske, ligeledes vil henvisningerne til sekundærlitteratur bære præg af, at jeg i så mange år har bevæget mig i det tyske forskningsmiljø. På én måde har jeg gjort et dårligt bytte ved at flytte fra det tyske til det danske forskningsmiljø: I Tyskland kunne jeg optræde som hård kritiker, der gik imod strømmen, imod den stærkt introspektivt orienterede forskningstens inden for den tyske lingvistik. I Danmark er det i meget højere grad en selvfølge, at man arbejder ud fra en bred empirisk basis.

2. Empiri

I denne sammenhæng forstår jeg ved empiri den viden om sprog, som lingvisten vinder under sin beskæftigelse med de genstande, som han anser for relevante for dannelsen af teorier, og de beskrivelser, som kan foretages på grundlag af de opstillede teorier. Den viden om sprog, som lingvisten kommer til, kalder jeg den empiriske basis. Denne basis indeholder data og en bedømmelse af de data, som lingvisten begrundes sine teoretiske antagelser og de

*Jeg vil takke Inger Bergenholtz for hendes kritik af det første udkast til denne artikel.

opstillede regler med.

Jeg vil her ikke gå ind på forholdet mellem induktiv og deduktiv fremgangsmåde. I praksis vælger jeg en primært induktiv arbejdsmetode som fx beskrevet af Toulmin og van de Velde.

Derimod må brugen af termen *regel* nødvendigvis gøres til genstand for omtale. Jeg anvender denne terminus i meget bred forstand: Enhver beskrivelse af et sprog forstås som en regel, også hvis denne beskrivelse ikke bliver givet som en explicit regel. Ved brugen af dette brede regelbegreb er det vigtigt at skelne mellem to grundlæggende regeltyper:

1. regler om det, der forekommer i et sprog
2. regler om det, der er umuligt i et sprog

Den sidste type af regler har været særlig populær de sidste 30 år. De er dog sårbare, da de skal gælde uden undtagelse. Når fx Pütz 1975,75 om pronomet *es* siger, at *wissen* hører til de verber, som altid optræder uden korrelatet *es*, er denne regel meget let at falsificere, sml.: *Ich wußte es ja, daß Sie als Literat ...* (Mann 1967, I, 212). Enten laver Th. Mann og mange andre, der skriver og taler tysk, grammatiske fejl ved brugen af *wissen*, eller også er reglen forkert. Jeg foretrækker den sidste mulighed.

Ved regler af den første type, regler om det, som forekommer, drejer det sig derimod ikke nødvendigvis om 100%-regler, det drejer sig om en beskrivelse af de forekomster, lingvisten indtil nu har registreret. Disse beskrivelser af forekomster kan fx i sprogundervisningen forenkles på en sådan måde, at man på et vist trin i tyskundervisningen lærer, at appositionen altid skal have kasuskongruens, når sprogeleven bruger en sådan konstruktion. Det drejer sig hermed om en *synteseregler*, som skal gælde i didaktisk sammenhæng og som ikke uden videre kan bruges som *analyseregel*. Derimod skal analyseregler kunne bruges til beskrivelse af alle kendte tekster med det pågældende (del-) sprog. Hvis en bestemt konstruktion i en konkret tekst ikke er forudset i analysereglerne, kan man enten tale om en fejl eller om en særlig sjældnen type.

3. Den empiriske basis

Hvordan er forskellige beskrivelser blevet til? Er nogle mere rigtige end andre? Det mener jeg nok, de er. I hvert tilfælde findes der en række regler, som ikke svarer til sprogbrogsbrugen i et bestemt sprog. Hertil nogle eksempler:

Admoni 1976,220 bringer bl.a. følgende regel: "Die unpersönlichen Sätze, die Witterung, Naturzustände und Zeitstufen bezeich-

nen, erfordern solches *es* immer unabhängig von der Form ihres Prädikats." (De upersonlige sætninger, som giver udsagn om vejret, naturlilstande og om tiden, kræver altid et *es* uafhængigt af prædikats form.). Denne regel passer ganske vist til *Jetzt ist es Winter*, men ikke til *Heute ist Donnerstag*, *Heute ist Nebel* eller *Jetzt ist Ebbe*. Altså ikke for alle sætninger, som giver udsagn om vejret, om naturlilstande eller om tiden.

Endnu mere klart misvisende er følgende beskrivelse i tabelform i Collins, en bog med tyske verbetabeller udkommet i første oplag i 1984: *gebären* (føde): *er gebiert, er gebar, er wird gebären, er hat geboren*. Det er vel tænkelig sprogbrug om fx en gletscher, der kælder. Men imperativerne er dog noget ejendommelige: *gebier!*, *gebären wir!*, *gebärt!*, *gebären Sie!*. Også konjunktiverne for 1. person singularis og 1. og 3. person pluralis er sælsomme, selv om disse former opføres i næsten alle grammatikker: *ich gebäre, wir gebären, sie gebären*. I overensstemmelse med reglen i de samme tyske grammatikker bruges disse former aldrig, da formerne for konjunktiv II (bortset fra *sein*) altid skal vælges i de tre tilfælde for at undgå sammenfald med indikativ.

