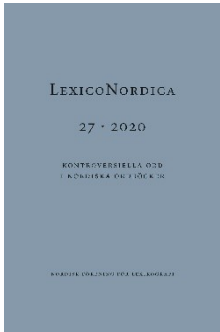


# LexicoNordica

Titel:	Svartsjuk tankelesing på vandresafari – en modell for bedømmelse av sammensatte ords gjennomsiktighet	
Forfatter:	Mikkel Ekeland Paulsen	
Kilde:	LexicoNordica 27, 2020, s. 161-182	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive">http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive</a>	

© 2020 LexicoNordica och författarna

## Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

# *Svartsjuk tankelesing på vandresafari* – en modell for bedømmelse av sammensatte ords gjennomsiktighet

*Mikkel Ekeland Paulsen*

Compounds are ubiquitous in Norwegian language use, and therefore also in Norwegian dictionaries. Whether a compound is selected for entry in the dictionary depends in part on the lexicographers' ability to distinguish transparent from nontransparent compounds. However, what decides if a compound is semantically transparent has never been fully investigated scientifically, at least not from a lexicographic point of view. This study outlines a model for assessing the semantic transparency of Norwegian compounds and presents the results of the model applied to a set of compounds.

## 1. Innledning

Sammensetninger utgjør en betydelig del av ordtilfanget i *Bokmålsordboka* (BOB) og *Nynorskordboka* (NOB), ifølge beregninger så mye som 60 % (Kjelsvik 2018). Et trunkert søk på ordet *arbeid* (arbeid\*) i Leksikografisk bokmålskorpus (LBK; se også Knudsen & Fjeld 2013) gir treff på over 2700 ulike lemma. Noen av disse er usammensatte, enkelte er ren støy, mens over 90 % er sammensetninger. I moderne leksikografi utgjør tekstkorpus et av de viktigste kilde-materialene. Her er utvelgelse av sammensatte oppslagsord notorisk problematisk. I norsk er orddanning via sammensetting høyproduktivt, og det er derfor viktig at leksikografer har velfunderte metoder for å vurdere sammensetningenes leksikografiske relevans.

Et viktig kriterium for å utelukke sammensetninger er semantisk gjennomsiktighet. Likevel fins det foreløpig ikke en velfundert

leksikografisk framgangsmåte for å bedømme slik gjennom-siktighet.<sup>1</sup>

I denne artikkelen presenterer og anvender jeg en modell for bedømmelse av gjennom-siktighet i sammensetninger, basert på faktorer som av ulike sammensetningsforskere hevdes å ha innflytelse på dette. Modellen anvendes til en undersøkelse av et utvalg sammensetninger som blir plassert på en skala fra minst til mest gjennom-siktig. Denne framgangsmåten tar sikte på å være et ledd i å utvikle en metodikk for å identifisere ordbokaktuelle sammen-setninger.

## 2. Teoretisk bakgrunn

I det følgende gjør jeg rede for termen *gjennom-siktige sammensetninger* og påpeker problematiske sider ved den tradisjonelle definisjonen av denne. Deretter diskuterer jeg ulike egenskaper som hevdes å påvirke gjennom-siktighetsgraden til den enkelte sammensetning.

Fjeld & Vikør (2008) trekker opp et skille mellom såkalte gjennom-siktige sammensetninger, til dømes *teaterbillett*, *postfunksjo-nær* og *jarnomn*, og sammensetninger med spesialiserte betydninger som ikke framgår direkte av sammensetningsledda, til dømes *teatersport*, *postgiro* og *jarnbane*.<sup>2</sup> Distinksjonen bidrar til å skille ordbokaktuelle sammensetninger fra sammensetninger som kan utelates.

*Gjennom-siktige* sammensetninger og motstykket *ugjennom-siktige (opake)* sammensetninger er mye brukte termer innenfor sammensetningslitteraturen, og kanskje særlig innenfor praktisk

1 Det fins andre kriterier også, til dømes frekvens i bruk. Dessuten avhenger det av ordbokas formål hvor viktig kriteriet om semantisk gjennom-siktighet er.

2 Døma til Fjeld & Vikør er nynorskformer. *Jarnomn* og *jarnbane* tilsvarer *jernovn* og *jernbane* på bokmål.

leksikografi (se bl.a. Fjeld & Vikør 2008). Termparet begrepsliggjør sammensetninger som et objekt med sider som vi i varierende grad er i stand til å se igjennom (Svanlund 2002). De mest regelbundne og forutsigbare sammensetningene har angivelig mer gjennomskuelige flater enn de mindre regelbundne. For å avgjøre om en sammensetning er gjennomiktig, bruker man ofte en adisjonsmetafor (se til dømes Bakken 1998 og Fjeld & Vikør 2008). Her framstilles betydninga til en sammensetning som en sum av ledda. Presumptivt blir da formelen *forledd + etterledd = betydning: jern + ovn = jernovn*. Motstykket er sammensetninger hvor denne formelen tilsynelatende ikke passer, til dømes *jern + bane ≠ jernbane*.

Svanlund (2002) kritiserer dette synet for å utgå fra et forenklet syn på ordsemantikk og å overse sammensetningenes betydningspotensial. *Jernovn* kan like gjerne bli brukt med betydninga 'ovn som man smelter jern i' (dette fins til dømes i bøker om jernproduksjon) som 'ovn av jern'. De er begge potensielle og aktuelle betydninger av sammensetninga, men dette kommer ikke fram når man behandler betydninga som om det var et enkelt regnestykke.

