



BERLINER BEITRÄGE ZUR SKANDINAVISTIK

Titel/
title: *Grenzgänger*
Festschrift zum 65. Geburtstag von Jurij Kusmenko

Autor(in)/
author: Sven Lange

Kapitel/
chapter: »Begrepp, betydelse och mängdlära –
några provsökningar i en elektronisk thesaurus«

In: Hornscheidt, Antje/Kotcheva, Kristina/Milosch, Tomas/Rießler,
Michael (Hg.): *Grenzgänger. Festschrift zum 65. Geburtstag von
Jurij Kusmenko*. Berlin: Nordeuropa-Institut, 1. Auflage, 2006

ISBN: 3-932406-24-9
978-3-932406-24-9

Reihe/
series: Berliner Beiträge zur Skandinavistik, Bd. 9

ISSN: 0933-4009

Seiten/
pages: 177–190

Feste URL: [<http://edoc.hu-berlin.de/...>]

© Copyright: Nordeuropa-Institut Berlin sowie die Autoren

© Copyright: Department for Northern European Studies Berlin and the authors

Diesen Band gibt es weiterhin zu kaufen. This book can still be purchased.

SVEN LANGE:
Begrepp, betydelse och mängdlära –
några provsökningar i en elektronisk thesaurus

1. Inledning

Målet för projektet Thesaurus Lex är att utarbeta ett s.k. hyperlexikon¹. Ett hyperlexikon är ett elektroniskt sammanlänkat system av ordböcker och encyklopedier. I dess centrum, som ett slags dockningsenhet eller ›moderskepp‹ finns en thesaurus (dvs. en begreppsordbok). Denna är nu under utarbetande och beräknas omfatta den moderna svenskan i form av ca 80.000 ord, strukturerade enligt systemet i Rogets Thesaurus² dvs. såväl systematiskt i begreppsklasser som alfabetiskt.

Som ett led i arbetet har (den nu något föråldrade) Brings »Svenskt ord-förråd ordnat i begreppsklasser« från 1930 inscannats³ och det är i denna elektroniska version av Bring som vi ska anställa en serie provsökningar. Härigenom hoppas vi få en bättre uppfattning om vilka sökalgoritmer den uppdaterade elektroniska produkten så småningom ska förses med.

Ord kan jämföras med varandra på många sätt. Det ena kan till exempel prövande sammanhållas med ett annat för att konstatera begreppslig eller betydelsemässig likhet. Att två ord har en begreppslig relation betyder dock inte nödvändigtvis att de är synonyma. De kan uttrycka olika aspekter av ett och samma begrepp men sinsemellan vara långt ifrån liktydiga. Det är om förhållandet begrepp och betydelse som den här artikeln ytterst ska handla.

Theasaurusanvändaren kan starta från såväl ordet som begreppsklassen. Låt oss säga att han/hon väljer ordet *överblicka* som utgångspunkt. I Bring återfinns ordet i sammanlagt två begreppsklasser (av 1000), dels i 441 ›Synsinne‹ i ett stycke (paragraf) innehållande verb:

(be, genom, å-, åter-, över-)se, (be-, genom-, å-, över-)skåda, (över-)blicka, syn-a (-as) betrakta osv.

Dels i 457 ›Uppmärksamhet‹:

..., visitera, mönstra, över-vaka (-blicka), efter- (till-)se, bevaka osv.

1 LANGE och TÖRNQVIST: 2003, 181–190.

2 ROGET: 1852.

3 BRING: 1930.

Som synes har vi inga ordförklaringar eller definitioner i texten. Många ord tillhör bara en enda begreppsklass medan andra kan lyda under långt över tio stycken. Det vi ska försöka komma underfund med i den här artikeln är hur vi ska ringa in ordbetydelser och betydelselikhet genom att utnyttja ordens begreppstillhörighet i thesaurusen.

