

Øivin Andersen

Om terminologi og tekstanalyse

Summary

In this lecture the relationship between term and text level of analysis is discussed. The first part takes us from term level to text level, discussing the relationship between informativity, redundancy and language economy. The second part takes us back from text level to term level, discussing documentary based thesaurus constructions as a mediator between the two levels. The relationship between term and descriptor are focussed in the framework of concepts such as text, competence and coherence. Various types of reduction analysis, such as van Dijks macrostructural theory is discussed.

1 Fra term til tekst

Innledning

I en fremstilling som søker å belyse forholdet mellom term og tekst er det naturlig å innlede med å si noe om forholdet mellom terminologi (TL) og fagspråk (FS). I min dobbeltartikkel "fagspråk som forskningsdisiplin" i tidsskriftet *Hermes* (1989-1990) skisserer jeg opp fire mulige grenseoppganger mellom fagspråk og terminologi:

A

- 1 *TL er en del av FS*
- 2 *FS er en del av TL*
- 3 *TL er autonom*
- 4 *FS og TL er likeverdige, men gjensidig avhengige*

Alternativ 1 representerer det tradisjonelle synet. TL er studiet av de fagspesifikke sektorene av leksikon, mens FS arbeider med større språklige enheter, som morfo-syntaktiske trekk, fraser, setninger og større enheter som paragrafer og tekster.

Forskning både innen TL og FS har i stadig større grad vist at dette skillet er svært vanskelig å opprettholde. Innen TL har studiet av forholdet mellom terminologiske begreper og fraseologi vist at hovedprinsippet om ett begrep - en term, altså et isomorforhold mellom begrepsside og termside, i mange tilfelle fravikes i faglige sammenhenger.

Hvis vi ser på utviklingen av synet på forholdet mellom syntaks og leksikon i generell lingvistisk teori, finner vi at "arbeidsdelingen" mellom de to komponentene har endret seg: Leksikon får en stadig mer sentral rolle å spille i de sentrale syntaktiske teoriene.

Videre synes det å være en klar tendens innen FS og TL at studier og systematiseringer av FS-leksikon i isolasjon er utilstrekkelige hvis ikke man tar hensyn til de tekstene leksikonet brukes i. Dessuten må tekstene studeres på bakgrunn av de faglige institusjoner og miljøer som er brukere av denne terminologien.

En annen forskjell mellom FS og TL er at FS tradisjonelt har hatt en deskriptiv innfallsvinkel, mens TL er mer normativt orientert.

Dette har sammenheng med de ulike funksjonene som terminologi har og som er nedfelt i en flertydighet i selve begrepet *terminologi*. Terminologi er et sett av metoder og retningslinjer for å utføre terminologisk arbeid. Slik sett er det naturlig at TL har (og nødvendigvis må ha) en normativ vinkling. Men terminologi er også en egen vitenskapelig disiplin med et definerbart studieobjekt og et teoretisk fundament. I denne betydningen har terminologi både en deskriptiv (og klart empirisk forankret) innretning, og også, som andre vitenskaper, en forklarende funksjon.

Det viktigste felleselementet i studieobjektene for FS og TL er det problematiske begrepet "faglighet". Denne fellesnevneren befinner seg i kjerneområdet i begge disipliner. På bakgrunn av dette er det grunn til å anta at løsning 4 ovenfor kanskje er den mest naturlige; dvs at FS og TL er likeverdige og selvstendige disipliner, men at de er gjensidig avhengige av hverandres data.

Informativitet, redundans og økonomi

Begrepet informativitet er nært forbundet med Shannon/Weavers klassiske informasjonsteori (Shannon et al 1949). Hvis begrepet skal ha relevans for analyse av menneskelige naturlige språk, må begrepet koples til den intensjonale overføring av informasjon ved hjelp av et etablert signalsystem. Det naturlige basisbegrepet blir da *kommunikasjon*.

Enhver språkteori som ønsker å inkludere kommunikasjon som en sentral teoretisk term må være pragmatisk innrettet: metatermer som avsender, mottaker, kontekst og intensjonalitet må ha en teoretisk avklart status.

Både de strukturalistiske og de generative lingvistiske teoriene har det felles at språkets kommunikative funksjon ikke har noen sentral plass. For strukturalismens vedkommende skyldes dette primært at den var opptatt av språklige former og strukturer, og at språkbruk, og dermed språkbrukerne, var utdefinert i utgangspunktet. For generativistene var språkbruk et performansfenomen som ikke engang hadde forklaringsverdi overfor kompetansen.