Når man vurderer om sammensetninger er gjennomiktige, er det viktig å presisere at det man vurderer, er gjennomiktighet på tvers av ulike språkbrukere og språklige kontekster. Det vil alltid være individuell variasjon i fortolkninga av sammensetninger (se til dømes Loenheim 2019 og Ryder 1994). Sammensetninger er prinsipielt mangetydige (Svanlund 2002) i den forstand at de via polysemi i forledd og etterledd og i relasjonen dem imellom kan anta et multiplum av gangbare betydninger. Likevel tyder mye på at noen sammensetninger *oppleves* som mer gjennomiktige enn andre (Svanlund 2002 og Loenheim 2019). Det er til dømes grunn til å tro at en sammensetning som *adamseple* oppleves som mindre gjennomiktig enn *treski*. Med dette menes det at det er vanskeligere å utlede den konvensjonelle betydninga til sammensetninga ut fra ledda i *adamseple* enn i *treski*. Hva denne forskjell-

len i opplevd gjennomsiktighet består i, kan blant annet forklares gjennom de faktorene jeg trekker fram her.

Svanlund (2002, 2009) presiserer forskjellen mellom *bedømmelse* og *opplevelse* av gjennomsiktighet. Opplevelse av gjennomsiktighet gjelder språkbrukerens første møter med en til da ukjent sammensetning. Leksikografer må som regel nøye seg med å bedømme gjennomsiktighet. Bedømmelsen skjer da utenfor kontekst og derfor reelt sett på tvers av ulike kontekster. Svanlund (2002) knytter kontekstfri gjennomsiktighet til *motiveringsgrad*. Motivering forstås her som en identifiserbar sammenheng mellom et ords betydning innenfor og utenfor en sammensetning. Man kan til dømes konstatere at betydninga av *blå* er å gjenfinne i sammensetninga *blåbær*, som tross alt denoterer et blått bær. Her framstår det ikke arbitrært, men snarere motivert at akkurat adjektivet *blå* og substantivet *bær* settes sammen for å denotere nettopp 'blått bær'. Såleis har sammensetninga *blåbær* høy motiveringsgrad. På den annen side burde *blåbær* med full gjennomsiktighet kunne bli brukt om alle blå bær, slik som den frie frasen *blått bær*. Den konvensjonelle betydninga av *blåbær* er imidlertid snevrere enn det som framgår av ledda alene, altså fins det en distinksjon mellom frasen *blått bær* og *blåbær* som gjør sistnevnte mindre gjennomsiktig, og derfor til en aktuell kandidat for ordbøker. Det vil si at en sammensetning kan ha høy motiveringsgrad, altså at leddas konvensjonelle betydninger er gjenfinnbare i sammensetningas konvensjonelle helhetsbetydning, men ikke være helt gjennomsiktig likevel. Førstnevnte er med andre ord en nødvendig, men ikke en tilstrekkelig betingelse for at en sammensetning skal bedømmes som helt gjennomsiktig.

Det at *blåbær* har en snevrere konvensjonell betydning enn ledda skulle tilsi, betegner Svanlund (2002) som at sammensetninga har tillegselementer. Et annet døme er sammensetninga *kvinnesak*, der det på ingen måte er tilstrekkelig å vite betydninga av ledda for å komme fram til helheten. En stor gruppe sammen-

setninger som har tilleggselementer, er fagtermer. For disse er den konvensjonelle betydninga snevrere enn det som framgår av ledda alene, tenk bare på termen *sammensetning* i grammatikken.

En forutsetning for ideen om motiveringsgrad og tilleggselementer er at sammensetninger har tydelige konvensjonelle betydninger som språkbrukere i hovedsak er enige om. Dette er ikke en uproblematisk forutsetning. Schmid (2015) understreker at det er et samspill mellom konvensjonalisering og eksponeringshyppighet (eller innprenting) som leder fram til en gitt språkbrukers tolkning av et gitt uttrykk. I den grad det fins slike konvensjonelle betydninger, fins det ingen garanti for at leksikografer har tid eller ressurser til å avdekke disse i alle tilfeller. Spørsmålet om hvordan man fullt ut kan identifisere konvensjonelle betydninger, må imidlertid besvares annetsteds. I kapittel 3 beskriver jeg nærmere hvordan jeg identifiserer den konvensjonelle betydninga til uttrykka som inngår i undersøkelsen.

Et annet poeng fra Svanlund (2002) er at sammensetninger består av ledd som utgjør en minimal kontekst for hverandre. En sammensetning kan såleis bidra til å disambiguere seg selv ved at ledda spesifiserer hverandre. Et døme på dette har vi i *blåfarge*, der etterleddet *-farge* bidrar til å spesifisere at betydninga til forleddet *blå-* ikke er 'politisk konservativ' eller 'melankolsk', men snarere 'fargen blå'.

Shoben & Gagné (1997) har funnet at ulike ledd har *relasjonspreferanser* som språkbrukerne anvender når de tolker og produserer sammensetninger. Relasjonspreferansene stammer fra eksponeringshyppighet, dvs. hvor ofte vi produserer og blir eksponert for en bestemt relasjon, men også fra ekstralingvistisk kunnskap om strukturen til forskjellige substanser og prosesser (til dømes at jern er et materiale mange gjenstander består av). Relasjonspreferansene er altså en type forhåndskunnskap vi aktiverer i møte med nye sammensetninger. Til dømes bruker *ffjell* som forledd å angi plasseringa til det som etterleddet betegner, som i *ffjellgeit*. Konklusjo-

nen til Shoben & Gagné (1997) er at språkbrukerne har lettest for å tolke sammensetninger der relasjonen mellom for- og etterledd rimer med relasjonspreferansene til forleddet. Som omtalt i Loenheim (2019) har denne konklusjonen blitt utdypet av blant andre Maguire, Maguire & Cater (2010) og Svanlund (2009). De har vist at også etterleddets relasjonspreferanser spiller inn.