En thesaurus är ingen synonymordbok även om synonymin på sätt och vis är implicerad och inbegripen genom ordens begreppstillhörighet. Thesaurusen är inriktad mot produktion. Den ska ge tankeimpulser, hjälp att tänka i nya banor, vägledning i för användaren tidigare okänd ordteräng osv. Så fungerar inte en synonymordbok. Den förutsätter i stället att satsen eller meningen formulerats i ett första steg och att den som talar eller skriver i ett andra steg byter ut ett ord mot ett annat i en fixerad, linnear sträng. Den vill servera synonymer, dvs. ge en lista av ord som är ›utbytbara i kontext‹ i en redan formulerad sträng. Vår elektroniska thesaurus däremot ska vid behov kunna ge en betydelsemässigt rangordnad lista av begreppsmässigt länkade ord som ska vara själva utgångspunkten för tanke och formulering. Orden är steg 1 i tankebearbetningen inte steg 2 i en korrigering. Hur denna lista ska tas fram ska utprövas i den här artikeln.

2. Metod

Inget ord i thesaurusen står isolerat. Samtliga tillhör en eller flera begreppsklasser och är därmed länkade till andra ord. Ord som tillhör samma begreppsklass är begreppsligt relaterade till varandra. Så t.ex. är ordet *vårflod* länkat till alla ord i begreppsklassen 348 ›Vattenflöde‹ och ordet *upprättelse* till samtliga ord i de sexton begreppsklasser som det är knutet till.

Att orden i thesaurusen är länkade till varandra (delar minst en begreppsklass) innebär att de begreppsligt överlappar varandra – mer eller mindre. Begreppslighet utgör således ett slags mängdobjekt. Två ord som är begreppsligt länkade till varandra bör därför kunna studeras med mängdalgebraisk metod. Artikeln vill pröva ett sådant analysförfarande. Som påpekats är siktet inställt på att få ett underlag för tillskapandet av ändamålsenliga sökrutiner i den planerade elektroniska thesaurusen i det här fallet för att åstadkomma en lista av ord rangordnade efter betydelselikhet med ett utgångsord.

3. Material

Brings begreppsordbok har en systematisk indelning i 1000 begreppsklasser (i likhet med tidigare utgåvor av Rogets Thesaurus) och ett ordförråd på ca 60.000 ord, fördelade på dessa. Orden återfinns tillsammans med sina begreppsklassnummer i det alfabetiska registret. En begreppsklass är en tanke- eller idékategori. Orden (eller snarare deras betydelser) uttrycker begrepp, ett eller flera. Vi kan tänka oss thesaurusen grafiskt illustrerad av ett rutnät med orden (från *a* till *övärld*) på y-axeln och begreppen (från nr 1 Tillvaro till nr 1000 Kyrkobyggnad) på x-axeln. Ett rutnät således bestående av 60.000.000 rutor eller celler. Varje ord har minst ett nedslag i någon ruta (cell), de flesta i flera (men sällan fler än i tio). Ordet *förstörelse* är ord nr 16.694 på y-axeln i vår elektroniska Bring och uttrycker de begreppsklasser som representeras av rutorna/cellerna 2, 49, 146, 162, 241, 649, 659, 663, 735, 756, 782 och 830 på x-axeln. Men inte nog med det. De ca 60 miljoner rutorna är (oftast) i sin tur uppdelade i underavdelningar, s.k. paragrafer.⁴ Ordet *förstörelse* ligger t.ex. i x-axelns andra kolumn, dvs. i begreppsklass 2 ›Intighet‹, närmare bestämt i den andra av de underavdelningar som endast innehåller substantiv (= s2). Det finns också v-paragrafer (verb) samt a-paragrafer (adjektiv, adverb och ibland idiom) i de flesta av de tusen begreppsklasserna.

Paragraferna består av räckor eller sekvenser av ord. Ordet *förstörelse* återfinns bl.a. i följande ordsekvens: *tillintetgörelse, undergång, utsläckning, utrotning, död, fördärv, förstörelse, förstöring, ödeläggelse, ödslighet, tabula rasa, nihilism, nirvana*. Orden i sekvensen föreslår jag att vi positionsbestämmer med samma teknik som vi tillämpar i en konkordansrad. Utgångsordet får siffran 0; orden till vänster minusnumreras (*fördärv* -1, *död* -2 osv.). Orden till höger plusnumreras (*förstöring* +1, *ödeläggelse* +2 osv.).

