

Sammlaren

Tidskrift för forskning om
svensk och annan nordisk litteratur
Årgång 139 2018

I distribution:
Eddy.se

Svenska Litteratursällskapet

REDAKTIONSKOMMITTÉ:

Berkeley: Linda Rugg
Göteborg: Lisbeth Larsson
Köpenhamn: Johnny Kondrup
Lund: Erik Hedling, Eva Hættner Aurelius
München: Annegret Heitmann
Oslo: Elisabeth Oxfeldt
Stockholm: Anders Cullhed, Anders Olsson, Boel Westin
Tartu: Daniel Sävborg
Uppsala: Torsten Petterson, Johan Svedjedal
Zürich: Klaus Müller-Wille
Åbo: Claes Ahlund

Redaktörer: Jon Viklund (uppsatser) och Sigrid Schottenius Cullhed (recensioner)
Biträdande redaktör: Niclas Johansson och Camilla Wallin Bergström

Inlagans typografi: Anders Svedin

Utgiven med stöd av Vetenskapsrådet

Bidrag till *Samlaren* insändes digitalt i ordbehandlingsprogrammet Word till info@svelitt.se. Konsultera skribentinstruktionerna på sällskapets hemsida innan du skickar in. Sista inlämningsdatum för uppsatser till nästa årgång av *Samlaren* är 15 juni 2019 och för recensioner 1 september 2019. *Samlaren* publiceras även digitalt, varför den som sänder in material till *Samlaren* därmed anses medge digital publicering. Den digitala utgåvan nås på: <http://www.svelitt.se/samlaren/index.html>. Sällskapet avser att kontinuerligt tillgängliggöra även äldre årgångar av tidskriften.

Svenska Litteratursällskapet tackar de personer som under det senaste året ställt sig till förfogande som bedömare av inkomna manuskript.

Svenska Litteratursällskapet PG: 5367–8.

Svenska Litteratursällskapets hemsida kan nås via adressen www.svelitt.se.

ISBN 978–91–87666–38–4

ISSN 0348–6133

Printed in Lithuania by
Balto print, Vilnius 2019

Språklig rytm i skönlitterär prosa

En fallstudie i Karin Boyes Kallocain

AV SARA STYMNE, JOHAN SVEDJEDAL

& CARIN ÖSTMAN¹

Problemet

Kallocain (1940) var Karin Boyes (1900–1941) femte och sista roman. Redan vid utgivningen blev den uppmärksammas och med åren har den blivit hennes mest spridda och omdebatterade roman.

Kallocain är en dystopi om en totalitär framtid och människors försök att frigöra sig från ett yttre förtryck som också har blivit internaliserat. Huvudpersonen är kemisten Leo Kall, fullt lojal mot de centralistiska och militaristiska stämningarna i Världsstaten – ett industriellt övervakningssamhälle med drag av både tidens Sovjetunionen och Tyskland. Leo Kall uppfinner en ”sanningsdrog”, benämnd kallocain, ett preparat som när det injiceras får människor att tala utan hämningar och avslöja sina inre hemligheter. Frispråkigheten inkluderar naturligtvis tvivel på statens linje i olika frågor och preparatet kan därför användas för att avslöja ”samhällsfiender”. Myndigheterna sätter igång försök med kallocain i stor skala, understödda av den statslojale Leo Kall. Men under händelsernas gång börjar han förlora tron på statens goda intentioner, uppmuntrad därtill av sin tvivlande hustru Linda och sin kollega Edo Rissen. Han får också vetskap om en hemlig sekt som i den hemlighetsfulla ”ökenstaden” prövar sätt att leva som präglas av en annan tillit och frihet än i Världsstaten. Leo Kall känner sig komma i en ny och oväntad kontakt med livets och växandets krafter.

Alltsammans slutar tragiskt. Men innan dess har Leo Kall genomgått en rejäl personlighetsförändring, som möjligen också påverkar romanens själva rytm. Lingvisten och vershistorikern Olof Gjerdmann har lanserat en intressant teori om effekterna av detta. Till sammanhanget hör att Gjerdmann som akademisk lärare i nordiska språk 1924 gav en experimentell kurs i fonetik, gällande kvantitets- och tonalitetsundersökningar, där Boye deltog. De diskuterade språkrätm och han noterade hennes starka intresse för frågan.² (Boye gav sedan själv några glimtar från kursen i essän ”Tendens och verkan”). När Gjerdmann behandlade ämnet i en minnesbok om Boye, pekade han också på hennes teoretisk-estetiska essäer, där hon diskuterar sambandet mellan per-

sonlighet och stil. Det finns goda skäl att anta att Boye ansåg det finnas nära förbindelser mellan språklig rytm och en personlighets egenart.