Loenheim (2019) finner at språkbrukere i hovedsak er mer samstemte i fortolkninga av frekvente sammensetninger, som de sannsynligvis har blitt eksponert for før, enn i fortolkninga av infrekvente sammensetninger. Når det er samstemmighet om forståelsen av infrekvente sammensetninger, skyldes det ofte at de tolkes i analogi med mer innarbeida sammensetninger med samme relasjonsmønster. *Politikvinne* kan til dømes tolkes i analogi med *politimann*. Gjennomsiktighet blir da også et spørsmål om hvorvidt det fins analoge sammensetninger som avkoderen har kjennskap til og kan generalisere ut fra.

Hittil har jeg diskutert gjennomsiktighet i sammensetninger uten å differensiere mellom ulike typer.<sup>3</sup> Substantiv-substantiv-sammensetninger er den vanligste typen i norsk, og den som med god margin er mest omtalt i litteraturen. Sammensetting begrenser seg likevel på ingen måte til denne typen. De leksikalske ordklassene – dvs. substantiv, verb, adjektiv, adverb og til en viss grad preposisjoner – kan med noen unntak kombineres fritt med hverandre (noen dømer er *balanseføre*, *nordetter*, *langveis* og *bofast*). Dette er en kjensgjerning for leksikografer, som ikke kan vie spesiell oppmerksomhet til én type sammensetninger. Selv om sammensetningstypen nok kan ha innvirkning på gjennomsiktigheten (Loenheim 2019), utelater jeg dette perspektivet her. Jeg kompenserer med å teste modellen på tre ulike typer forledd, som uunngåelig dekker minst tre ulike typer sammensetninger.

3 *Type* dreier seg her og videre i teksten om ordklassetilhørigheten til ledda. Sammensetninger kan også kategoriseres etter andre kriterier.

I opplevelsen av gjennomsiktighet gjør også distribusjonsaspekter seg gjeldende (Svanlund 2009 og Loenheim 2019). Disse aspektene faller imidlertid utenfor rammene av denne artikkelen og må heller sees i sammenheng med vurderinger av sammensetningers utbredelse i korpus.

### 3. Modell for bedømmelse av gjennomsiktighet

I dette kapitlet presenterer jeg en gruppe gjennomsiktighetsfaktorer som ble diskutert i kapittel 2, og beskriver en metode for hvordan disse faktorene kan brukes til å bedømme gjennomsiktighet i sammensetninger. Metoden tester jeg i kapittel 4 på en samling sammensetninger som jeg introduserer i slutten av dette kapitlet.

For å vurdere graden av gjennomsiktighet vekter modellen sammensetninga etter følgende kriterier:

- **tilleggselement:** hvorvidt fortolkninga av en sammensetning fordrer kunnskap om mer enn betydninga til ledda
- **motiveringsgrad:** hvorvidt ledda bidrar med sine konvensjonelle betydninger
- **disambiguering:** hvorvidt ledda disambiguerer hverandre
- **relasjonsmønster:** hvorvidt sammensetninga kan forstås ut fra leddas relasjonspreferanser
- **analogi:** hvorvidt sammensetninga kan forstås i analogi med mer frekvente sammensetninger

I det følgende forklarer jeg hver faktor nærmere.

#### 3.1. Metode

**Tilleggselement:** Har sammensetninga konvensjonell(e) betydning(er) som er videre eller snevrere enn det som framgår av led-



da alene? **Framgangsmåte:** Sammensetning X får +1 hvis den på tilfredsstillende vis kan parafraseres uten at man tar i bruk andre innholdsord enn de som inngår i sammensetninga (sånn som *hårfjerning* → *fjerning av hår*).

**Motiveringsgrad:** Bidrar ledda med sine konvensjonelle betydninger i sammensetninga? Til dømes er en som er svartsjuk, verken konvensjonelt sjuk eller konvensjonelt svart. **Framgangsmåte:** For å operasjonalisere vurderinga av hva som er et gitt ledds konvensjonelle betydning, anser jeg som en nødløsning betydning 1 i BOB som den mest konvensjonelle. Om ordet ikke står i BOB, vil jeg konsultere *Det Norske Akademis ordbok* (NAOB).<sup>4</sup> Framgangsmåten blir slik: Sammensetning X får +1 hvis betydning 1 av forleddet er aktiv og + 2 hvis betydning 1 av etterleddet er aktiv i sammensetninga. Ellers får den 0. Kopulative sammensetninger, som ikke har en distinkt semantisk kjerne, kan kun få +1 for hvert ledd.

**Disambiguering:** Disambiguerer ledda hverandre? **Framgangsmåte:** Sammensetning X gis +1 i gjennomsiktighet hvis ledd A har flere betydningsnivåer i BOB og ledd B bidrar til å identifisere hvilket av disse betydningsnivåene av ledd A som er aktivt i sammensetninga, eller vice versa.

**Relasjonsmønster:** Er sammenstillinga av ledd A og B i tråd med ledd As eller ledd Bs relasjonspreferanser? Faktoren *relasjonsmønster* er inkludert i modellen med bakgrunn i psykolingvistiske studier av Shoben & Gagné (1997) og Maguire, Maguire & Cater (2010), som indikerer at relasjonspreferansene til ledda letter og stabiliserer tolkninga av sammensetningenes betydning. For å bruke et relasjonsrammeverk som er empirisk utviklet med basis i norsk, benytter jeg meg av Eieslands (2015) 14 relasjonstyper for

4 Det er ikke ideelt å bruke betydning 1 i NAOB, siden denne ordboka eksplisitt oppgir at betydning 1 er den *eldste* og ikke den mest konvensjonelle betydninga. Disse er stundom, men slettes ikke alltid, sammenfallende. Når NAOB likevel brukes i tillegg til BOB, er det i mangel av andre dugelige bokmålsressurser.