4. Val av utgångsord och jämförelseord

Utgångspunkten för våra provsökningar är ordet *förstörelse*. Det representerar *utgångsordet* i den kommande jämförelsen. Ett ord är anknutet till minst en av de 1000 begreppsklasserna. Ordet *förstörelse* finns i Bring noterat för 12 olika begreppsklasser. Vi säger därför att ordet har en

4 LANGE: 2000.

›begreppslatitud« på 12. För att det ska vara intressant att jämföra två ord med varandra måste de ha minst en (1) begreppsklass gemensam. Ord är begreppsmässigt länkade om de delar åtminstone en begreppsklass. Med detta i åtanke går vi nu på jakt efter de ord i thesaurusen som är länkade till utgångsordet *förstörelse*.

Vi börjar med en sökning i vår databas, där de ca. 60.000 orden ligger kodade med sina begreppsklasser. Databasens första fem kolumner är:

- A Nr
- B Ord
- C Förklaring (inskjutna förklaringar i Brings alfabetisk register)
- D Ordklass (s, v eller a)
- E Begreppsklasser

För att få en ordsamling där orden är begreppsligt länkade och tillhöriga samma ordklass som utgångsordet *förstörelse* villkorar vi sökningen med ett ›s« i kolumn D. Därefter söker vi de ord som i likhet med *förstörelse* har begreppsklasserna 2 eller 49 eller 146 eller 162 osv. (Kolumn E). Sökningen ger 1182 olika ord.

Dessa båda steg har alltså gett oss en samling ord, samtliga länkade (minst en begreppsklass gemensam) med vårt utgångsord *förstörelse*.

Nu kan vi lätt notera:

- F antal begreppsklasser som ett jämförelseord ur denna samling har.
- G vilka begreppsklasser som jämförelseordet har gemensamt med utgångsordet (*förstörelse*.)

och dessutom fastställa:

- H Jämförelseordets antal gemensamma begreppsklasser med utgångsordet.
- I Jämförelseordets antal icke-gemensamma begreppsklasser med utgångsordet
- J Den procentuella andelen begreppsklasser som jämförelseordet delar med utgångsordet.
- K Den procentuella andelen begreppsklasser som jämförelseordet icke delar med utgångsordet.

Vi antar att för graden av betydelselikhet (synonymi) mellan två begreppsligt länkade ord spelar följande faktorer in (som vi förklarar närmare under resans gång). Dock okänt hur.

- jämförelseordets andel av utgångsordets begreppslatitud
- snittmängden
- differenserna $A \setminus B$ och $B \setminus A$
- paragraftillhörigheten
- position i paragrafens ordsekvens

5. Jämförelseordets andel av utgångsordets begreppslatitud

Det verkar rimligt att tänka sig att ett ord som har högre gemensam andel av utgångsordets begreppslatitud också är mer synonymt med detta än ett annat med lägre gemensam andel.

Låt oss därför sortera de 1182 begreppsmässigt länkade substantiven efter vilka som har högst procentuell gemensam andel begreppsklasser med utgångsordet *förstörelse* (= J). Resultatet framgår av tabell 1 som vi kan läsa nerifrån och upp: 901 ord har endast en begreppsklass gemensam med utgångsordet *förstörelse*. Det ger en andel av utgångsordets begreppslatitud på 8,3 %. 208 ord har 2 begreppsklasser gemensamma med utgångsordet, vilket ger 16,7 % av dess begreppslatitud. Endast 16 ord har en andel över 25 % av begreppslatituden hos *förstörelse*.

Andel gem b-lat med u-ord i %	Antal gem begreppsklasser m u-ord	Antal ord	Rang
66,7	8	1	1
58,3	7	1	2
50	6	2	3
41,7	5	2	4
33,3	4	10	5
25	3	57	6
16,7	2	208	7
8,3	1	901	8

Tabell 1: Antal ord länkade till utgångsordet *förstörelse* och deras andel av dess betydelselatitud

De 16 orden med andel över 25 % är följande:

fördärv (66,7 %), förstöring (58,3 %), skada, ödeläggelse (50 %), förbannelse, undergång, (41,7 %), brist, fall, plåga, plågoris, ruin, sammanstörtning, skadegörelse, skeppsbrott, tillintetgörelse, upplösning (33,3 %).