En av de grunnleggende antakelsene fra informasjonsteorien er at signalinformasjonsinnhold er omvendt proporsjonalt med sannsynligheten for forekomst i en konkret kontekst. Jo større sannsynligheten er for at et bestemt signal forekommer i en kontekst, desto mindre informasjon inneholder det. Hvis signalet er fullstendig forutsigbart er informasjonsinnholdet lik null.

Shannon og Weaver prøvde å kvantifisere forholdet mellom informativitet og forutsigbarhet (også kalt *redundans*), men det er nå allmenn enighet om at dette ikke uten videre lar seg overføre til naturlige menneskelige språk.

Imidlertid har de grunnleggende prinsippene blitt nedfelt i særlig kognitivt psykologisk

orienterte retningene innenfor tekstlingvistikken. Disse retningene legger vekt på de prosessuelle sidene ved tekstbegrepet, som produksjon og forståelse av tekster (jf Beaugrande et al 1981).

I vår sammenheng er følgende aspekter av betydning: All form for språklig strukturering øker redundansen, og dermed forutsigbarheten i tekster. Det gjelder alle de lingvistiske nivåene, som fonotaks, morfofonologi, morfologi, syntaks og semantikk innen rammen av "språkssystemet" i snever forstand. Men det gjelder også pragmatiske strukturer som tekstkohesjon og koherens (for eksempel tema-remastrukturer), tekstuell lay-out, teksttype og genre. Noen av disse begrepene kommer vi tilbake til senere.

Her skal vi først konsentrere oss om forholdet mellom leksikon og grammatikk. La oss sette opp følgende enkle modell:

B	
<i>leksikon</i>	<i>grammatikk</i>
.....	
<i>informativitet</i>	<i>redundans</i>
<i>lav forutsigbarhet</i>	<i>høy forutsigbarhet</i>

Ifølge Lyons (1977:33f) innebærer symbolsk innhold valgmuligheter. Et signal er kun kommunikativt hvis avsenderen har til hensikt å gjøre mottakeren oppmerksom på noe som mottakeren fra før av ikke var klar over. Hvis avsender ikke har valgmuligheter, er ikke signaloverføringen kommunikativ. Hvis person P har feber, vil det være synlig for en tilstedeværende iakttaker Q. Men P kommuniserer ikke noe til Q fordi P ikke kan velge om han/hun har feber eller ikke. Hvis P derimot sier "Jeg har feber" foreligger kommunikasjon, nettopp fordi P har valget mellom å kommunisere og å la være.

Hvis kommunikasjon av symbolsk innhold impliserer valg, kan disse betraktningene koples til forskjellen mellom leksikalske ord og formord. Som kjent medfører store

valgmuligheter store paradigmer (i Saussures forstand). Hvis paradigmet er lik null har enheten ingen informasjonsbærende betydning. Et eksempel er infinitivsmerket *å* på norsk. Det er karakterisert ved ikke å kunne erstattes av noe annet enn kanskje null:

C			
I	II	III	IV
<i>Per</i>	<i>liker</i>	<i>å</i>	<i>arbeide</i>
<i>Ole</i>	<i>hater</i>		<i>leke</i>
<i>Mannen</i>	<i>elsker</i>		<i>synge</i>
...

Vi ser at i kolonne I, II og IV i C har vi rike substitusjonsmuligheter, mens vi ikke har noen muligheter i kolonne III. Derav følger at infinitivsmerket *å* i norsk ikke har noe innhold.

Generelt kan vi si at funksjonsord har høy forutsigbarhet (og dermed lite informasjonsinnhold) i tekster, mens leksikalske ord har lav forutsigbarhet (og dermed høyt informasjonsinnhold), jf B. Gjennom B har jeg også prøvd å få frem det viktige poenget at det ikke er noen vanntett grenseoppgang mellom leksikon og grammatikk på dette punktet. Visse ordtyper som for eksempel preposisjoner og visse pronomener kan sies å stå i en mellomstilling.

Beaugrande et al (op.cit) skiller mellom tre typer informativitet i tekster: første ordens informativitet med høy forutsigbarhet, andre ordens informativitet med lav forutsigbarhet og dessuten tredje ordens informativitet. Sistnevnte skaper kognitive vansker for mottakeren, som for eksempel motsigelser, sekvenser som bryter med lagret kunnskap eller representerer ny kunnskap for mottakeren. For å gjøre tredje ordens informativitet forståelig må mottakeren nedgradere tredje orden til andre orden. Forskjellige strategier er tilgjengelige for mottakeren til dette formål: Han kan gå tilbake i tidligere lest tekst (backward downgrading); han kan fortsette (forward

downgrading) eller han kan gå utenom teksten (outward downgrading).