substantiv-substantiv-sammensetninger. Andre sammensetningstyper, som adjektiv-adjektiv, adjektiv-substantiv, substantiv-verb og adjektiv-verb, er viet langt mindre oppmerksomhet i litteraturen. Her er det derfor ikke mulig å ta i bruk et veldefinert rammeverk på samme måte. I sammensetninger der et verb er sammenstilt med et substantiv, dreier relasjonsmønsteret seg om den semantiske rollen til substantivet. Det samme gjelder når et verbavleddet ledd er sammenstilt med et substantiv. *Folk* er til dømes agens i sammensetninga *folkelesing*, mens *avis* er patiens i *avislesing*. Når det gjelder adjektivsammensetninger, skiller Loenheim (2019) mellom kopulative og determinative adjektiv-adjektiv-sammensetninger. *Gulgrønn* kan til dømes tolkes kopulativt som 'gul og grønn' eller determinativt som 'gulaktig grønn'. I substantiv-adjektiv-sammensetninger skiller hun mellom en sammenliknende og en årsaksbeskrivende relasjon mellom ledda. *Iskald* ('kald som is') er døme på det første og *solvarm* ('varm av sola') på det andre.

**Framgangsmåte:** Sammensetning X gis +2 i gjennomsiktighet hvis ledd A og ledd B har en relasjon til hverandre som er i tråd med begge leddas relasjonspreferanser. X gis +1 hvis ledd A eller ledd B har en relasjon til motsatt ledd som følger relasjonspreferansene. Preferansene utledes fra de fem mest frekvente sammensetningene i LBK med ledd A som forledd, og tilsvarende fra de fem mest frekvente med ledd B som etterledd. Dersom minst tre av de fem har samme relasjon, er dette relasjonspreferansen til ledd A eller B.

**Analogi:** Fins det analoge sammensetninger som letter tolkninga av sammensetning X? Denne faktoren nevnes som vesentlig for en sammensetnings gjennomsiktighet av både Loenheim (2019) og Svanlund (2002). Førstnevnte konkluderer med at «hurvida en sammansättning som är ny för den enskilde språkbrukaren framstår som genomskinlig sammanhänger med vilken tillgång hon har till analogibaser» (2019:281). Her kan det i enkelte tilfeller være tilstrekkelig med én eneste analog sammensetning, som i tilfellet med *blåstrømpe* 'høyrevridd feminist', som trolig tol-

kes i analogi med *rødstrømpe* ‘venstrevridd feminist’, eller tilfellet *forkvinne* som analogi til *formann*. Begge disse tilfellene er imidlertid enveiskjorte i den forstand at den minst etablerte termen forstås i analogi med den mest etablerte, og ikke motsatt. Analoge sammensetninger kan i prinsippet identifiseres på mange nivåer, men jeg inntar her et restriktivt syn på analogi og medregner kun sammensetninger med en helt åpenbar mønster- eller betydningslikhet. For at sammensetning X skal være analog med sammensetning Y, må sammensetning X være mindre frekvent enn sammensetning Y. Videre må X og Y ha ledd A eller B til felles, og det resterende leddet må inngå i samme leksikalske felt. Alternativt må sammensetning X og Y ha tilsvarende definisjoner i enten BOB eller NAOB, samtidig som de har minst ett ledd til felles. *Fjellhare* er en analogi til *fjellrev* fordi den er mindre frekvent, fordi den har forleddet *fjell-* til felles, og fordi *hare* inngår i samme leksikalske felt som *rev*. **Framgangsmåte:** Sammensetning X får +1 i gjennomsiktighet hvis det fins en analog i LBK til sammensetning X etter disse kriteriene.

*Motiveringsgrad* og *relasjonsmønster* vektes høyere enn andre faktorer (her er det mulig å få opptil henholdsvis 3 og 2 poeng) siden disse trolig har størst innflytelse på gjennomsiktigheten til sammensetninger.

For alle flertydige sammensetninger blir den presumptivt minst gjennomsiktige betydninga bedømt. Det er nemlig denne som på semantisk grunnlag motiverer ordbokstatusen til en gitt sammensetning.

### 3.2. Data

I den påfølgende undersøkelsen, som viser modellen i bruk, vil jeg analysere gjennomsiktigheten til tre grupper sammensetninger, nemlig sammensetninger som begynner med henholdsvis *svart-*, *vandre-* og *tanke-*. Utvalget består av samtlige toledda sammenset-

ninger med disse forledda i BOB, komplementert med de 15 mest frekvente sammensetningene i LBK (se oversikt i tabell 1). Disse utvalgskriteriene skaper en skeivhet i utvalget i retning av etablerte sammensetninger. Ulempen oppveies imidlertid av at det alltid er det mest frekvente segmentet av sammensetninger med et gitt forledd som er aktuelt for ordbokoppføring, i det minste i arbeid med allmennordbøker, som har i oppgave å fange det mest sentrale ordforrådet i et språk.

## 4. Resultater

I dette kapitlet vil jeg kort presentere resultatene for de tre sammensetningsstrekke, før jeg i neste kapittel går over til å diskutere modellens fordeler og ulemper. En oversikt over gjennomsiktighetsgraden til hver enkelt sammensetning beregnet etter modellen gis i tabell 1 mot slutten av kapitlet. I denne oversikten mangler kolonnen lengst til høyre, altså den mest gjennomsiktede enden av skalaen, siden ingen sammensetninger fikk maksimal uttelling på gjennomsiktighet.