Detta perspektiv får Gjerdmann att diskutera rytmförändringar inom Boyes författarskap, men också inom *Kallocain*. Kort uttryckt menar han att det i hennes lyriska författarskap uppträder en ”avgjort böjligare” allmänrytm i hennes fjärde diktsamling, *För trädets skull* (1935), än i hennes tre tidigare.³ En liknande rytmisk förändring urskiljer han i *Kallocain*:

Samma allmänrytm som i ”För trädets skull” förekommer också i en god del av Karin Boyes prosa från senare år. Omkring år 1934 tycks den vara helt genomförd. Den fastare rytmen och den därmed ofta förbundna svalare tonen är dock inte därför definitivt borta ur hennes diktning. Den finns i ”Kallocain” men märkligt nog inte genomgående, utan första hälften av romanen har den stramare, senare hälften den böjligare, mera rörliga rytmen. Denna tar vid sid. 182, rad 10.⁴

Den scen i romanen som Gjerdmann avser är ett tillfälle när Leo Kall – helt oväntat – reser sig upp vid ett offentligt möte och yttrar några högst utmanande meningar. I romanens värld betecknar scenen hans ”omvändelse”, ett genombrott för hans tvivel på samhället som är så plötsligt och överraskande att Kall själv inte riktigt kan tolka det.

Gjerdmans påstående är djärvt – om än utifrån ordentlig bekantskap med Boye och hennes verk – men har egentligen inte följts upp i den omfattande forskningen kring *Kallocain*.⁵ Undantaget är Staffan Björck, som i *Romanens formvärld* (1953) nämner Gjerdmans hypotes och menar att den ”torde innehålla en viss sanning”. Skälet ser han i författarens inlevelse i huvudpersonen: ”att Karin Boye så starkt identifierat sig med Leo Kall i hans utveckling, att det ögonblick s. 182 som slår en första reva i hans plikt- och skräckpansar också bräcker den spända hållningen i hennes egen stil.”⁶

En grundläggande forskningsfråga är därmed om det stämmer att det går en gräns mellan en ”stram” respektive en ”böjlig” och mer ”rörlig” rytm” tvärsigenom *Kallocain*. Förutom den självklara första frågan vad som avses med ”stram”, ”böjlig” och ”rörlig”, så väcker detta spörsmål i sin tur en rad bredare frågor. Hur ska rytm i skönlitterär prosa definieras? Hur mäts den bäst? Fallet *Kallocain* får alltså aktualisera ett bredare problemkomplex som vi i denna studie angriper med datorlingvistiska metoder. Materialet utgörs av de digitala versioner av Karin Boyes verk som finns tillgängliga i Litteraturbanken.⁷ Ett grundläggande grepp har alltså varit att pröva vad som kan kallas ett informerat intuitivt yttrande om språklig rytm mot de resultat som kan nås med huvudsakligen automatiserade datorlingvistiska metoder. Vetenskapsteoretiskt utgår bägge analyserna från empiriska fakta, men insamlade och bearbetade på ytterst olika sätt. Två frågor i sammanhanget är dels om analyserna pekar i samma riktning, dels om den senare kan ge en mer nyanserad och exakt bild.

Kampen mellan de två stilarna

I vidare mening hänger frågan om rytmen i *Kallocalain* samman med mer övergripande frågor om stilen i romanen. Hur denna stil är beskaffad har belysts dels i Margit Abenius biografi om Boye, dels i en strukturalistiskt inspirerad analys av Anton Fjeldstad.

Margit Abenius talar om ”de två stilarna” i romanen, en som har att göra med förtryck och rädsla och en som är förbunden med liv och växande. De representerar två krafter som kämpar mot varandra i samhället och inom människorna. Med närmast prosalyrisk diktion summerar Abenius:

Allt det rustade, raka och pansarslutna i sin stil har författarinnan renodlat till en mekanisk omänsklighet som skrämmer oss – i rytmens effektiva marschtakt, i det småskurna ordvalet, i syntaxens slätkammade normalprosa. I det andra språket möter vi den stora lyrikern. Det är guds- och människospråket, så som det talas av försökspersonerna i rus och i dröm och Leo Kall själv efter genombrottet, än spänt och lidelsefullt, än trevande avbrutet, än sorgset, än trotsigt eller tillitsfullt – den ena skiftningen är inte den andra lik.⁸

Anton Fjeldstad följde upp denna beskrivning med en undersökning av stilen i romanen som i praktiken utformades som en analys av ordförrådet och satsbyggnaden. Den grundläggande konflikten beskrev han som motsättningen mellan det organiserade och det organiska, vilken språkligt sett tar sig uttryck i konflikten mellan stereotypa uttryck och spontana (där de förra oftast kommer från myndighetslojala personer och de senare flödar fram exempelvis i repliker från personer i *kallocalainrus*).⁹

Fjeldstad behandlar inte rytmen i romanen, men det ter sig uppenbart att hans synpunkter är förenliga med Gjerdmans beskrivning, även om Fjeldstad (liksom Abenius) inte tänker sig någon direkt stilistisk vändpunkt i romanen, utan snarare ser en pågående och inre konflikt genom hela framställningen.