Det er den sistnevnte typen, dvs "outward downgrading" vi skal ta utgangspunkt i når vi nå skal se nærmere på forholdet mellom term og tekst i fagspråklig sammenheng.

Dette bringer oss til begrepet *økonomi*.

Begrepet *økonomi* har en kontroversiell historie i lingvistikken. Mest kjent er vel Martinets *økonomibegrep*, koplet med hans *inertiabegrep* og *funksjonell belastning* (Martinet 1955). Disse idéene har senere også dukket opp innenfor *språkplanleggingsteori* (Tauli 1968).

Tre av Taulis *økonomiprinsipper* er verd å se nærmere på i forbindelse med forholdet mellom termer, definisjoner og forklaringer i FS-tekster:

For det første, *optimal språklig økonomi* tilsier at antallet språklige uttrykk må være lavest mulig uten at informasjon går tapt. Det innebærer for FS/TL at antallet leksikalske enheter både i systemet og i teksten må være så lavt som mulig så lenge entydighet kan opprettholdes.

For det andre må uttrykket være så kort som mulig. For det tredje må ikke uttrykket formidle mer informasjon enn det som er nødvendig (Tauli op.cit.:32).

Disse *økonomiprinsippene* finner man igjen i Grices *pragmatiske teori om implikaturer*, dvs en analyse av spenningsfeltet mellom det sagte (uttrykket) og det impliserte i konversasjoner (Grice 1975). Selve *økonomiprinsippet* formulerer Grice i form av 4 *konversasjonsmaksimer* som danner fundamentet for fremmingen av effektiv kommunikasjon av proposisjonalt innhold (som jo er et overordnet mål i FS-tekster):

D

Quantity. (i) Make your contribution as informative as is required (for the current purposes of the exchange); (ii) do not make your contribution more informative than is required.

Quality. Try to make your contribution one that is true: (i) Do not say what you believe to be false; (ii) Do not say that for which you lack adequate evidence.

Relation. Be relevant.

Manner. Be perspicuous: (i) Avoid obscurity of expression, (ii) Avoid ambiguity; (iii) Be brief (avoid unnecessary prolixity); (iv) Be orderly.

Taulis tre prinsipper faller inn under Grices kvantitetsmaksime. Kvalitetsmaksimen angir forholdet mellom avsenders og mottakers kunnskapsdybde.

Relasjonsmaksimen om relevans er et grunnlag som tematisk og makrostrukturell analyse må hvile på, og som er avgjørende for tekstens koherens. Den er derfor en sentral maksime for tematisk tekstanalyse og indeksering i dokumentasjonsbasert tesaurusarbeid.

Modalmaksimen (manner) harmonerer godt med terminologiens normative idealprinsipper om entydige og velordnede begrepssystemer.

Det mest interessante med Grices perspektiv i vår kontekst er at dette inngår som del av en analyse av den ideelle språkbrukerens kommunikative kompetanse. Perspektivet blir dermed annerledes. Det er ikke lenger bare snakk om påbud og forbud som man må forholde seg til hvis man skal drive terminologiarbeid og skrive FS-tekster, men en analyse av språkbrukerens kompetanse. Det er ikke lenger snakk om normer vi lager oss for å fremme effektiv kommunikasjon, men egenskaper ved naturlig språk (både AS og FS) som er sentrale og nødvendige for enhver overføring av kommunikasjon ved hjelp av språk.

I FS-sammenheng optimaliseres i større grad Grices økonomiske maksimer enn i AS.

Som del av den kommunikative kompetansen kan dette danne en del av rammene og utgangspunktet for studiet av FS-tekster. FS vil da neppe kunne betraktes som en autonom disiplin med et eget studieobjekt i kvalitativ forstand. Men studiet av FS-tekster, så langt det er mulig å definere dem som FS-tekster, vil kunne yte verdifulle bidrag til pragmatikken.

Den praktiske nytten av dette har man allerede sett konsekvensen av innenfor kunstig intelligens og ekspertsystemer, hvor økonomi og entydighet er av avgjørende betydning for tilrettelegging for datamaskinell behandling.