### 4.1. *Svart-*

*Svart*-sammensetningene (heretter *s*-sammensetninger) fordeler seg over så godt som hele gjennomsiktighetsskalaen, fra de mest gjennomsiktede (til dømes *svartbrun* og *svarthåret*) til en større samling som er minimalt gjennomsiktede (til dømes *svartsjuk* og *svartemarje*). Bakgrunnen for spredninga er trolig de polyseme egenskapene til forleddet *svart-*, som medfører at ulike mentale forestillinger aktiveres i ulike sammensetninger. I materialet opptrer *svart-* med disse betydningene:

- ‘som har svart farge’, til dømes *svartjord*, *svartspett*, *svart-ovn*, *svarthåret*
- ‘som kjennetegnes av noe med fargen svart’, til dømes *svartedauden*, *svartemarje*, *svartkopp*, *svartskjorte*
- ‘djevlesk’ eller ‘okkult’, til dømes *svartemannen*, *svartebok*, *svartekunst*
- ‘dyster’ eller ‘nedstemt’, til dømes *svartsjuk*, *svartsyn*, *svart-sinn*
- ‘ulovlig’ eller ‘bannlyst’, til dømes *svartebørs*, *svarteliste*

Kun 12 av 39 s-sammensetninger kan parafraseres mer eller mindre tilfredsstillende uten andre innholdsord. Til dømes er *svarthåret* og *svartovn* mer eller mindre ekvivalente med frasene *med svart hår og ovn som er svart*, mens *svartebørs* ikke kan erstattes av frasen *børs som er svart*.

Når det gjelder motiveringsgrad, samsvarte 19 av 39 sammensetninger med betydning 1 i BOB: ‘med farge som sot eller kull’. Her fins bl.a. *svartand*, *svartbak* og *svartrot* som denoterer noe med slik farge, mens *svartebok*, *svartmale* (i overført betydning) og *svartemannen* ikke er motivert av den konvensjonelle betydninga av *svart*.

5 av 39 sammensetninger får uttelling på bakgrunn av kriteriet disambiguering. Her har vi sammensetningene *svarthvit*, *svart-brun*, *svartsmusket*, *svartsladd* og *svartlakkere*. Disse er disambiguert med bakgrunn i at etterledda understreker fargebetydninga til forleddet *svart*.

Ingen av s-sammensetningene har fått uttelling på bakgrunn av relasjonspreferansene til forleddet. Dette har sin rot i at de 5 mest frekvente sammensetningene i LBK, *svart-hvitt*, *svartebørs*, *svartkledd*, *svarthåret* og *svarteper*, ikke utviser noen entydig relasjonspreferanse for *svart*.

7 sammensetninger har fått uttelling for analogier. Disse er *svartand* (analogi: *svarttrost*), *svartebok* (*hvitebok*), *svarthvit*<sup>5</sup> (*gråhvit*), *svarthåret* (*rødhåret*), *svartskjorte* (*brunskjorte*), *svartspett* (*svarttrost*) og *svartbrun* (*rødbrun*). Mest tvilsom er kanskje *svart-hvit* (*gråhvit*), siden analogien som oftest betyr 'grålig hvit'. I og med at begge kan ha kopulativ betydning 'a og b', har *svarthvit* likevel fått uttelling for analogi.

Som helhet fordeler s-sammensetningene seg utover 7 av 8 trinn på gjennomsiktighetskalaen. En viss overvekt av disse har lav gjennomsiktighet. De ikke-ordbokførte sammensetningene havner i den gjennomsiktige enden. Ellers finner man her også en overvekt av adjektiv-adjektiv- og adjektiv-verb-sammensetninger. I den ugjennomsiktige enden fins en del metaforiske og metonymiske sammensetninger. Til dømes er *svart-hvitt* i NAOB definert som 'unyansert'.

#### 4.2. Tanke-

*Tanke*-sammensetningene (t-sammensetningene) konsentrerer seg mer mot sentrum av skalaen enn s-sammensetningene, med en liten overvekt i retning av ugjennomsiktighet. Forleddet *tanke* har ikke distinkt ulike betydninger i t-sammensetningene, men noe betydningsvariasjon er det definitivt å spore. Det virker til dømes rimelig at *tankeretning* bygger på en annen forståelse av tankebegrepet enn *tankegymnastikk*. Førstnevnte har en ekvivalent med forleddet *idé*-, altså *idéretning*, mens *idégymnastikk* ikke er et godt synonym til *tankegymnastikk*.

8 av 30 sammensetninger lar seg parafasere forholdsvis tilfredsstillende. Blant dem er typer med verbavlede etterledd, som *tankeoverføring* og *tankelesing*. Disse er parallelle med *overføring av tanker* og *lesing av tanker*. Videre kan *tankerekke* omsettes

5 Merk at *svarthvit* og *svart-hvitt* ikke er samme sammensetning. De har ulike oppslag i både BOB og NAOB.

til *rekke med tanker* og *tanketom* til *tom for tanker*. *Sprang i/med tanker* og *gods av tanker* er derimot lite tilfredsstillende erstatninger for *tanke­sprang* og *tanke­gods*.

Alle t-sammensetningene er motivert av betydning 1 av *tanke* i BOB, ‘tanke­virksomhet, forestilling i bevisstheten’.

Det er få t-sammensetninger der et av ledda disambiguerer det andre. De eneste tilfellene er *tanke­bygning* og *tanke­vekkende*, der forleddet *tanke-* peker ut betydninga til etterleddet.

Ingen av t-sammensetningene blir gradert som mer gjennom-siktige på bakgrunn av relasjonspreferansene til forleddet *tanke-*. De mest frekvente sammensetningene med *tanke-* i LBK er *tanke­gang*, *-vekkende*, *-gods*, *-rekke* og *-kors*. I førstnevnte tolker jeg *gang* som en nominalisering av verbet gå. Forleddet *tanke-* er dermed agens til dette verbet, mens forleddet i *tanke­vekkende* er patiens. De tre gjenstående, *tanke­gods*, *-rekke* og *-kors*, fanges ikke uten videre opp av Eieslands (2015) relasjonskategorier. Er til dømes et *tanke­kors* et kors *i* tankene, et kors *av* tanker eller et kors som er *fra* eller *tilhører* en tanke? Det er ikke åpenbart hva relasjonen er her, og jeg konkluderer derfor med at den jevne språkbruker ikke vil oppleve disse sammensetningene som mer gjennom-siktige på bakgrunn av en tydelig relasjonspreferanse for forleddet *tanke-*. Enkelte av sammensetningene, deriblant *tanke­leser*, *tanke­overføring* og *tanketom*, gir uttelling for gjennom-siktighet på bakgrunn av relasjonspreferansen til etterleddet.