Hur stämmer denna lista med de synonymförslag till *förstörelse* som vi får från en konventionell synonymordbok? »Ord för ord« ger följande:

förstörelse förstöring, se *ödeläggelse*, se *förödelse*⁵

Ett uppslag på *ödeläggelse* ger:

ödeläggning, tillintetgörelse, utplåning, bortträdning, förödelsens styggelse, sabotage, förödelse

5 PALMÉR: 1977.

Ett uppslag på *förödelse* ger:

härjning(ar), skövling, förhärjning, vandalism, katastrof, tabelras, skada, skadegörelse, åverkan

Om vi jämför våra listor noterar vi att endast 5 av de 18 givna synonymerna från synonymordboken finns med bland de 16 på vår lista med ord med andel över 25 % av begreppslatituden hos *förstörelse* från vår sökning i den elektroniska Bring. De fem representerade orden är *förstöring* (58,3 %), *ödeläggelse*, *skada* (50 %), *tillintetgörelse* samt *skadegörelse* (33,3 %). På vilka procentnivåer hamnar då resten av synonymordbokens förslag? En mönstring visar att *katastrof* i Bring har en 25-procentig andel av begreppslatituden hos *förstörelse*. Tre ord (*förödelse*, *förhärjning*, *skövling*) hamnar på en andel på 16,7 % och fem ord (*härjning(ar)*, *sabotage*, *tabelras*, *utplåning*, *åverkan*) på en andel av 8,3 %. Fyra ord från synonymordbokens lista (*bortradering*, *förödelsens styggelse*, *vandalism och ödeläggning*) saknar länk (dvs. minst en gemensam begreppsklass) till *förstörelse* i Bring.

Slutsatsen måste således bli att jämförelseord med hög andel (exempelvis över 25 % som i vårt experiment) av utgångsordets begreppslatitud inte på något entydigt sätt är mer synonyma med utgångsordet än ord på lägre procentnivåer. Ja, synonymerna kan tydligen finnas på alla nivåer, även om ett visst samband tycks finnas.

Vi går därför vidare och närmast med en mer detaljerad granskning av några av orden med andelar över 25 % av begreppslatituden hos *förstörelse*. Vi börjar med det jämförelseord som har den allra högsta andelen av utgångsordets begreppslatitud, nämligen *fördärv* med 66,7 %. Det bör anmärkas att ordet inte fanns med på listan från synonymordboken »Ord för ord«.

Inledningsvis analyserar vi ordens snitt ($A \cap B$, gemensamma begreppsklasser). Analysen innefattar bl.a. en kartläggning av i vilken utsträckning de två orden uppträder i samma paragrafer. Om så är fallet bestämmer vi ordens positioner visavi varandra i de delade paragrafernas ordsekvenser. Därefter anställer vi också några reflexioner om differenserna $A \setminus B$ och $B \setminus A$. Förhoppningsvis kan vi härigenom komma närmare målet att få den elektroniska thesaurusen (databasen i den planerade »Thesaurus Lex«) att på automatisk väg skapa en godtagbar rangordnad lista av begreppsligt länkade ord efter grad av betydelselighet med en vald glosa (= utgångsord).

5.1 Begreppsklassanalys av utgångsord och jämförelseord

Jämförelseordet *fördärv* har 16 begreppsklasser. Åtta av dessa är gemensamma med utgångsordets (*förstörelse*) 12 begreppsklasser. Det betyder att jämförelseordet *fördärv* har en andel på 66,7 % (8 av 12) av utgångsordets (*förstörelse*) begreppsplatitud (= uppsättning av begreppsklasser).

Låt oss ta tabell 2 som utgångspunkt:

$A \setminus B$	$A \cap B$	$B \setminus A$
4	8	8

Tabell 2: Mängd för A. *förstörelse* och B. *fördärv*

Av tabell 2 framgår att $A \cup B$ är 20 (4+8+8). Den ger också vid handen att $A \cap B$ utgörs av åtta element (begreppsklasserna: 2, 146, 162, 649, 659, 663, 735, 830). Differensen $A \setminus B$ består av fyra element (begreppsklasserna: 49, 241, 756 och 782). Differensen $B \setminus A$ består av åtta element (begreppsklasserna: 67, 619, 645, 651, 732, 776, 828 och 945).

