

Packat och oklart

Informationspackningens roll för begriplighet i text

Lars Melin, Stockholms universitet

Inledning

På 1980-talet skedde en omorientering i begriplighetsforskningen: från ytspråk till perspektiv, retorisk strategi etc. Denna svängning blev så definitiv att den klassiska begriplighetsforskningen, som baseras på ytspråkliga egenskaper i texter, i Sverige dog ut, trots att det faktiskt finns frågor kvar.

Men kvar levde förstås den psykolingvistiska forskningen där laboratoriemässiga mätningar av försökspersoners reaktioner på språkliga stimuli hade andra syften, främst att klarlägga hur mänskliga perceptionsstrategier tillämpades på språkligt material. Sådan forskning baseras på mycket korta stimuli, oftast enstaka ord. Det finns dock en poäng med att bedriva begriplighetsforskning om ytspråk och på längre, sammanhängande text. Alla textstrategier slutar i ytspråk, och så gott som alla meningsfulla meddelanden är så långa att de har både kontext och struktur.

En fråga som ständigt aktualiseras i samband med ytspråkssvårigheter, t ex Klarspraksgruppens arbete (<http://justitie.regeringen.se/klarsprak/>), är komplexitet. Problemet som denna uppsats angriper är om komplexiteten verkligen är det största problemet. Innehållspackning är en stark motkandidat.

Syfte

Syftet är att experimentellt undersöka i vilken mån sådana språkdrag som traditionellt förknippas med komplexitet respektive innehållspackning bidrar till att öka lästid och minska minnesbehållning och textuppskattning vid läsning av sammanhängande text.

Hypotesen, grundad på data från Lindblom (2000), är att informationspackning i högre grad än komplexitet bidrar till att sänka läshastighet, minnesbehållning och textuppskattning.

Bakgrund

Att språklig komplexitet bidrar till att göra texter mindre tillgängliga har vid olika tider förklarats på olika sätt. Den klassiska psykolingvistikens såg antalet processteg mellan djup- och ytstruktur som den avgörande faktorn, och detta var drivkraften bakom de första experimentserierna i modern psykolingvistik (t ex Fodor & Garret 1967). Platzack (1974) såg problemet i att vissa transformationer gjorde djupstrukturen mindre genomskinlig för läsaren. Vänsterförgreningar och andra komplexa strukturer har setts som mer svårprocessade (Yngve 1960; Wearing 1970). Korttidsminnet får svårare att hantera information som inte är strukturellt självförklarande (Waern 1978), och innehållsmängden (mätt i antal propositioner) är avgörande för processmodan (mätt som lästid) (Kintsch 1974).

Bortser vi från teorier som är förankrade i tidiga generativa teorier har vi två huvudkandidater som erbjuder sig att förklara svårt språk:

1. Svårt språk är komplext och påfrestar därför korttidsminnet genom att anstränga vår parsingkapacitet.
2. Svårt språk är innehållspackat och kräver mentalt processande av stora propositionsmängder på kort tid.

Båda förklaringarna verkar intuitivt rimliga, speciellt som komplext språk ofta är innehållspackat och innehållspackat språk ofta är komplext. Men hur blir det om man separerar: vilket anstränger oss mest, komplexitet eller innehållspackning?

Förklaringen med innehållspackning brukar knytas till propositionsmängden i en text. Eftersom det inte är alldeles självklart hur man räknar propositioner har Lindblom (2000) prövat ett annat grepp. Han har formaliserat innehållsmängd som hög nominalkvot och därför laborerat med att höja och sänka nominalkvoten i en text och därefter kontrollera lästiderna. NQ är ju ett etablerat mått på just innehållspackning i text (Einarsson 1978). Lindbloms experimenttexter på 2,4 resp 0,6 i NQ genererade en skillnad i medellästighet på 40 procent, men eftersom skillnaden i textlängd var 44 procent gick de två läsargrupperna i mål ungefär samtidigt. Detta ser ut som en perfekt verifiering av Kintsch' (1974) teori att propositionsmängden är avgörande för läsningen.

Men packningsteorin duger inte för att förklara åtskilliga av de klassiska begriplighetsexperiment som Platzack (1974) och Telemann (1973) gjorde med till exempel ordföljd och meningslängd.