Som et sidste eksempel fra Collins kan man prøve at forestille sig, hvornår man kan bruge *sprießen* (at vokse (om planter)) i imperativ og i alle personer i ental og flertal.

sprießen [strong, *sein*]

PRESENT PARTICIPLE sprießend	PAST PARTICIPLE gesprossen
PRESENT INDICATIVE ich sprieße du sprießest er sprießt wir sprießen ihr sprießet sie sprießen	PRESENT SUBJUNCTIVE ich sprieße du sprießest er sprieße wir sprießen ihr sprießet sie sprießen
IMPERFECT INDICATIVE ich sproß du sprosstest er sproßt wir sprossen ihr sproßt sie sprossen	IMPERFECT SUBJUNCTIVE ich sprösse du sprössest er sprösse wir sprossen ihr sprösset sie sprossen

FUTURE INDICATIVE

ich werde sprießen
du wirst sprießen
er wird sprießen
wir werden sprießen
ihr werdet sprießen
sie werden sprießen

CONDITIONAL

ich würde sprießen
du würdest sprießen
er würde sprießen
wir würden sprießen
ihr würdet sprießen
sie würden sprießen

PERFECT INDICATIVE

ich bin gesprossen
du bist gesprossen
er ist gesprossen
wir sind gesprossen
ihr seid gesprossen
sie sind gesprossen

PLUPERFECT SUBJUNCTIVE

ich wäre gesprossen
du wär(e)st gesprossen
er wäre gesprossen
wir wären gesprossen
ihr wär(e)t gesprossen
sie wären gesprossen

IMPERATIVE: sprieß(e)! sprießen wir! sprießt! sprießen Sie!

Vi vender tilbage til temaet: På hvilket grundlag er de forskellige korrekte og ikke korrekte beskrivelser kommet i stand?

Der findes mindst fire måder at skaffe sig data og bedømmelser af data på:

- (1) Lingvisten går ud fra foreliggende beskrivelser, som er foretaget af andre lingvister.
- (2) Lingvisten bruger sin egen sprogkompetens som grundlag for beskrivelsen: den introspektive fremgangsmåde.
- (3) Lingvisten spørger andre, som behersker det pågældende sprog: den multispektive fremgangsmåde.
- (4) Lingvisten undersøger sprogbrogen i konkrete tekster.

4. Introspektion

Normalt vælges ikke en af disse fire grundmuligheder, men en kombination, hvor en af mulighederne spiller en større rolle end de andre. Brugen af lingvistens egne sproglige kompetens er til enhver tid uundværlig ved analyse af de sproglige data, hvorimod det er tvivlsomt, om den alene er i stand til at fremskaffe de data, der skal bedømmes.

Når jeg i mange artikler (bl.a. i Bergenholz 1983a, 1983b, 1984, Bergenholz/Mugdan 1989) stærkt har kritiseret den introspektive fremgangsmåde, er det at forstå som en reaktion på den metode, som betragter introspektionen som den sikreste og bedste måde at opnå indsigt i sproglige fænomener. Schnelle 1970,53 har sagt det på følgende måde:

"Mir scheint, daß es den generativen Linguisten (...) gelungen ist, nachzuweisen, daß (...) die eigentlich bedeutsamen sprachwissenschaftlichen Aussagen auf heuristischen und intuitiven Wegen gefunden werden. Das praktische Vorgehen dieser Sprachwissenschaftler zeigt, daß sie die Heuristik der Beispiele und Gegenbeispiele an Hand von Urteilen als empirisches Auswahlverfahren meisterhaft beherrschen." (Efter min mening er det lykkedes for den generative skole at påvise, at de virkelig betydningsfulde sprogvindskabelige udsagn er opnået ved hjælp af en heuristisk og intuitiv fremgangsmåde. I praksis har disse lingvister vist, at de på overbevisende måde behersker heuristikken i empirisk udvælgelse og indordning af eksempler og modeksempler.)

Siden den tid er der flere gange blevet rokket ved troen på den "overbevisende brug af intuitionen". Fx har Levelt 1972 i en test bedt 24 lingvister om at betegne 14 eksempletsætninger fra kendte lingvistiske arbejder som grammatiske eller ugrammatiske. Ved sammenligning mellem bedømmelsen i den oprindelige artikel og i testen var det meget påfaldende, at af de sætninger, der oprindeligt blev indordnet som grammatisk rigtige, blev 30% af et flertal betegnet som ugrammatiske, samtidig blev kun 15% af de sætninger, som oprindeligt blev betegnet som ugrammatiske, af de 24 lingvister også indordnet på samme måde. Selv har jeg også lavet en lille undersøgelse af den slags. Et par eksempler, som jeg tog fra Pütz 1982, har jeg forelagt for 80 informanter (tykstuderende i Essen) og bedt dem indordne sætningerne som grammatiske eller ugrammatiske. I følgende tilfælde (ud af otte) havde flertallet af de adspurgte en anden bedømmelse end lingvisten:

	Pütz	Informanter
1. <i>Ich liebe, daß Blumen Schmuck sind.</i>	+	-
2. <i>Ich liebe es, daß Blumen Schmuck sind.</i>	+	-
3. <i>Ich bewundere, daß es hier so warm ist.</i>	+	-
4. <i>Es ist zu umständlich, daß das Auto in Serie geht.</i>	+	-

Man kan hertil ikke spørge, hvem der har "ret". Mere præcist ville det være at spørge:

- (a) (som beskrivelse af det, der forekommer i tekster) Findes der sætninger af den pågældende type? Hvis ja, i hvilke typer af tekster?
- (b) (som regel til brug ved produktion af tekster (som svarer til en lærers bedømmelse af elevproducerede tekster)) Er sætninger af denne type, henholdsvis regler, som fører til sådanne sætninger, af en sådan art, at de skal høre til det, der skal læres?