Som vi tidigare påpekat har jämförelseordet (*fördärv*) således en gemensam andel av utgångsordets begreppsplatitud på 66,7 % (8 av 12). Men det finns också en 33,3-procentig andel begreppslighet hos *förstörelse* (differensen $A \setminus B$) som *fördärv* inte delar (4 av 12).

Vi kan också notera att *fördärv* har en andel av sin begreppsplatitud som utgångsordet *förstörelse* saknar helt (differensen $B \setminus A$). Den andelen uppgår till 50 % (8 av 16).

5.2 Paragrafanalys av utgångsord och jämförelseord i snittet

Paragrafanalysen ska redogöra för i vilken utsträckning som utgångsord och jämförelseord delar paragrafer samt i förekommande fall på vilket avstånd de ligger från varandra i ordsekvensen. Paragrafanalysen kan givetvis utföras enbart på de paragrafer som utgångsordet och jämförelseordet har gemensamma (dvs. på snittet). I paragrafens ordsekvens ligger som påpekats ovan utgångsordet på plats 0 (noll). Jämförelseordet står antingen på position vänster (minus) eller höger (plus) om utgångsordet.

Av tabell 3 nedan framgår att orden *förstörelse* och *fördärv* ligger i samma paragrafer. När det gäller position i paragrafens ordsekvens är det kortaste avståndet 0 (= noll; i begreppsklass 830, *för-därv* (-störelse)). Det längsta är +21 (i begreppsklass 659). Snittvärdet är 5,1 (plus- eller minusposition är troligen irrelevant för snittvärdet).

	Utgångsord	Jämförelseord
	§	§ och position
Begreppsklass	<i>förstörelse</i>	<i>fördärv</i>
2 Intighet	s2	s2 -1
146 Omstörtning	s3	s3 -2
162 Förstöring	s1	s1 +8
649 Ont	s1	s1 -6
659 Försämring	s1	s1 +21
663 Skadlighet	s1	s1 -1
735 Olycka	s3	s3 -2
830 Obehag	s2	s2 +/-0

Tabell 3: Paragraftillhörighet för utgångsordet *förstörelse* och jämförelseordet *fördärv* samt jämförelseordets position i paragrafens ordsekvens

Vi kan också notera att placering i den första av flera substantivparagrafer (dvs. i s1) antyder att ordet ligger i den ordsekvens vars ord mest generellt uttrycker begreppsklassen och betydelsemässigt ligger närmast begreppsklassenamnet. Paragraferna 2, 3 och därefter är snarast sidoordnade specificeringar av olika slag.

5.3 Analys av differenserna A\B och B\A

5.3.1 Differensen A\B

Differensen A\B, avser fyra element (begreppsklasser).

Begreppsklass	§
49 Splittring	s2
241 Formlöshet	s1
756 Annulering	s1
782 Mistning	s3

Tabell 4: Paragraftillhörighet för differensen A\B

Differensen A\B står för de begreppsklasser där thesaurusen föreslår *förstörelse* men inte *fördärv*. Dessa representerar tillsammans ungefär en tredjedel av begreppsplatituden hos *förstörelse*. Av tabell 4 framgår i vilka paragrafer ordet hör hemma i respektive begreppsklass. Här har vi alltså en indikation på att betydelskillnad föreligger mellan de två orden. Vi kan vidare konstatera att *förstörelse* i två fall ligger i den första substantivparagrafen – i begreppsklasserna ›Formlöshet‹ och ›Annulering‹. Att ett ord ligger i den första substantivparagrafen betyder att det ligger i samma paragraf som begreppsnamnet självt. Denna paragrafhemvist ger, som vi

påpekade ovan, en viss antydning om att ordet höra hemma mer i begreppets generella del än i dess mer specifika. Ordet *förstörelse* är således ett inte alltför långsökt ord att ganska allmänt uttrycka ›Formlöshet‹ och ›Annulering‹ med. Ju större differensen $A \setminus B$ är desto fler kontexter existerar där *förstörelse* inte är utbytbar mot *fördärv*.