Problem som dessa bör ses mot bakgrund av de teorier om parsing som kommit i stället för dem som låg till grund för de klassiska experimenten. En vanlig väg är att utgå från s k garden path-meningar för att se vad som är

läsarens defaultväg genom syntaxen. Dessa experiment brukar visa att late closure (defaulten är att påbörjad fras fortsätter) och minimal attachment (förutsätt så få noder som möjligt i strukturen!) är de bärande principerna för att guida en läsare genom meningen (Frazier & Rayner 1982). Dessa ger stöd för att komplexitet är försvårande eftersom en mening som bryter minimal attachmentregeln är resultat av någon form av inbäddning.

Likaså verkar late closure-regeln i samma riktning, det vill säga mot mer inbäddning i längre fraser.

Å andra sidan är det naturligt att betrakta sådana meningar som bryter mot parsingprinciperna som kortvarianter av andra naturligare alternativ, det vill säga informationspackade meningar.

En konkurrerande parsingteori menar att huvudarbetet utgår från den information som lexikon ger om ordens benägenhet att ingå i syntaktiska strukturer, t ex verbens valens (Clifton, Frazier & Randall 1983). Denna teori är helt neutral till komplexitet versus informationspackning.

En tredje teori fokuserar spår (*traces*) av antecedenter. Spår är belastande för korttidsminnet, men olika spår har visat sig vara olika lätt att fylla (Frazier & Clifton, 1989). Denna teori stöder både komplexitetsteorin eftersom flyttningar brukar bidra till komplexitet och packningsteorin eftersom förekomsten av *gaps* är informationskoncentrerande.

Metod

Min undersökning baseras på texter som manipulerats för att innehålla många drag som traditionellt förknippas med komplexitet respektive innehållspackning. Sådana texter, jämte kontrolltext som saknar överdriven komplexitet och innehållspackning, har presenterats för försökspersoner som läser texterna, i ett försök på papper och i ett annat försök på datorskärm. I båda försöken har lästiden mätts, och i papperstestet har dessutom minnesbehållning och attityd till texten kontrollerats.

1. Texter

Underlaget för experimenttexterna utgörs av en DN-debattartikel om farorna med snusande. Denna text har bearbetats i tre versioner, en ”normaltext”, en komplicerad och en innehållspackad. Alla tre texterna omfattar en tätskriven A4-sida, ca 450 ord. För en språklig analys se tabell 1.

Därtill har de tre versionerna klippts ihop till textfiler som tre försökspersoner läser med ett self-paced-reading program (Svartholm 1999). Här har normal, komplex och packad varvats. För varvningsprinciper se Material nedan.

2. Försökspersoner

Vid samtliga försök har försökspersonerna utgjorts av studenter vid Nordiska språk i Stockholm. Studenterna har läst grundkurs i Svenska eller Praktisk svenska. Testen gjordes i anslutning till undervisning.

3. Testmetoder

Vid papperstestet har försökspersonerna fått ett papper med text på ena sidan och vissa frågor på andra sidan. De har instruerats att mäta sin egen lästid, oftast med tidtagarfunktionen på mobiltelefon. De har också instruerats att läsa texten som om de var intresserade av ämnet, alltså varken jäkta igenom texten eller läsa överdrivet långsamt för att försöka memorera detaljer. Läsmålen (Gunnarsson 1982) har alltså satts tämligen lågt.

Self paced reading går ut på att försökspersonerna själva matar fram texten i önskad fart. Datorn mäter tiden mellan frammatningarna och kan alltså ge detaljerad information om tidsåtgången för varje rad och för texten som helhet.

Antalet försökspersoner har begränsats till tre. Det innebär att det är fråga om ett roterat test där tre fp läser tre textversioner som redigerats efter tre principer. Varje faktor (person, text, textvariant) kan alltså alltid vägas mot alla andra. Det är möjligt att mäta relativa differenser, men omöjligt att signifikansberäkna dessa.

Minnesbehållningen har kontrollerats med kontrollfrågor på detaljer (minnessökning). Attitydfrågorna fungerar som maskering mellan läsningen och kontrollfrågorna.

Material

I detta avsnitt beskrivs texterna och testmetoderna.

1. Papperstexterna

När texterna redigerats har jag strävat efter att uppfylla följande kriterier:

Komplex text: långa meningar, vänstertyngd och inskott av underordnad sats eller långa adverbial mellan primära satsled, det vill säga språkdrag vars sammantagna effekt gör att texten blir längre och får en mer svårparsad struktur.