Endelig vil jeg nævne KVL, som er en af de få ordbøger, der i forordet explicit går ind på introspektionens problematik ved udarbejdelsen af en ordbog. I forordet til denne ordbog bliver det fremhævet, at eksemplerne er af en sådan art, at man kan anbefale deres brug for udlændinge. Endvidere bliver det oplyst, at disse eksempler og deres beskrivelse er opnået ved den individuelle kompetens, af og til efter diskussion med andre medarbejdere. Det indrømmes dog, at nogle sætninger er blevet "stilistisk uheldige" (*stilistisch recht unglücklich*). Et særligt "godt" eksempel er: *Er behauptet es, mich zu lieben*. Det er ikke bare uheldigt, men meget lidt anbefalelsesværdigt for udlændinge; i forbindelse med *behaupten* optræder der ikke noget foregribende objekt *es*.

Noget helt andet er, at de selvkonstruerede eksempler i denne ordbog for en stor del er af typen: *Moderen vasker op / Faderen er ude ved bilen / Moderen lægger børnene i seng / Far læser avis*. Jeg kalder normalt sådanne sætninger "lingvistpoesi", ofte meget afsørgende for dem, der danner sætningerne. I øvrigt er eksemplerne i denne ordbog ganske vist stereotyp, men dog ikke helt så kunstige som de "vittige" eksempler i generative arbejder.

Hovedargumentet er, at lingvistens sprogformmelse, sprogkompetens, ikke er i stand til både at fremskaffe data og at bedømme data. Det går ganske vist i meget enkle tilfælde, men i alle interessante sammenhænge har sprogkompetensen nok at gøre med at indordne og bedømme foreliggende data. Eller sagt på en anden måde: Der bør skelnes mellem sprogkompetens¹, som sætter et menneske i stand til at tale og forstå et sprog, og sprogkompetens², som betegner et menneskes evne til at bedømme en sproglig konstruktion. Og mellem disse to former for kompetence findes der væsentlige forskelle.

5. Multispektion

Min kritik af den introspektive fremgangsmåde kan i vid udstrækning overføres til tests med informanter. Det er i denne sammenhæng interessant, at det særligt er syntaktikere, som har brugt tests med "native speakers". Men også inden for leksikologien og leksikografien er der i de sidste 20-30 år gentagne gange blevet henvist til sådanne tests. Det er dog påfaldende, at der som regel kun udspørges et lille antal (ca. 20 er normalt, 100 er snarere undtagelsen, som fx i den amerikanske ordbog AHD).

Hvorfor man i sprogsørgsmål kan klare sig med et meget mindre antal informanter, end det er normalt ved opinionsundersøgelser, diskuteres aldrig, ligesom spørgsmålenes art og udformning, deres indvirkning på svarene osv. aldrig gøres til genstand for alvorlige overvejelser. Jeg har selv i tidens løb gennemført en lang række tests, som til dels har en tvivlsom værdi, hvorfor jeg heller ikke har offentliggjort ret meget om dem. Det første eksempel er en test med 80 tyskstudierende på universitetet i Essen, som blev foretaget i forbindelse med mine undersøgelser af kasuskongruens. De to første sætninger er taget fra "der Spiegel", de øvrige udvalgt fra kendte grammatikker.

Kreuzen Sie bitte für die folgenden Sätze an,

- (a) ob Sie die jeweiligen Sätze als *gemeinsprachlich korrekt* ansehen
- (b) ob Sie den Satz als *grammatisch falsch* einstufen
- (c) ob Sie sich nicht ohne weiteres für (a) oder (b) entscheiden können.

	normal	falsch	weiß nicht
(1) Danach fiel der Kurs der Playboy-Aktien - <u>einem Papier</u> , daß mit einem nackten "Sharemate" verziert ist.	33	35	12
(2) Er sieht sich als Vollstrecker des politischen Erbes von Kemal Atatürk, <u>des legendären Gründers</u> der Republik Türkei.	54	17	9
(3) Das schadet dem Ansehen meines Onkels als <u>ehrbarer Kaufmann</u> .	33	36	11
(3) Herr Plessing erwies sich als <u>ausgezeichneter Verwaltungsbeamter</u> .	78	1	1
(4) Er erwies sich <u>als ausgezeichneten Verwaltungsbeamten</u> .	1	76	3
(5) Am Montag, <u>dem 23. November 1981</u> , erhielt ich endlich Antwort auf meinen Brief.	61	16	3
(6) Am Montag, <u>den 10. Juni 1960</u> lief das Schiff vom Stapel.	32	39	9