5.3.2 Differensen $B \setminus A$

Differensen $B \setminus A$, avser åtta element. Av tabell 5 framgår vilka begreppsklasser som *fördärv* har som inte *förstörelse* delar. Tabellen visar också i vilka paragrafer ordet hör hemma.

Begreppsklass	§
67 Slut	s4
619 Nackdel	s1
645 Gagnlöshet	s4
651 Ofullkomlighet	s3
732 Misslyckande	s2
776 Förlust	s3
828 Lidande	s3
945 Last	s1

Tabell 5: Paragraftillhörighet för differensen $B \setminus A$

Vi kan konstatera att differensen $B \setminus A$ är dubbelt så stor som differensen $A \setminus B$, vilket betyder att *förstörelse* har en mindre andel av begreppslatituden i ordet *fördärv* (50 % 8 av 16) än vad *fördärv* har av ordet *förstörelse* (66,7 % 8 av 12).

Vi kan också notera att *fördärv* har lika många begreppsklasser i snittet som i differensen (8 st). Men en närmare jämförelse mellan snittet och differensen $B \setminus A$ visar att ordet ligger i fler s1-paragrafer i snittet (4 av 8 i snittet mot 2 av 8 i differensen).

Eftersom ordet har lika många begreppsklasser i snittet som i differensen $B \setminus A$ men fler förekomster i s1-paragrafer i snittet antyder det att ordet i snittet uttrycker allmännare innebörder i begreppen och i differensen mer specificerande.

5.4 Slutsats (*förstörelse* – *fördärv*)

Undersökningen kommen så här långt visar tyvärr inte vilken inverkan ›grad av överensstämmelse i fråga om paragraftillhörighet‹ har, eftersom orden ligger i samma paragrafer i mängdsnittet. Tidigare har vi som bekant

inte tillmätt faktorn ›hög andel av utgångsordets begreppslatitud‹ någon avgörande betydelse för betydelselikhet.

Vi måste därför som nästa logiska steg välja ut och jämföra två jämförelseord som har samma procentuella andel av utgångsordets begreppslatitud men som förhoppningsvis inte överensstämmer i fråga om paragrafhemvist med utgångsordet. Ur listan som vi presenterade ovan väljer vi *undergång* och *förbannelse*.

6. Delexperiment

förstörelse – undergång och förstörelse – förbannelse.

Orden *undergång* och *förbannelse* har samma andel av utgångsordets begreppslatitud, nämligen 41,7 %. Genom analysen av ordet *fördärv* ovan kunde vi som vi såg inte komma åt vilken inverkan paragraftillhörigheten har på betydelselikheten eftersom utgångsord (*förstörelse*) och jämförelseord (*fördärv*) låg i samma paragrafer. Genom att nu hålla ›andel av utgångsordets begreppslatitud‹ konstant men förhoppningsvis ha variation när det gäller paragrafhemvist i förhållande till utgångsordet skulle vi kunna få ett bättre underlag för ett avgörande.

Tabellerna 6 och 7 åskådliggör mängdförhållandena visavi utgångsordet *förstörelse* för jämförelseorden *undergång* och *förbannelse*.

$A \setminus B$	$A \cap B$	$B \setminus A$
7	5	9

Tabell 6: Mängd för A. *förstörelse* och B. *undergång*.

$A \setminus B$	$A \cap B$	$B \setminus A$
7	5	1

Tabell 7: Mängd för A. *förstörelse* och B. *förbannelse*.

Av tabellerna framgår att såväl *undergång* som *förbannelse* vardera förfogar över 5 av utgångsordets 12 element, dvs. har en andel på 41,7 % av utgångsordets begreppslatitud. Den stora skillnaden ligger i differensen $B \setminus A$: 9 element för *undergång* respektive 1 element för *förbannelse*.

Låt oss nu granska hur paragrafhemvist och position i ordsekvens ser ut för de båda jämförelseorden visavi utgångsordet och därmed kanske få vår fråga om hur de olika faktorerna inverkar på betydelselikheten bättre besvarad.

Vi börjar med *undergång*.