4 av 30 t-sammensetninger er mer gjennom-siktige på bakgrunn av analoge sammensetninger. Disse er *tanke­bane* (analogi: *tanke­gang*), *tanke­gymnastikk* (*tanke­øvelse*), *tankemodell* (*tanke­system*) og *tanke­verden* (*forestillings­verden*).

I motsetning til de ikke-ordbokoppførte s-sammensetningene plasserer ikke de tilsvarende t-sammensetningene seg nær den gjennom-siktige enden av skalaen. Om noe har de en svak tendens til ugjenomsiktighet, og de er såleis gode ordbokkandidater etter dette kriteriet.

### 4.3. *Vandre-*

*Vandre*-sammensetningene (v-sammensetningene) fordeler seg utover den midtre delen av gjennomsiktighetskalaen. Forleddet *vandre-* uttrykker ulike nyanser i de forskjellige sammensetningene, men samtlige relaterer seg på et eller annet nivå til bevegelse eller forflytning. V-sammensetningene er langt mindre frekvente enn s- og t-sammensetningene. Den mest frekvente, *vandrehistorie*, har 72 treff i LBK. Til sammenlikning har de mest frekvente s- og t-sammensetningene *svartkledd* og *tankegang* henholdsvis 198 og 964 treff. I motsatt ende har den minst frekvente v-sammensetninga i undersøkelsen skarve 4 treff i LBK, mens tilsvarende tall for s- og t-sammensetningene er henholdsvis 29 og 18.

11 av 23 v-sammensetninger lar seg parafasere og får dermed uttelling for gjennomsiktighet. Av disse kan *vandrefolk* = *folk som vandrer*, *vandrehall* = *hall til å vandre i* og *vandrestav* = *stav til å vandre med* nevnes. På motsatt side kan *vandremandag* ≠ *mandag som vandrer*, *vandrepokal* ≠ *pokal som vandrer* og *vandrehistorie* ≠ *historie som vandrer* nevnes. I de to sistnevnte trengs tilleggene mellom *vinnere* og *mellom folk* for å få fullgode parafaseringer.

Også 6 av 23 v-sammensetninger er motivert av betydning 1 av *vandre* i BOB, 'gå rolig'. Disse er *vandremandag*, *hall*, *safari*, *-stav*, *-sløyfe* og *-tur*.

Blant v-sammensetningene er det kun én som etter min tolkning oppviser noen form for intern disambiguering ledda imellom, nemlig den mest gjennomsiktige, *vandretur*. Her peker forleddet *vandre-* ut betydning 2 av *tur* i BOB, 'kortere eller lengre reise, ferd, utflukt'.

De mest frekvente sammensetningene i LBK med *vandre-* er *vandrehistorie*, *-falk*, *-utstilling*, *-år* og *-pokal*. Med unntak av i *vandreår* er etterleddet her agens til verbet *vandre*. Etterleddet spesifiserer altså hva det er som vandrer, og vi kan i henhold til kriteriene i modellen slå fast at relasjonspreferansen til forleddet *vandre* er



å stå i et verb–agens-forhold til etterleddet. 16 av de 23 v-sammensetningene følger dette relasjonsmønsteret. Blant de negative dømene finner vi *vandrehall* (som ikke er en hall som vandrer), *vandremandag* og *vandresafari*.

5 v-sammensetninger har fått gjennomsluktighetsuttelling for mer frekvente analoge sammensetninger. Disse er *vandrefigur* (*gjennomgangsfigur*), *vandrefolk* (*flyttefolk*), *vandremotiv* (*gjennomgangsmotiv*), *vandrefugl* (*trekkfugl*) og *vandretur* (*spasertur*, *gåtur*). Analogien består i at et av ledda er realisert med et synonym (*gåtur*), eller at definisjonen til sammensetninga som helhet tilsvarende definisjonen til v-sammensetninga, slik tilfellet er med *gjennomgangsfigur*.

Det er ingen av v-sammensetningene som plasserer seg i skalaens ytterpunkter. Snarere hopper de seg opp mot sentrum, men med en betydelig skeivhet i retning lav gjennomsluktighet. Dette skyldes at de fleste av v-sammensetningene ikke bygger på den konvensjonelle betydninga av *vandre*.

#### 4.4. Sammendrag av resultatene

Resultatene fra analysen (se tabell 1) av et sammensetningsutvalg med forledda *svart-*, *tanke-* og *vandre-* viser at man for hver forleddstype kan forvente seg en viss gjennomsluktighetsvariasjon. Ingen sammensetninger får full uttelling for samtlige faktorer, og kolonnen helt til høyre på skalaen er derfor ikke inkludert i tabell 1. S-sammensetningene varierer klart mest når det gjelder gjennomsluktighet, siden disse plasserer seg utover 7 av 8 trinn på gjennomsluktighetsskalaen.