- (7) Eine Horde randalierender Halbstarker hatte das Geschäft vollkommen verwüstet.
- (8) Vergiß nicht, ein Dutzend frische Eier zu kaufen.
- (9) Er vergaß nicht, ihr einen Strauß duftende Rosen mitzubringen.
- (10) Bei einem Glas gutem Wein werden wir uns wohl einigen können.
- (11) Der Preis eines Zenfner Weizen ist sprunghaft gestiegen.
- (12) Der Preis eines Pfund gekochten Schinkens ist sprunghaft gestiegen.
- (13) Der Weggang des Rektor Meyers hat sehr viel Unruhe in die Schule gebracht.
- (14) Die Verehrung gilt dem idealen Schwunge, dem edlen Stolz, dem feurigen Vaterlandsgefühl des Dichters, Vorzüge, neben welchen wir seine Mängel billig zurechtlegen sollen.
- (15) Man ging durch die Katharinenpforte, ein ehemaliges Tor und seit Erweiterung der Stadt ein offener Durchgang.
- (16) Es ist mehr als fraglich, ob die Bewertung dieses Romans als ein Zeitdokument zutreffend ist.
- (17) Die Verhaftung des Generals als des eigentlichen Drahtziehers ließ zu lange auf sich warten.

Der er ganske vist en del tendenser, der bliver tydelige, men de svarer hverken til kendte regler i grammatikker eller til sprobrugen i skrevne tekster. Fx gælder både (6) og (7) i så godt som alle grammatikker for korrekte, men halvdelen af de adspurgte anser (7) for at være ukorrekt. I skriftsproget forekommer type (7) i særlig høj grad i handelsprog, hvor type (6) er mindre brugt. Derimod findes type (6) stort set altid i dødsannoncer, hvor man næsten aldrig finder type (7) (i 500 undersøgte dødsannoncer var der to eksempler af type (7) mod 498 med type (6)).

I det næste eksempel var 100 tyskstuderende på universiteterne i Essen og Bochum testpersoner. Sætningerne blev konstrueret af Bjørne Ulvestad over belæg fra romaner. I denne tekst skal der ikke tages stilling til den problematiske skalering "normal-falsk-ved ikke", men kun til eksempler med en maksimal mængde af es-pronomener. Tallene angiver mængden af informanter, der anså det pågældende (e)s for ikke at være muligt:

Durchstreichen Sie jedes (e)s, das Ihrer Meinung nach nicht zulässig ist.

1. "Ach Chef, ich hab's (78) vergessen, es (63) Ihnen zu sagen, daß Karin heute nicht kommt." "Sie hat also angerufen. Welchen Grund gibt sie an?" "Sie sagt es (94), sie sei krank." (oder: "Sie sagt es (92), daß sie krank sei.")
2. Die einfachste und ungefährlichste Ausrede ist es (52), du sagst es (86) dem Chef, du bist krank. Wer kann es (48) beweisen, daß du es nicht bist?
3. Warum sagst du es (35) erst heute, daß du (es) (84) vor hast, morgen zu heiraten?
4. Ob Egon es ernst mit mir meint? Das ist schwer zu sagen; aber zumindest sagt er es (57) immer wieder, daß er mich liebt.
5. Wer kein Fieber hat, darf es (96) nicht sagen, er ist krank.
6. Frau Adriatica kennt sich gut aus in der Geschichte des Spätmittelalters, und sie sagt es (51) jedem, der es (3) hören wil, daß eine Mylendonk Mitte des 13. Jahrhunderts Äbtissin eines Stiftes zu Bonn am Rheine war.
7. Warum bist du so erstaunt? Georg hat es (57) dir doch gleich gesagt, daß es so kommen würde.
8. Es ist leicht es (97) zu sagen, daß man krank ist, aber in dieser Firma muß man es (64) auch beweisen, daß man krank ist.
9. Natürlich glaube ich es (72), daß Winnetou kommt, er hat es (28) ja selbst gesagt, daß er spätestens morgen in der Frühe hier sein wird.
10. Siehst du, du sagst es (43) selbst, daß du gelogen hast!

Disse tests siger efter min mening mere om, hvad informanter mener under indtryk af deres skolegang og kulturelle baggrund end om den virkelige sprogbrug. Det er fx påfaldende, at nogle informanter afviser næsten alle sætninger, mens andre informanter kun sjældent anser en sætning for grammatisk forkert. Hvad der mangler, er ikke flere tests, men teoretiske overvejelser, som de fx har fundet sted inden for de politiske specialvidenskaber. Hvordan skal informanter udvælges i forhold til emnet? Hvor stort skal dette antal være? Vil de 1.-2.000 informanter, der som oftest bruges ved politiske meningsmålinger, være nok? Hvordan skal spørgsmålene stilles? Hvor mange spørgsmål kan en informant besvare uden at blive for stærkt påvirket af spørgsmålets art? Disse og andre problemer bliver næsten aldrig diskuteret ved fremlæggelsen af sproglige tests. I stedet for møder man udtrykket "hvad også tests med informanter viser" (fx Pütz 1975,10). Men når der samtidig ikke siges noget om spørgsmålets art og om de adspurgtes antal, alder, køn og erhverv, må udtalelsen betragtes som et immuniseringsforsøg, som en garanti uden indhold.