Packad text: hög NQ, långa ord, nominaliseringar, långa nominalfraser, sammansatta substantiv, det vill säga språkdrag vars sammantagna effekt blir att texten blir kortare och redundansen minskar.

Normal text: låga värden av det som karakteriserar komplicerad och packad.

Redigeringen av texterna har lett till de data som redovisas i tabell 1.

2. Self paced readingtexterna

Texterna har varvats enligt följande:

version 1: normal, packad, komplicerad
version 2: packad, komplicerad, normal
version 3: komplicerad, normal, packad

Varje modul består av mellan 50 och 60 ord, så self-paced-texterna är väsentligt kortare än papperläsningstexterna. Samma språkliga nyckeltal som för papperstexterna redovisas i tabell 2.

Tabell 1: Nyckeltal för texterna i papperstestet

	normal	packad	komplicerad	Δp mot k i %
Grunddata				
ord	455	329	481	46
tecken	2253	1919	2388	24
antal men	25	21	22	4
komplexitet				
fund	2,12	2,38	7,64	221
inskott	0	7	19	171
vänster *	47	49	45	-8
ord/men	18,2	15,7	21,8	38,8
packning				
Nplängd **	2,19	2,68	1,83	-31
nom	8	23	11	-52
ordl	4,9	5,8	4,9	-15
nq	0,6	3,06	0,85	-71
sms***	26	43	25	-42

* vänster är antalet bestämningar till vänster om huvudord

** medelnominalfraslängd

*** antal sammansatta substantiv

3. Blanketten

Vid papperstestet fick försökspersonerna mäta sin egen lästid, och markera attityd till texten i denna tabell:

Texten var lätt att läsa	nej	nja	ja	absolut
Texten ingav förtroende	nej	nja	ja	absolut
Texten var lätt att förstå	nej	nja	ja	absolut
Texten gav ett positivt intryck	nej	nja	ja	absolut

Minnesretentionen kontrollerades med ett par detaljfrågor.

Tabell 2: Nyckeltal för self-paced-reading-modulerna.

	N1	P1	K1	N2	P2	K2	N3	P3	K3
grunddata									
ord	57	36	53	50	36	51	46	40	68
tecken	316	235	308	279	215	295	256	245	321
meningar	4	3	3	3	2	2	2	2	2
komplexitet									
fund	1,5	2,6	5,3	1	1	5,5	2	1,5	22,5
vänster	3	3	7	2	3	11	3	4	1
ord/men	14,5	12	17,7	16,7	18	25,5	23	20	34
inskott	0	0	6	0	0	9	0	0	6
packning									
ordl	5,5	6,5	5,8	5,6	6	5,8	5,6	6,1	4,7
nom	1	1	1	3	2	2	2	3	2
Nplängd	2	2,9	2,2	2,7	4	4	1,7	2,7	1,5
NQ	0,6	1,8	,7	2	4,5	1,7	1,4	2,9	,8
sms	2	5	2	4	3	5	4	7	3

Resultat

Resultaten av de två testen redovisas i tabell 3 och 4.

1. Papperstestet

Tabell 3. Redovisning av lästid, attityd och minnesbehållning. N = 66

	tid	lätt	förtr	förstå	intryck	minne
normal	18,17	3,05	2,62	3,05	2,15	3
packad	15,74*	2,73*	2,73	3,09	1,77***	2,73*
kompl	17,18	3,15	2,95	3,10	2,20	2,82

Den informationspackade versionen skiljer sig på flera punkter från den komplicerade. Den största differensen gäller positivt intryck, men också de för läsbarheten avgörande faktorerna, tidsåtgång och minnesretention var signifikanta i t-test. Att den komplicerade texten gav ett positivt intryck och ingav förtroende är en förväntad effekt. Detta är ofta fallet med texter med en hög grad av formalitet (Melin & Lange, 2000).

2. Self paced readingtestet

Detta är ett roterat test där tre personer läst tre texter i tre versioner. Det är stora avvikelser mellan läsarnas läshastighet, 42 procent mellan snabbaste och långsammaste och 10 procent mellan den långsammaste och mellanprestationen. Alla läsarna har stora variationer i läshastigheten. Det är den snabbaste läsaren som har högst variationsindex, 26 procent. De övriga har 13 resp 17 procent.