Som nevnt har kunstige språk flere trekk til felles med FS, blant annet entydighet og økonomi. Dette ser vi klarere hvis vi prøver å oversette for eksempel predikatskalyler i symbolsk notasjon til naturlig språk. Hvis entydigheten skal opprettholdes, blir syntagmene (det lineære uttrykk) lengre i det naturlige språket.

På samme måte som man kan se på kunstig språk som en "oversettelse" fra naturlig språk, kan definisjoner og forklaringer sees på som en oversettelse fra FS til AS. Dette synet finner vi eksempel på i leksikografien hos Knudsen et al (1957) hvor man foreslo å anvende Hjelmleivs kommutasjonstest for å måle definisjoners brukbarhet i Norsk Riksmålsordbok. En ordboksdefinisjon burde være istand til å erstatte et ord i en vanlig setning (Knudsen et al op. cit.:124).

Det samme kravet til definisjoner finner man også i FS og TL, for eksempel hos Laurén (1985:84). Han krever blant annet at definisjonen må formuleres slik at den uten endringer kan erstatte termen i løpende tekst.

I moderne leksikografi har man innsett at denne kommutasjonstesten i svært mange tilfeller er umulig, fordi et leksem (som en ordboksdefinisjon representerer) tilhører et mer abstrakt nivå enn ordformer i tekster. Som Greimas påpeker gis leksemet kun en partiell realisering i en tekst eller en diskurs (Greimas 1981:196). Leksemet er altså, som termen, en abstraksjon, dvs en typebetegnelse, mens det er de realisererte

ordformene, eksemplarene på tekstnivå, vi observerer. Det er kun sistnevnte som kan være referanseuttrykk.

Men selv om kommutasjonsprinsippet bunner i manglende evne til å holde leksemnivå og tekstnivå fra hverandre, sier den likevel noe om definisjonens kommunikative funksjon.

La oss derfor se litt nærmere på implikasjonene av Lauréns krav. I TL er definisjonen en språklig (forklarende) formulering av et fagbegrep. Fagbegrepet representeres språklig av en term i FS-tekster. Av det følger at fagterm og definisjon skal referere til samme begrep. Sagt på en annen måte: definiendum, det som skal defineres (termen) skal kommunikativt sett være ekvivalent med definiens, det som definerer (definisjonen).

Hovedforskjellen blir da at definiendum er kort mens definiens normalt er lengre enn definiendum, nettopp fordi definiens skal forklare definiendum.

Forenklet sett kan vi si følgende: Hvis fagbegrep a skal kommuniseres fra person P1 til P2, kan den korte definiendum brukes hvis P2 kjenner begrepet (og termen som brukes). Dermed økes økonomien i fremstillingen. Hvis termen brukes og P2 ikke kjenner begrepet, må P2 foreta en "outward downgrading" i Beaugrandes terminologi. Det betyr at han må konsultere terminologiske oppslagsverk eller andre kilder som ligger utenfor teksten for å opparbeide seg en forståelse. Alternativet er å bruke den lengre definiens for å erstatte den uforståelige definiendum med resulterende tap av økonomi i fremstillingen.

Jo færre definiendumer P2 kjenner innenfor et emne, desto flere fagspesifikke (korte) definiendumer (termer) må erstattes med lengre definienser eller tilsvarende lange omskrivninger og forklaringer. Noen av de samme resonnementene er Jan Roald inne på i en av sine artikler (Roald 1985:257).

Men språklig økonomi er det logisk komplementære til språklig redundans. I følge Platzack (1977:6) kan man gjøre en meddelelse mindre følsom for forstyrrelser ved å bygge ut meddelelsen med språklig "semantisk" redundans eller tilleggsinformasjon. "Tilleggsinformasjon" defineres som den informasjonen som tilføres meddelelsen i dens kortest tenkelige utforming. På en måte kan dette sees på som et slags sløseri fordi man gjør meddelelsen lengre uten å øke informasjonsinnholdet, dvs tap av økonomi og økning i redundansen. Denne økningen har likevel den funksjon at den sikrer mot forvekslinger og misforståelser. Mottakeren får altså økte muligheter til å oppfatte innholdet (og dermed slippe å foreta en "outward downgrading") selv om deler av meddelelsen ikke når frem. Det blir altså lettere å gjette seg til de delene man ikke oppfattet i selve signaloverføringen.

Hvis det er mulig å tenke seg en FS-tekst uten terminologi og kunstig språk (som tabeller, modeller og andre formalismer), vil mye mer av den språklig økonomiske gevinst gå tapt. Økonomiegenskapen ved FS hviler altså i stor grad nettopp på innslaget av terminologi og kunstig språk.