Minst gjennomsliktig					Mest gjennomsliktig	
svart-	svart-	svart-	svart-	svart-	svart-	svart-
s-e-dauden	s-bak	s-e-børs	s-e-bok	s-and	s-sladd	s-hvit
s-e-mannen	s-male	s-jord	s-sinn	s-hyll	s-smusket	s-håret
s-e-marje	s-e-per	s-e-kunst		s-or	<u>s-brent</u>	<u>s-brun</u>
s--hvitt	s-skjorte	s-kvist		s-rot	<u>s-kledd</u>	<u>s-lakkere</u>
s-katt		s-e-liste (v)		s-spett		
s-kopp		s-e-liste (s)		s-still		
s-sjuk		s-spett		s-trost		
s-sjuke		s-år		<u>s-ovn</u>		
s-syn						
	<b>tanke-</b>	<b>tanke-</b>	<b>tanke-</b>	<b>tanke-</b>	<b>tanke-</b>	
	t-flukt	t-bane	t-arbeid	t-gymna- stikk	t-leser	
	t-gods	t-korn	t-bygning	t-vekkende	t-over- føring	
	t-kors	t-lesing	t-eksperi- ment		t-rekke	
	t-retning	t-modell	t-gang		t-tom	
	t-stiller	t-strek	t-sprang			
		t-vekker	t-spredd			
		<u>t-sett</u>	t-verden			
		<u>t-smie</u>	t-virksom- het			
			<u>t-mønster</u>			
			<u>t-prosess</u>			
			<u>t-spinn</u>			
	<b>vandre-</b>	<b>vandre-</b>	<b>vandre-</b>	<b>vandre-</b>	<b>vandre-</b>	
	v-historie	v-figur	v-mandag	v-bibliotek	v-folk	
	v-pokal	v-motiv	v-sagn	v-falk	<u>v-fugl</u>	
	v-skjold	v-stjerne	<u>v-arbeider</u>	v-nyre	<u>v-stav</u>	
		<u>v-sløyfe</u>	<u>v-due</u>	v-utstilling	<u>v-tur</u>	
		<u>v-år</u>		<u>v-hall</u>		
				<u>v-maur</u>		
				<u>v-safari</u>		

Tabell 1: Gjennomsliktighetsgradering for sammensetninger med forledda *svart-*, *tanke-* og *vandre-*. Understrekede sammensetninger er ikke oppført i BOB. Markeringene (v) og (s) står henholdsvis for verb og substantiv.

## 5. Diskusjon

Gjennomsiktighetsmodellen som er presentert og anvendt her, gir en finkornet analyse av ulike sammensetninger og plasserer hver av dem på en skala. Ulike sammensetninger kan av ulike grunner bli oppfattet som gjennomsiktige, og det kan derfor være at vidt forskjellige sammensetningstyper, altså sammensetninger som har ulike kombinasjoner av ordklasser i ledda sine, får samme gjennomsiktighetsgradering, samtidig som sammensetninger av samme type kan få forskjellig gjennomsiktighetsgradering. Det er en u diskutabel styrke ved modellen at den for hver sammensetning uavhengig av type kan bedømme gjennomsiktigheten ut fra de samme faktorene. Om ulike sammensetningstyper har ulik tilbøyelighet til å være gjennomsiktige, vil dette bli fanget inn av faktorene som er innebygd i modellen, bl.a. tilleggselement og relasjonsmønster. Anvendeligheten på tvers av ordklasser er dessuten gunstig for leksikografer siden den muliggjør en arbeidsmåte der samme redaktør redigerer hele sammensetningsstrekk, som ofte kan inneholde ulike typer.

Det er liten grunn til å tro at gjennomsiktighet er systematisk fordelt på bakgrunn av ordklassetilhørigheten til ledda. Som tabell 1 viser, ligger adjektiv-adjektiv-sammensetningene *svartsjuk* og *svartbrun* i hver sin ende av gjennomsiktighetsskalaen.

Modellen tar også høyde for formlighet mellom ulike sammensetninger. Mindre frekvente sammensetninger (til dømes *svartand*) kan forstås gjennom mer frekvente sammensetninger (til dømes *svartspett*), men ikke vice versa. Dette kan føre til at *svartand* blir bedømt som mer gjennomsiktig enn *svartspett*. Semantisk er dette rimelig, for det vil nettopp være flere faktorer som bidrar til å øke gjennomsiktigheten til *svartand* enn til *svartspett*. Det viser imidlertid hvor nødvendig det er å kombinere denne modellen med en distribusjonsanalyse som fanger opp det mest aktive ordforrådet, der *svartspett* trolig går foran *svartand*. Poenget

er uansett ikke å skille mellom sammensetninger som står rett ved siden av hverandre i tabellen, men snarere å undersøke omtrent hvor en gitt sammensetning havner.

Enkelte andre resultater kan framstå som kontraintuitive. Hvorfor er til dømes *svartbrun* blant de mest gjennomsliktige sammensetningene når den har en uklar relasjon mellom ledda og derfor er flertydig mellom betydningene ‘svartaktig brun’, ‘brunaktig svart’ og ‘brun og svart’? Dette belyser muligens at modellen ikke tar nok høyde for ulike potensielle totalbetydninger. På den annen side kan man hevde at *svartbrun* er et godt døme på en sammensetning med liten leksikografisk relevans til tross for høy frekvens. Det at den er flertydig, kommer delvis av at den kan være både kopulativ og determinativ, men delvis også av lav konvensjonalitet, som dessuten kan tjene som en del av forklaringa på hvorfor ledda ikke har antatt noen spesialiserte betydninger, og hvorfor relasjonen mellom ledda er uklar.

De operasjonelle kriteriene i kapittel 3 gir langt på vei klare instruksjoner for hvordan en gitt sammensetning rent praktisk skal vektas i henhold til gjennomsliktighetsfaktorene. Faktoren relasjonsmønster er vanskeligst å operasjonalisere siden den krever distinkte relasjonspreferanser for både for- og etterleddet i hver enkelt sammensetning. I mange tilfeller er det tidkrevende å utlede disse. En mulig løsning er å operere med relasjonspreferanser med mest mulig grovmaskede kategorier.