6.1 Förstörelse – undergång

Om vi enbart håller oss till snittet och efterser i vilken grad orden tillhör samma eller olika paragrafer kan vi notera i tabell 8 att jämförelseordet *undergång* ligger i samma paragrafer som utgångsordet *förstörelse* (5 fall av 5). Det kan vi uttrycka som att jämförelseord och utgångsord har ›hundra-procentig överensstämmelse i snittets paragrafstruktur‹.

	Utgångsord	Jämförelseord
	§	§ och position
Begreppsklass	<i>förstörelse</i>	<i>undergång</i>
2 Intighet	s2	s2 -5
146 Omstörtning	s3	s3 +1
649 Ont	s1	s1 -4
735 Olycka	s3	s3 +3
830 Obehag	s2	s2 +2

Tabell 8: Paragraftillhörighet för utgångsordet *förstörelse* och jämförelseordet *undergång* samt jämförelseordets position i paragrafens ordsekvens.

Av tabell 8 framgår också på vilket avstånd jämförelseordet ligger från utgångsordet, dvs. jämförelseordets position i paragraferna. Avståndet borde stå i omvänd proportion till betydelselikheten. Antagandet försvagas dock betänkligt av det faktum att paragraferna (hos Bring) inte är mekaniskt uppställda synonymräckor. I begreppsklass 649 ›Ont‹ där båda orden hamnar i paragraf s1 ser sekvensen ut så här:

Ont: (...) *fördärv (-lighet)*, *försämring*, ***undergång***, *död*, *förstöring (sbeğär, -sverk)*, *förstörare*, *förstörelse* (...);

och med positionerna tydligare angivna så här:

-4	-3	-2	-1	0	
<i>undergång</i> ,	<i>död</i> ,	<i>förstöring (-sbeğär, -sverk)</i> ,	<i>förstörare</i> ,	<i>förstörelse</i>	o.s.v.

Ordet *förstörare* (-1) bryter som synes synonymsekvensen *undergång*, *död*, *förstöring*, *förstörelse*. Med detta i åtanke måste vi därför tills vidare förhålla oss tämligen skeptiska till tanken att ett kortare avstånd mellan utgångs- och jämförelseord i ordsekvensen automatiskt skulle medföra högre grad av betydelselighet än ett längre. Hursomhelst, i det här fallet har vi ett snittvärde på 3 ($5+1+4+3+2=15/5=3$).

Låt oss därefter övergå till *förbannelse*.

6.2 Förstörelse – förbannelse

Ordet *förbannelse* ligger i samma paragrafer som utgångsordet *förstörelse* i två av fem fall. Som framgår av tabell 9 gäller det begreppsklasserna ›Skadlighet‹ och ›Obehag‹.

	Utgångsord	Jämförelseord
	§	§ och position
Begreppsklass	<i>förstörelse</i>	<i>förbannelse</i>
162 Förstöring	s1	s5
649 Ont	s1	s8
663 Skadlighet	s1	s1 +4
735 Olycka	s5	s2
850 Obehag	s2	s2 -5

Tabell 9: Paragraftillhörighet för utgångsordet *förstörelse* och jämförelseordet *förbannelse* samt jämförelseordets position i paragrafens ordsekvens.

Det är också naturligtvis så att det är bara i de fall där orden ligger i samma paragrafer som det är meningsfullt att granska positionerna och därmed synonymräckorna närmare. Och då kan vi notera i samma tabell att jämförelseordet i förekommande fall (›Skadlighet‹ och ›Obehag‹) ligger på en fjärde- (+4) och en femteposition (-5).

Ordet *förbannelse* har bara ett enda element (en begreppsklass) i sin differens B\A. Intressant nog representeras det av begreppsklassen ›Förbannelse‹. Vi ser alltså att ordet och begreppsklassbeteckning sammanfaller. Det bör nog tolkas så att ordets kärnbetydelse ligger utanför snittet, dvs. de element (uppsättning av begreppsklasser) som ordet har gemensamma med utgångsordet *förstörelse*.

7. Preliminära slutsatser

Snittmängden säger att utgångsordet och jämförelseordet kan uttrycka samma begrepp. Att orden är begreppsligt länkade (= har minst en begreppsklass gemensam) säger emellertid inte att de är synonyma i den vedertagna lingvistiska meningen ›utbytbara i språklig kontext‹. Eftersom vår thesaurus helt saknar orddefinitioner, sådana som vi möter i vanliga alfabetiska ordböcker som »Nationalencyklopedins ordbok (NEO)« och »Svensk Ordbok« m.fl. är det angeläget att fastställa i vilken mån orden i den elektroniska thesaurusen enbart med hjälp av den systematiska makrostrukturen ändå låter sig rangordnas efter betydelselikhet.