I enhver språkkommunikativ sammenheng er det altså et nærmest kybernetisk samspill mellom økonomi og redundans som er styrt av behovet for entydighet og presisjon på den ene siden, og behovet for forståelighet på den andre siden.

Term, tekst og popularisering

Dette aktualiserer populariseringsaspektet i FS-tekster, et spørsmål som har relevans for avgrensingsproblematikken vår: Når slutter FS og når begynner AS? Som Jan Roald påpeker gjør populariseringsaspektet ved FS avgrensingen mellom AS og FS uklar: Hvis mottakeren av en fagtekst ikke er en fagmann på det aktuelle området foreligger en fagmann-lekmann-kommunikasjon. I slike tekster vil avsender, for å oppnå en vellykket kommunikasjon, justere utsagn om begrepshierarkier etter mottakergruppens antatte forutsetninger (omskrivninger i stedet for tekniske termer, eksplisitering på

tekstnivå av en rekke karakteristiske trekk ved begrepet o.l.) Roald 1986a:184).

Vil en sterkt popularisert FS-tekst i hovedsak skille seg fra en ikke-popularisert FS-tekst ved at termer og bruk av kunstig språk i den ikke-populariserte versjonen er erstattet av definisjoner, forklaringer og omskrivninger i den populariserte versjonen (hvis vi antar at det denotative innholdet er identisk i de to variantene)?

Kan man for eksempel si at det er et omvendt proporsjonalt forhold mellom bruk av syntagmene (bruk av definisjoner og forklaringer) og bruk av paradigmene (slik vi bruker termen i C ovenfor), det vil si bruk av termer, for å få frem begreper i FS-tekster? Vil dette være variabler som leder fra FS til AS via popularisering? Vil en popularisert FS-tekst m.a.o. belaste språkets syntagmatiske dimensjon på bekostning av den paradigmatiske (i lingvistisk forstand), og dermed gi avkall på språklig økonomi for å lette forståeligheten?

Hvis man som fagmann skal omtale et emne eller problemkompleks til en faglig sett utenforstående, er det mange termer man ikke kan bruke. Man må da erstatte termene med definisjoner og forklaringer for å opprettholde forståelsen hos mottakeren, samtidig som presisjonsnivå og entydighet skal opprettholdes.

Hvis dette er tilfellet, så blir definisjonen en oversettelse fra FS til AS; en kommunikativ brobygger mellom de to. Av det følger logisk at definisjoner ikke kan vurderes fyldestgjørende i et abstrakt vakuum, men må formuleres med utgangspunkt i situasjon og brukergruppe.

Dette harmonerer med en antakelse om at FS-varianter er parasittiske på AS, men synes å bryte med Thomas Kuhns tese om at observasjoner er "paradigmeimpregnerte" i den forstand at data og observasjoner kun er forståelige gjennom det paradigmet, eller den faglige matrisen, de tilhører. Hvis paradigmer skal kunne konkurrere med hverandre, må de være gjensidig forståelige.

Hvis man, som et tankeeksperiment, forestiller seg at to ulike FS-varianter overlapper med to ulike faglige matriser, så vil disse to variantene i alle fall være indirekte kommensurable, fordi de begge er parasittiske på AS, som jo har et gjennomgripende sett av normer som FS-variantene må relatere seg til.

Som konklusjon kan man si at termer i tekster er økonomiindikatorer, og at definisjoner og forklaringer, i den grad de kan substituere for termer i tekst, er faktorer som viser forholdet mellom det abstrakte termnivået, og det konkrete tekstnivået gjennom populariseringsaspektet.

2 Fra tekst til term

Dokumentasjonsbasert tesaurus som formidler mellom term og tekst

Mens definisjoner og forklaringer kan sies å formidle fra term til tekst, kan dokumentasjonsbaserte tesauruskonstruksjoner sies å formidle den andre veien: fra tekst til term. For å anskueliggjøre dette er det nødvendig å gå litt nærmere inn på hva en dokumentasjonsbasert tesaurus er. La oss angi den internasjonale standardiseringsorganisasjonen ISOs definisjon som utgangspunkt:

E

A thesaurus may be defined either in terms of its function or its structure.

In terms of function, a thesaurus is a terminological control device used in translating from the natural language of documents, indexers, or users into a more constrained "system language" (documentation language, information language).

In terms of structure, a thesaurus is a controlled dynamic vocabulary of semantically and generically related terms which cover a specific domain of knowledge (ISO 2788).