Det er ligeledes ikke overbevisende, at Folsom 1966,62 kan

hævde, at han ikke har fundet en eneste informant, som har godkendt en sætning med *lernen + es + objektsætning* for at være rigtig. Jeg har spurgt 100 informanter, tyskstuderende fra universiteterne i Essen og Bochum; af dem anså 80 følgende sætning for at være normal: *Ich habe es noch nicht gelernt, mit dem Ding umzugehen*. Jeg vil med andre ord ikke afvise, at informanter kan give interessante oplysninger, men det er ikke så meget henvisninger til den virkelige sprogbrug (hvis det ikke drejer sig om enkle fænomener), men snarere meninger om, hvordan den burde være. Det er meget muligt, at informanterne selv bruger visse konstruktioner i bestemte kontekster, men at de i den sterile laboratorisituation, hvor de skal bedømme kontekstløse sætninger, ikke er i stand til at forestille sig den nødvendige og måske mulige kontekst.

6. Brugen af den lingvistiske litteratur...

...er en selvfølge. Men hvem kan være sikker på, at han har taget hensyn til den nødvendige og vigtige litteratur? Det er lidt informativt at læse artikler, som kun refererer til de allerede kendte udsagn. Mere væsentligt i denne sammenhæng er, at foreliggende lingvistisk litteratur også bør bedømmes ud fra kriteriet: Hvilken empirisk basis har den? Hvordan er den empiriske fremgangsmåde? Alt efter svarene på disse to spørgsmål ofrer jeg personligt mere eller mindre tid på læsningen af lingvistiske arbejder til brug for et bestemt tema. Konkret vil det sige, at jeg normalt læser de arbejder særligt grundigt, som også refererer en tekstanalyse.

7. Undersøgelser af tekster

Der findes ganske vist lingvister, som fx Itkonen 1976, som anses undersøgelser af tekster for at være "en overflødig ceremoni", men sådanne lingvister danner uden tvivl undtagelsen og ikke reglen.

Det er ikke noget nyt, at lingvister samler belæg til brug for deres sprogvidenskabelige arbejder, men når det ikke drejer sig om sprog med en meget lille og helt afsluttet (skrevet) tekstmængde (altså om et "dødt" sprog), bliver der normalt ikke aflagt regnskab for den præcise fremgangsmåde ved udvalget af belæg og for det meste heller ikke om det opnåede antal belæg. Det er selvfølgelig let ud fra de muligheder, computeren byder os i dag, at nøjes med at have hånlige bemærkninger tilovers for de store anstrengelser, som lingvister indtil den allernyeste tid har måttet gøre sig for at

fremskaffe belæg. Jeg vil derfor gøre opmærksom på, at min kritik af "belægsamlinger", som de normalt er blevet sat sammen og brugt, helt svarer til det, som fx Behaghel skriver i forordet til sin syntaks fra 1923.

Traditionelt er der overvejende blevet brugt tekster med de "gode" forfattere, først i nyere tid har man også taget aviser, fagbøger og trivallitteratur med. Det talte sprog har så godt som ingen rolle spillet. Fx siger Duden-GWB i forordet, at man har gennemset millioner af belæg fra det nuværende tyske sprog, på omslaget bliver tallet to millioner belæg nævnt, i alt bliver der opført ca. 650 tekster, som belæggene er taget fra. Tilsyneladende har de, der har udvalgt belæggene, i særlig grad skullet søge eksempler med usædvanlig og/eller sjælden ordbrug (Duden-GWB I, 19). Instruktionerne til Duden-medarbejderne er aldrig blevet offentliggjort; jeg går ud fra, at instruktionen om belægsøgning i OED svarer til den, som man har brugt i Dudenredaktionen: "Make a quotation for every word that strikes you as rare, obsolete, old-fashioned, new, peculiar, or used in a peculiar way. (...) Make as many quotations as you can for ordinary words (...) (OED xv). Uden tvivl er en samling af belæg nyttig og i alle tilfælde en mere sikker basis end tests med informant-er eller introspektion. Men én ting kan en samling af belæg fra sporadisk gennemsete tekster ikke: Den kan ikke sige noget om et fænomens hyppighed og slet ikke, om fænomenet ikke findes i alle tekster eller ikke i tekster af en bestemt type. Og der kan heller ikke siges noget om hyppighedsforholdet mellem konkurrerende former. Belægsamlinger kan være særdeles nyttige, i mange tilfælde uundværlige, nemlig som supplement til korpusundersøgelser. Men analyse af et tekstkorpus er stadig den mest sikre udgangsposition.

8. Tekstkorpus

Ved et tekstkorpus forstår jeg en endelig mængde af tekster eller sammenhængende dele af tekster, som er blevet produceret i en kommunikativ proces. Med denne definition ville en samling af egne eksempler ikke kunne kaldes et korpus, hvad fx Greenbaum 1984, 194 gør. Denne definition indebærer også, at der forudsættes en undergrænse for størrelsen af den sammenhængende del af en tekst, som kan udgøre en egen tekst i et korpus. Denne grænse kan dog ikke fastsættes kvantitativt, fx mindst 500 ord; der menes hermed, at et sammenhængende afsnit i en tekst ikke bør opdeles på en sådan måde, at udsnittet bliver uforståeligt. Der er her tale om en upræcis bestemmelse; i praksis vil det sige, at i tvivlstilfælde må en

lidt større deltekst foretrækkes fremfor en lidt mindre som tekst i et korpus.

Det er i de fleste tilfælde hurtigere og sikrere at arbejde med maskinlæsbare korpora, hvortil man let kan fremstille hyppighedslist-ter og konkordans. Men også et ikke-maskinlæsbart korpus kan være brugbart og ved nogle problemstillinger udgøre den eneste mulighed.