En annen ulempe er at analysen tar utgangspunkt i ordbøker som premissleverandør for hva som er den konvensjonelle betydninga av et ord. Det er ingen garanti for at den øverste betydninga til et oppslagsord er den som oppleves som mest konvensjonell av et flertall språkbrukere. Definisjoner og betydningsrekkefølge kan være arvet fra et tidligere ordbokverk som oppstod i en helt annen språklig kontekst. Dette kan lede til gale antakelser om konvensjonalitet, som igjen leder til at man får upresise beregninger av sammensetningers motiveringsgrad.

En annen mulig svakhet ved gjennomsiktighetsmodellen er at den ikke kan anvendes mekanisk uten at leksikografens språkkunnskaper kommer i spill. Dette er til en viss grad tilfellet for alle faktorene. Ulike leksikografer kan dermed tenkes å ende opp med ulike resultater når de bruker modellen. Det er imidlertid ikke sikkert at denne variasjonen er problematisk. Når det kommer til stykket, er det nettopp dette som, i alle fall tradisjonelt, kjennetegner ordbøker – at de er sammensatt og kvalitetsvurdert av ulike redaktører med mer eller mindre ulike språklige bakgrunner og formeninger om ords betydning. At ulike redaktører kommer til ulike resultater, er en del av det som forhindrer leksikografisk ensretting, og som sørger for at mer enn ett perspektiv er representert i ordboka.

Gjennomsiktighetsmodellen er utformet i tråd med et syn på ordbøker som hevder leksikografenes kvalitetsvurderinger som vesentlige og nødvendige. Målet er dermed ikke at gjennomsiktigheten i sammensetninger skal kunne regnes ut mekanisk uten språklige vurderinger, men snarere at gjennomsiktighetsmodellen kan gi leksikografen konkrete holdepunkter for hvilke aspekter ved sammensetninger som er relevante å trekke inn når man bedømmer gjennomsiktighet.

## 6. Oppsummering og videre forskning

Den mest pålitelige evalueringa av gjennomsiktighetsmodellen er trolig et fortolkningseksperiment der man kan observere empirisk om språkbrukeres vurderinger er i tråd med modellens påstander om gjennomsiktighet. Bare med en slik metode kan man verifisere om modellens faktorer faktisk påvirker gjennomsiktighetsgraden, og avdekke om det eventuelt fins andre faktorer som er i spill.

Videre må man supplere gjennomsiktighetsmodellen med undersøkelser av den enkelte sammensetnings utbredelse på tvers av tekstsjangre, kilder og tid.

## Litteratur

### Ordbøker

BOB = *Bokmålsordboka*. Bergen: Universitetet i Bergen og Språkrådet. <<http://ordbok.uib.no>> (august 2019).

NAOB = Tor Guttu mfl. (red.): *Det Norske Akademis ordbok*. Oslo: Det Norske Akademi for Språk og Litteratur. <<https://www.naob.no>> (august 2019).

NOB = *Nynorskordboka*. Bergen: Universitetet i Bergen og Språkrådet. <<http://ordbok.uib.no>> (august 2019).

### Annen litteratur

Bakken, Kristin (1998): *Leksikalisering av sammensetninger: en studie av leksikaliseringsprosessen belyst ved et gammelnorsk diplommateriale fra 1300-tallet*. Det historisk-filosofiske fakultet, Universitetet i Oslo.

Eiesland, Eli Anne (2015): *The semantics of Norwegian noun-noun compounds: a corpus-based study*. Det humanistiske fakultet, Universitetet i Oslo.

Fjeld, Ruth V. & Lars S. Vikør (2008): *Ord og ordbøker: ei innføring i leksikologi og leksikografi*. Kristiansand: Høyskoleforlaget.

Kjelsvik, Bjørghild (2018): *Retningslinjer for sms i BOB og NOB*. Upublisert internt notat for Revisjonsprosjektet.

Knudsen, Rune L. & Ruth V. Fjeld (2013): LBK2013: A Balanced, Annotated National Corpus for Norwegian Bokmål. I: *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA 2013. NEALT Proceedings Series* 19, 12–20.

LBK = Leksikografisk bokmålskorpus. Tekstlaboratoriet, Institutt for lingvistiske og nordiske studier, Universitetet i Oslo. <<http://www.hf.uio.no/iln/tjenester/kunnskap/samlinger/bokmal/veiledningkorpus/>> (august 2019).

- Loenheim, Lisa (2019): *Att tolka det sammansatta. Befästning och mönster i första- och andraspråkstalaras tolkning av sammansättningar*. Meijerbergs arkiv för svensk ordforskning 43, Göteborgs universitet.
- Maguire, Phil, Rebecca Maguire & Arthur W.S. Cater (2010): The influence of interactional semantic patterns on the interpretation of noun–noun compounds. I: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36:2, 288–297.
- Ryder, Mary (1994): *Ordered Chaos: The Interpretation of English Noun-Noun Compounds*. University of California Publications in Linguistics 123. Berkeley: University of California Press.
- Schmid, Hans-Jörg (2015): A blueprint of the entrenchment-and-conventionalization model. I: *Yearbook of the German Cognitive Linguistics Association*, 3:1, 3–26.
- Shoben, Edward J. & Christina L. Gagné (1997): Thematic relations and the creation of combined concepts. I: Thomas B. Ward, Steven M. Smith & Jyotsna Ed Vaid (eds.): *Creative thought: An investigation of conceptual structures and processes*. American Psychological Association, 31–50.
- Svanlund, Jan (2002): Lexikalisering. I: *Språk och stil* 12, 7–45.
- Svanlund, Jan (2009): *Lexikal etablering: En korpusundersökning av hur nya sammansättningar konventionaliseras och får sin betydelse*. Acta Universitatis Stockholmiensis.

Mikkel Ekeland Paulsen  
doktorgradsstipendiat i nordisk  
Institutt for lingvistiske, litterære og estetiske studier  
Universitetet i Bergen  
Haakon Shetelig's plass 7  
NO-5007 Bergen  
Mikkel.Paulsen@uib.no