Definisjonen bør innenfor denne diskusjonsrammen også presiseres og utdypes på følgende måte: en dokumentasjonstesaurus er en systematisk ordnet hierarkisk liste over emner som brukes til indeksering av og søking etter referanser i konkrete

dokumentsamlinger. Indeksering er i den forbindelse en type informasjonsreduksjon, hvor tekstinhold reduseres til en liste av indekseringsord (en indeks). Denne indeksen behandles terminologisk og gjennomgår normering. Et normert indekseringsord kalles en *deskriptor* og kan sies å være det dokumentasjonsbaserte motstykke til en normert term innen et fagområde. Det fagområdet som dette feltet tilhører kalles informasjon og dokumentasjon (iod).

Terminologi- og fagspråksforskningen spiller en vesentlig rolle i iod. Som Picht et al (1985:193) påpeker er terminologilærens prinsipper av grunnleggende betydning for iod, fordi iod er avhengig av mest mulig motiverte og entydige betydningsindikatorer (indeksord) for effektiv gjenfinning av dokumenter. Indeksordene i disse systemene må med andre ord være styrt og normert, som i terminologilæren.

Men fordi indekser bygges opp ut fra informasjonsinnholdet (eller mer presist, det denotative innholdet) i store mengder av tekster, vil fagspråksforskningen også stå sentralt.

Termer og deskriptorer

Indeksering i iod består av en innholdsanalyse av *hvert* enkelt av en mengde dokumenter. Den refereres til som en *dokumentasjonsprosess*, dvs tilrettelegging og behandling av dokumenter i informasjonsgjenfinningsøyemed. Målet for denne prosessen er å kunne formulere (mest mulig) av informasjonsinnholdet ut fra en begrenset mengde temaer. Temaet uttrykkes gjerne som enkle ord eller fraser. Disse kalles *deskriptorer*. Deskriptorene må organiseres på en slik måte at informasjonssøkerne kan relatere dem til sine søkebehov. Målet er at alle dokumenter som er relevante fra brukerens synspunkt, blir avdekket eller funnet ved søking. I motsatt fall sies systemet å ha *ballast*, dvs gjenfinningsmangler (jf Sparck Jones et al 1973, Lang 1980).

Utgangspunktet for dokumentasjonsprosessen kalles *kildetekst* (originaltekst), og sluttproduktet kalles *indeksbeskrivelsestekst*.

Indekseringen av tekster gir en liste av indekserte ord. Disse blir så klassifisert tematisk i ordnede begrepsfelt ved å anvende et "systemspråk" som styringsredskap. Et systemspråk er et sett av instruksjoner for etablering av leksikalske relasjoner (kalt *tesaurusrelasjoner*) innenfor et avgrenset begrepsfelt eller mellom begrepsfelt. Et begrepsfelt kan tilsvare et fagområde (for eksempel geologi, jus eller lingvistikk), eller en del av et fagområde. Det kan også tilsvare et teknologisk område som for eksempel reservoarteknikk, et rørsystem eller et elektroteknisk installasjonssystem. Hva som utgjør et slikt felt avhenger i siste instans av dokumentenes art og brukernes informasjonsbehov (som igjen er betinget av deres faglige kompetanse).

Begrepsfeltene struktureres så videre til tematiske hierarkier med assosiative relasjoner på tvers av hierarkiene hvor det er "nyttig" i gjenfinningsammenheng.

I terminologiske begrephierarkier søkes en (så vidt mulig) fullstendig beskrivelse av et gitt begrepsområde. Hierarkiene søkes derfor fremstilt uten "huller". Tesauroshierarkiene, derimot, bygger på tematikken i et klart avgrenset korpus av tekster. Av den grunn kalles de *temahierarkier*. Derfor kan et tesauroshierarki fortone seg "ufullstendig" og "hullet" sett i relasjon til et begrepssystem. Vi kan si at en dokumentasjonsbasert tesaurus viser et gitt fagområde utelukkende slik det fremtrer i de dokumentene som inngår i dokumentasjonsbasen. Tematisk vektning av informasjonsinnhold er av sentral betydning: Det som forfatteren trekker frem i sin fremstilling som sentralt, vil også få en sentral plass i dokumentets informasjonsstruktur.

En tesaurus består som ferdig struktur (jf strukturdefinisjonen under E) av en alfabetisk del (hvor deskriptorene er alfabetisk sortert) og en hierarkisk, systematisk del. Det finnes assosiative forbindelseslinjer både fra alfabetisk del, og mellom hierarkiene i systematisk del.