Tabell 4. Redovisning av tid i sekunder för self paced readingtestet

	normal	pack	komp	Δp mot k	Δn mot p
test 1	19,15	17,90	19,75	10,34	6,98
test 2	21,57	18,25	18,62	2,03	18,17
test 3	25,13	23,72	25,53	7,61	5,93
medel	21,95	19,96	21,30	6,73	9,97

Skillnaderna mellan versionerna framgår av tabell 4 ovan, där det framgår att den packade texten alltid tar längst tid att läsa. Om samtliga normalrader, packningsrader och komplikationsrader adderas blir medellästiden 25,5, 19,7 resp 20,9 sekunder. Alltså läses den packade texten alltid med lägre hastighet oberoende av hur man mäter. Differensen mot normaltexten är 29,5 procent, medan differensen mellan packad och komplicerad är obetydlig.

Radanalys

Self paced readingprogrammet ger inte bara information om den totala lästiden utan också möjlighet att se lästiderna på radnivå. Här redovisas de rader som avviker mer än en standardavvikelse uppåt från medellästiden.

1. Avvikande rader i packad text

långt NP: en bild av snuset som en välsignelse för folket”

långt ord: tobakstillverkare

långt NP en bild av snuset som ett oskadligt rökalternativ

långt NP och långt ord: nikotinläkemedlens förtjänster

2. Rader i komplex text

verbkedja: anges kunna enkelt förklara

långt NP: till människor i andra nationer i västvärlden

långt NP: ett oskadligt alternativ till rökning

omfattning: ett med nikotintuggummi, plåster och spray ...

... jämförbart rökavvänjningsmedel

I den normala texten avviker ingen rad mer än en standardavvikelse uppåt.

Om en läsare tvingas sänka sin läshastighet mer än en standardavvikelse tycks det alltid vara fråga om långa enheter (ord eller nominalfraser), och dessa fraser kan ibland vara komplicerade, men de är alltid informationspackade.

Resultatdiskussion

Papperstestet visar att packad text tar längre tid att läsa, uppskattas lägre och bevaras sämre i läsarnas minne än komplex text som i sin tur visar sämre värden än normaltext.

Self paced readingtestet bekräftar relationerna för lästid. Det senare ger också underlag för att på radnivå konstatera att när den komplicerade texten tar lång tid att läsa kan tidsutdräkten lika lätt förklaras med packnings- som komplexitetsdrag.

Komplexitet gör uppenbarligen en text svårare att läsa, men informationspackning är ännu mer läsarovänlig.

Refrenser

- Einarsson, Jan (1978) *Talad och skriven svenska*. Sociolingvistiska studier. Lundastudier i nordisk språkvetenskap. Serie C, Studier i tillämpad nordisk språkvetenskap.
- Frazier, L & Clifton, C & Randall, J (1983) Filling gaps: decision principles and structure in sentence comprehension. *Cognition*, 13.
- Frazier, L & Rayner, K (1982) Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, 14.
- Gunnarsson, Britt-Louise (1982) *Lagtexters begriplighet*. En språkfunktionell studie av medbestämmandelagen. Liber.
- Kintsch, W (1974) *The representation of Meaning in Memory*. Hillsdale NY.
- Lindblom, Magnus (2000) Nominalkvotens betydelse. En studie i hur informationspackningen påverkar läsarens attityd till texten. B-uppsats i Information i text och bild. Nordiska språk. Stockholms universitet.
- Melin, Lars & Lange, Sven (2000) *Att analysera text*. Studentlitteratur.
- Platzack, Christer (1974) *Språket och läsbarheten*. Gleerups.
- Svartholm, Gotfrid (1998), *Deka*. Ett datorprogram för self-paced-reading.
- Teleman, U, Forkner, M, Mitsell, H, Molin, B & Wikman, I (1973) *Åtta texter ur Dagens Nyheter. En läsbarhetsstudie*. Meddelanden från avdelningen för tillämpad nordisk språkvetenskap vid Lunds universitet 11. Lund.
- Wearing, A J (1970) *The Storage of Complex Sentences*. Journal of Verbal Language and Verbal Behavior 9.
- Waern, Yvonne, *Begrips?* Rapport nr 21, 1978. Psykologiska inst vid Stockholms universitet.
- Yngve, V H (1960) *A Model and an Hypotheses for Language Structure*. Proceedings of the American Philosophical Society 104.