Det är ganska troligt att en thesaurusbrukare vill få fram en lista av synonymer till ett visst ord. Resultatet i undersökningen pekar på att det mest ändamålsenliga är då att bygga svaret på de ord som är länkade till utgångsordet och som är av samma ordklass. De jämförelseord med ›högre andel av utgångsordets begreppslatitud‹ bör beaktas men inte tillmätas avgörande betydelse i fråga om betydelselikhet.

Om två ord har samma procentuella andel av utgångsordets betydelselatitud bör det ord som har flest paragrafer gemensamma med utgångsordet rankas högre. Ordet *undergång* med hundra procentig paragraföverensstämmelse med *förstörelse* skall hamna högre på listan än *förbannelse* med blott fyrtioprocentig paragraföverensstämmelse. Detta bekräftas för övrigt också av samtliga konsulterade synonymordböcker.

I fråga om ›position i ordsekvens‹ är det svårare att dra några säkra slutsatser. Om allt annat för övrigt är lika för två jämförelseord, dvs. om de har samma andel gemensam begreppslatitud med utgångsordet och samma antal gemensamma paragrafer med utgångsordet kan vi möjligtvis räkna med att genomsnittsvärdet för position i paragrafsekvens kan fälla utslag. Ett ord med lägre snittvärde skulle i så fall rankas högre i fråga om betydelselikhet än ett ord med högre snittvärde. Något sådant fall har vi inte ännu undersökt.

Även i fråga om differenserna och deras betydelse för rankningen återstår en del forskning. Men analysen av ordet *förbannelse* visar ändå tydligt att ett jämförelseord med långt fler begreppsklasser i snittet än i differensen B\A kan detta till trots ha sin ›kärnbetydelse‹ förlagd till differensen B\A. Det minskar drastiskt möjligheten till synonymi mellan jämförelseord och utgångsord. Det visar också att ett ord noterat för många begreppsklasser i thesarusen inte har sin betydelse ›utsmetad‹ i lika proportioner över samtliga begreppsklasser det tillhör. Dess kärnbetydelse kan vara förlagd till ett fåtal, eller kanske en enda av dessa. Kommande experiment ska inriktas till att ytterligare belysa paragrafhemvist och differensernas inverkan på frågan om betydelselikhet närmare.

LITTERATUR

- ALLÉN, Sture et al.: *Svensk Ordbok*. Utarbetad vid Språkdata, Göteborgs universitet. 3., rev. uppl., Stockholm: Norstedts ordbok, 1999.
- BRING, Sven Casper: *Svenskt ordförråd ordnat i begreppsklasser*. Stockholm: Geber, 1930.
- LANGE, Sven och Lars TÖRNQVIST: »Thesaurus Lex»: En ny svensk elektroniskt länkad thesaurus.« I: HANSEN, Zakarias Svabo och Annfinnur JOHANSEN (red.): *Nordiske studier i leksikografi 6. Rapport fra Konference om leksikografi i Norden, Tórshavn 21.–25. august 2001*. (= *Skripter udgivna av Nordiska föreningen för leksikografi*; 7), Tórshavn: Nordisk forening for leksikografi, 2003 181–190.
- LANGE, Sven: *Tesaurusen i ordförrådsforskningens tjänst*. (= Rapporter från ORDAT; 2), Göteborg: Göteborgs universitet, Institutionen för svenska språket, 2000.
- Nationalencyklopedins ordbok 1–3*. Utarbetad vid Språkdata, Göteborgs universitet. Höganäs: Bra Böcker, 1995–1996.
- PALMÉR, Johan: *Ord för ord: svenska synonymer och uttryck*. 2., moderniserade uppl., utarb. av Herbert FRIEDLÄNDER. Stockholm: Norstedt, 1977.
- ROGET, Peter Mark: *Thesaurus of English Words and Phrases*. 1852.