Et tekstkorpus indeholder en endelig mængde af tekster - det gælder også, hvis man principielt forudsætter, at korpus skal kunne udvides. På et bestemt tidspunkt er omfanget endeligt. Det indebærer, at en endelig mængde af tekster ikke vil kunne indeholde alle muligheder for brugen af sproget, at vigtige muligheder måske ikke forekommer. I så fald vil en udvidelse af tekstgrundlaget være påkrævet, man vil også kunne inddrage samlinger af belæg eller søge belæg i foreliggende beskrivelser af sproget. Fx havde Joachim Mugdan og jeg under forberedelsen af prøveartikler med bl.a. *jagen* ingen eksempler med typen *Damit kannst du mich jagen* (Bliv mig fra livet med det...), en vending, vi selvfølgelig kendte. Vi tog derfor de nødvendige eksempler fra de store kendte ordbøger.

9. Brugen af korpora i praksis: tre eksempler

Ud fra Limas-tekstamligen og Freiburger-Korpus med alt i alt ca. 3,6 mio. tekstord blev alle eksempler med angst-ord (dvs. *Angst*, *Furcht*, *Entsetzen*, *fürchten*, *ängstlich*, *panisch* osv.) søgt og trykt ud. Med *Angst* blev der fundet 928 belæg, som ligesom stablerne af de andre ord blev gennemset, sorteret og beskrevet under hensyntagen til forskellige problemstillinger, fx:

- Hvilke verber findes der i sætninger, hvor et angstsubstantiv er grammatisk objekt?
- Hvilke verber findes der i sætninger, hvor et angstsubstantiv er grammatisk subjekt?
- Hvilke attributive adjektiver findes der til angstsubstantivet?

I alt blev der stillet ni spørgsmål af denne art. Principielt skulle en anden lingvist, der arbejdede med det samme materiale og brugte samme metode, opnå samme svar. Hvormed ikke er sagt, at to lingvister også nødvendigvis ville beskrive hver deres lister fuldstændig ens. Denne fremgangsmåde er dog meget langsom og dag vil jeg bruge en halvautomatisk fremgangsmåde som fx beskrevet i Bergenholtz 1985. Med udgangspunkt i en sætningskonkordans ville klassificeringen af eksempler og kollokationer dels

kunne foretages af maskinen, mens lingvisten skal bekræfte (eller afvise) forslagene, dels vil lingvisten under gennemsyn af konkordansen kunne rubricere belæggene. Alle former for statistikker vil blive foretaget af maskinen, den konkrete beskrivelse af lingvisten.

I en anden undersøgelse kunne et maskinlæsbart korpus ikke bruges. Det drejer sig om mine undersøgelser af kasuskongruensen ved brugen af appositioner. Her blev i stedet brugt et læsekorpus, suppleret med et lille hørerekorpus. Alt i alt indeholdt det brugte korpus ca. 2 mio tekstord. Metoden var traditionel: Ud fra en bestemt definition af apposition læste jeg bøger, ugeblade og aviser. Hele tiden med en blyant i hånden understregedes appositioner. Undervejs gjorde jeg mig notater om interessante eksempler. Når jeg var færdig med en tekst, indførte jeg kongruenstyperne i et skema, der fik følgende slutresultat:

← N (oKm)	941	G ← N	27	N	← D	0
← G (oKm)	85	D ← N	4	G	← D	8
← D (oKm)	164	A ← N	23	A	← D	0
← A (oKm)	208	G ← aN	3	N	← aD	0
N ← N	24	D ← aN	1	D	← aD	3
G ← G	232	A ← aN	5	A	← aD	0
D ← D	108	pG ← N	2	pG	← aD	0
A ← A	558	pD ← N	42	pA	← aD	0
N ← aN	125	pA ← N	11	pG	← D	1
G ← aG	13	pG ← aN	0	pA	← D	6
D ← aD	15	pD ← aN	0	Adv	← D	1
A ← aA	113	pA ← aN	2	N	← A	0
pG ← G	2	N ← G	0	G	← A	1
pD ← D	535	D ← G	0	G	← A	0
pA ← A	188	A ← G	2	pG	← A	0
pG ← aG	0	pD ← G	4	pD	← A	1
pD ← aD	36	pA ← G	0	N	← aA	0
pA ← aA	15	N ← aG	2	G	← aA	0
Satzappos.	18	D ← aG	0	D	← aA	1
		A ← aG	0	pG	← aA	0
		pD ← aG	0	pD	← aA	1
		pA ← aG	0	pA	← aA	1

Forkortelserne står for kasus, ordklasser eller for enkelte ord: N = nominativ, G = genitiv, D = dativ, A = akkusativ, oKm = kasuskongruens ikke morfologisk udtrykt (fx ved navne), aN = *als/wie* + nominativ, aG = *als/wie* + genitiv, pG = præposition + genitiv, Satzapp. = sætningsapposition (apposition til en hel sætning), Adv = adverb, N ← N = en apposition i nominativ med en appositionsbasis i nominativ, N ← aN = apposition i nominativ og en appositionsbasis med

als eller *wie* + nominativ.

Undersøgelsen endte med et angreb på nyere grammatikere, som mente, at ikke-kongruerende appositioner skulle være blevet normalt i moderne tysk. Det kunne påvises, at det ikke er tilfældet, men at man i visse teksttyper snarere skal forvente inkongruens end i andre (konkret i fagtekster, og ikke i litterære tekster). Interessant i vores sammenhæng er størrelsen af det brugte korpus: Hvorfor to mio tekstord? De to mio var ikke på forhånd fastlagt, men efterhånden som denne mængde nåedes, kunne det forudses, at det samlede resultat ikke mere ændredes væsentligt.

I en tredje undersøgelse, som blev foretaget sammen med Bjarne Ulvestad drejede det sig om pronomenet *es*, når det foregriber objektet ved bestemte verber. Ved hjælp af computeren kunne alle eksempler med *es* søges og trykkes ud; i Limas-korpus var det 7096 belæg, hvoraf der var 261 belæg med *es*, som korrelerede med en objektsætning, dvs. at kun 3,7% af *es*-belæggene var med et *es*, som interesserede i den pågældende sammenhæng, resten var eksempler med anden brug af *es*.

Af de 261 belæg var der 90 forskellige verber og verbekombinationer. 261 belæg er meget lidt, men mere fandtes ikke i et tekstkorpus med en million tekstord. Endnu mere problematisk var det, at alle eksemplerne uden *es* som korrelat til en objektsætning ikke kan søges ved hjælp af *es*. Selv om der søgtes ved hjælp af de nu fundne verber og kombinationer med verber, ville vi ikke finde alle eksempler uden *es*-korrelat, da vi ikke kendte alle verber, der kunne optræde med *es*. Vores liste med sådanne verber indeholdt til slut langt mere end 1000 verber og kombinationer med verber. Desuden havde vi den tese, at brugen af *es*-sammenhæng kunne være et fænomen, som kun ville kunne forklares som del af en større tekstsammenhæng. Altså måtte vi læse hele tekster. Det foregik på samme måde, som beskrevet ovenfor, ved gennemlæsning, understregning og sammentæning på skemaer. Alt i alt læste vi ca. 40 mio tekstord, i praksis blev den største tekstmængde læst af Bjarne Ulvestad, der godt kunne være standset ved ca. 20 mio tekstord, som bagefter viste sig at være grænsen for virkelig nye indsigter. Herefter begyndte antallet af nye verber med et pronomen *es* som korrelat til en objektsætning at tage af ved gennemlæsning af nye tekster.

I ingen af de tre undersøgelser kan resultaterne hævdes at gælde uden indskrænkninger for moderne tysk, hverken for alle andre tyske tekster eller for endnu ikke producerede tekster. Man kan kalde

metoden **minimalistisk** (efter van de Velde 1979,25), da resultaterne i snæver forstand kun gælder for de undersøgte tekster. En sådan minimalisme står i modsætning til en maksimalistisk samling og forklaring af flest mulige "ægte" eller selvkonstruerede eksemplertyper, hvad der fx i generativ-transformationelle arbejder ofte fører til en diskussion ud fra usædvanlige konstruktioner. Den her foretrukne minimalistiske fremgangsmåde vælger derimod at søge et maksimalt udbytte af en **fuldstændig analyse** af en vis mængde af hele tekster eller større tekstenheder; kun derved gives mulighed for nøjagtige udsagn om forholdet mellem det regelmæssige og det uregelmæssige og også om den manglende forekomst af principielt mulige udtryk i den undersøgte tekstmængde. I et videre arbejds-skridt vil de opnåede analyseregler kunne bruges som udgangspunkt for synteseregler, hvis yderligere tekstundersøgelser ikke giver væsentligt nye indsigter.

10. Konklusioner

Til slut skal i hovedpunkter opridses, hvad der må være konklusionen på denne artikels metodiske holdning:

Introspektion, brugen af egen sprogkompetens, er en nødvendig del af bedømmelsen af det foreliggende empiriske materiale. Jeg mener dog, at man tiltror muligheden for brugen af egen kompetens for meget, hvis den både bliver brugt til dannelsen af det empiriske materiale og til bedømmelsen heraf. Uden tvivl kan man øve brugen af sin sprogkompetens; det er uden tvivl også sådan, at nogle lingvister er hurtigere og bedre til at finde gode eksempler end andre. Men efter min erfaring er denne evne særlig vigtig i diskussioner og ikke så vigtig i den normale forskningsproces. Hvis man i en længere forskningsproces til stadighed danner og bedømmer egne konstruktioner, vil ens egen sprogkompetence også ændre sig. Efter et vist tidsrum vil en sådan lingvist tendere til at bedømme alle selvlavede eksempler ud fra teorien og ikke ændre teorien ud fra eksemplerne.

Hvad angår tests med informanter, så gennemføres der masser af tests, om hvilke man kan sige, at de viser interessante resultater. Men det er såre uklart, hvad de viser resultater om. De siger snarere noget om, hvordan bestemte sproglige fænomener menes at burde være, i mindre grad siger resultaterne noget om den virkelige sprogbrug. Der mangler her overvejelser

- om måden at spørge på

- om udvalget af informanter (Der bliver næsten altid brugt studenter som informanter)
- om antallet af informanter (Er tyve virkelig nok? Hvor mange er nok? Mon det ikke er sådan, at man til forskellig slags spørgsmål behøver et forskelligt antal informanter?)

Itkonen kalder *tekstundersøgelser* for "en unødigt ceremoni", jeg vil kalde dem "tidskrævende og anstrengende", men samtidig påpege, at de efter min mening er nødvendige, hvis man skal nå til resultater, som går ud over en sammenfatning af allerede opnåede resultater eller en nyvurdering af allerede forelagte beskrivelser.

Litteratur

- Admoni, Wladimir (1976): Es handelt sich um es. Zur gegenwärtigen Lage in der Grammatiktheorie, i: *Wirkendes Wort* 26, 1976, 219-227.
- AHD = *The American Heritage Dictionary of the English Language*. Boston. Atlanta. Dallas. Geneva. Illinois. Hopewell, New Jersey. Palo Alto 1976.
- Behagel, Otto (1923): *Deutsche Syntax. Eine geschichtliche Darstellung. Bd. I: Die Wortklassen und Wortformen. A: Nomen. Pronomen*. Heidelberg: Winter 1923.
- Bergenholtz, Henning (1983a): Grammatik im Wörterbuch: Zur Terminologie und zur empirischen Basis, i: *Kopenhagener Beiträge zur germanistischen Linguistik* 21, 1983, 70-72.
- Bergenholtz, Henning (1983b): Sind systematische Textuntersuchungen eine unnötige Zeremonie?, i: *Zeitschrift für germanistische Linguistik* 11, 1983, 211-213.
- Bergenholtz, Henning (1984): Grammatik im Wörterbuch: Syntax, i: *Studien zur neuhochdeutschen Lexikographie V*. Hrsg. von Herbert Ernst Wiegand. Heidelberg/New York: Olms 1984, 1-46 (=Germanistische Linguistik 3-6/84).
- Bergenholtz, Henning (1985): DFG-Forschungsantrag "Grammatisches Wörterbuch". Essen 1985.
- Bergenholtz, Henning/Mugdan, Joachim (1989): Formen und Probleme der Datenerhebung II: Gegenwartsbezogene synchronische Wörterbücher, i: *Wörterbücher. Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Hrsg. von Franz Josef Hausmann o.a. Berlin/New York: de Gruyter (i trykken).
- Collins (1984) = *Collins Gem German Verb Tables With Nominal Declensions*. London/Glasgow: Collins 1984.
- Duden-GWB = *Duden. Das große Wörterbuch der deutschen Sprache in sechs Bänden*. Hrsg. und bearb. vom Wissenschaftlichen Rat und den Mitarbeitern der Dudenredaktion unter Leitung von Günther Drosdowski. Mannheim. Wien. Zürich. Bd. 1: A-Ci 1976, Bd. 2: Ci-F 1976, Bd. 3: G-Kal 1977, Bd.

- 4: Kam-N 1978, Bd. 5: O-So 1980, Bd. 6: Sp-Z 1981.
- Folsom, Marvin H. (1966): *The Syntax of Substantive and Non-finite Satellites to the Finite Verb in German*. The Hague/Paris: Mouton 1966.
- Greenbaum, Sidney (1984): Corpus Analysis and Elicitation Tests, i: *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research* ed. by Jan Aarts and Willem Meijs. Amsterdam: Rodopi 1984, 193-201.
- Itkonen, Esa (1976): Was für eine Wissenschaft ist die Linguistik eigentlich?, i: *Wissenschaftstheorie der Linguistik*. Hrsg. von Dieter Wunderlich. Kronberg: Athenäum 1976, 56-76.
- KVL = Ulrich Engel/Helmut Schumacher: *Kleines Valenzlexikon deutscher Verben* unter Mitarbeit von Joachim Ballweg et al. Tübingen: Narr 1978 (1. Aufl. 1976).
- Levelt, W. J. L. (1972): Some psychological aspects of linguistic data, i: *Linguistische Berichte* 17, 1972, 18-30.
- Mann, Thomas (1967): *Der Zauberberg*. Frankfurt/M: Fischer 1967.
- OED = *The Oxford English Dictionary*. Ed. by James A. Murray, Henry Bradley, W. A. Craigie, C. T. Onions. Volume I. A-B. Repr. Oxford 1970. [first publ. 1933].
- Pütz, Heribert (1975): *Über die Syntax der Pronominalform es im modernen Deutsch*. Tübingen: Narr 1975.
- Pütz, Heribert (1982): Objektprädikate, i: *Satzglieder im Deutschen. Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung*. Hrsg. von Werner Abraham. Tübingen: Narr 1982, 331-367.
- Schnelle, Helmut (1970): Theorie und Empirie in der Sprachwissenschaft, i: *Theorie und Empirie in der Sprachforschung*. Hrsg. von Herbert Pilch und Helmut Richter. Basel/München/Paris/New York: Karger 1970, 51-65
- Toulmin, Stephen (1975): *Der Gebrauch von Argumenten*. Aus dem Englischen übersetzt von Ulrich Beck. Kronberg/Ts.: Scriptor 1975.
- Veldt, Roger G. van de (1974): *Zur Theorie der linguistischen Forschung. Ein Beitrag zur Grundlagenproblematik der Linguistik*. München: Hueber 1974.
- Veldt, Roger G. van de (1979): Probleme der linguistischen Theoriebildung einer empirischen Textwissenschaft, i: *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora*, hrsg. von Henning Bergenholtz/Burkhard Schaefer. Königstein/Ts.: Scriptor 1979, 10-27.