Ordet ”rytm” i Karin Boyes författarskap

Som poet reflekterade Karin Boye naturligtvis över den rytm hon praktiserade. Men själva föreställningen om rytm hade också en vidare betydelse för henne än den rent verstekniska, vilket framgår av en genomgång av ordets förekomst i hennes författarskap.¹⁰

I diktsamlingarna är antalet belägg som följer: *Moln* 1, *Gömnda land* 0, *Härdarna* 3 och *För trädets skull* 1. En intressant aspekt är att ordet ”rytm” i dikterna gärna förknippas med attraktionskraften hos en älskad person, dels i ”Nattens djupa violoncell” (*För trädets skull*), dels i den diktsvit som brukar kallas ”Tillägnan” (*Härdarna*),

ägnad en tillbedd men avvisande person: ”vandrar du förbi den i din svingande rytm, / så tänker du nog inte på mig”; ”I rytmen av din rytm, i takten av din gång / slår min puls i hunger.”¹¹ Här uttrycks alltså tanken att varje människa har sin egen personliga rytm. (Citatet ur *Härdarna* får en särskild psykologisk laddning av att dikten kan sättas i samband med en intensiv och olycklig förälskelse från Karin Boyes sida under studentåren.¹²)

I novellsamlingarna förekommer ordet flitigare (*Uppgörelser* 4; *Ur funktion* 2). Mönstret är inte helt entydigt, men ordet tenderar att förknippas med dels personligheten, dels själva livskraften (”Lägg livet i kistor! Lägg rytmen i kistor!”; i en annan novell beskrivs konst som ”en linje inifrån, som en bjudande medfödd rytm att ge ut åt världen”).¹³

I romanerna förekommer ordet rytm ofta i storstadsskildringen *Astarte* (10), då ofta med koppling till dans och jazzmusik, men bara tillfälligtvis i *Merit vaknar* (1) och *För lite* (1). Förekomsterna är inte många i *Kris* (4), men desto mer centrala – tre av dem kopplas till huvudpersonen Malin Forsts starka kärleksupplevelse av sin älskade, den fjärde till hennes upplevelse av skönheten i Bachs musik. I *Kallocain*, slutligen, förekommer ordet inte alls.

Som redan har nämnts, diskuterade Boye rytmens betydelse i några essäer, främst i ”Språket bortom logiken” (1932), ”Tendens och verkan” (1935) och ”Diktning och moral” (1935). Hennes grundtanke är alltså att det finns en djupgående förbindelse mellan språklig rytm och personlig äkthet. Tankegången kan enklast summeras med ett par meningar ur den först nämnda essän, som har karaktären av ett estetiskt manifest (ordet ”rytm” används dock där bara två gånger):

Skulle författaren plötsligt infoga något som för honom inte hörde in i samma känslösammanhang – kanske gripen av plikt-känsla, eller därför att det är lättare att skriva med uttryck, bilder, rytmer, händelser, som redan förut är använda och tillrättalagda, då märker man ett avbrott, en död punkt i diktverket. Det där är inte äkta! Och det hjälper inte, att de döda orden, de döda händelserna medvetet innehåller författarens bestämda övertygelse – som dikt är de ändå inte äkta, när de inte på hans eget givna symboliska språk uttrycker hans egna känsloupplevelser.

Den underjordiska betydelsevärlden i ett diktverk, det hemliga och personliga språket innanför det logiska, är det som avgör, om diktverket skall ha förmåga att gripa eller inte.¹⁴

Som citatet visar, ansåg Boye rytmen tillhöra en ”underjordisk” och bara halvt medveten värld av estetiska verkningsmedel som uttrycker den egna personligheten. Och även om tanken inte kan utläsas med samma diskursiva tydlighet i hennes skönlitterära verk, är den som har framgått fullt förenlig med hur ordet ”rytm” används där. Det är

inte orimligt att koppla denna användning av ordet ”rytm” till hennes intresse för psykoanalysen och vad den kunde uppdaga om människors djupare psykologiska skikt.¹⁵

Karin Boyes uttalanden om rytm tyder alltså på att hon såg den som djupt förbunden med personligheten, vilket innebär att Olof Gjerdmans tes om rytmförändringen åtminstone på ett teoretiskt plan står i samklang med Boyes litteraturuppfattning.

Den tidigare forskningen och Boyes egna uttalanden har lett oss till några hypoteser som kommer att prövas i den följande undersökningen:

1. Två stiltendenser och rytmiska tendenser varvas med varandra i *Kallocalain*: de kan preliminärt kallas ”den strama” respektive ”den levande”.
2. Stiltendenserna är förbundna med gestalternas personlighet och lojaliteter, men dessa personligheter kan genomgå temporära eller permanenta förändringar – exempelvis genom ett kallocalainrus.
3. Av detta följer att rytmen i berättarpartierna bör ändras när Leo Kall har genomgått sin ”omvändelse” (just som Olof Gjerdmann har förutsatt), men att rytmen i romanen som helhet – berättarpartier och repliker – bör vara mer sammansatt. Ett skäl till detta är att de kallocalainrus som beskrivs främst ligger i romanens början.

Litterär och språklig rytm

Den undersökning av rytmen som vi vill genomföra kräver att termen operationaliseras på ett rimligt sätt. Definitionen av ”rytm” skiljer sig åtskilligt åt mellan olika ämnesskolor. Narratologer tenderar att fokusera berättelsens och berättandets rytm, lingvister snarare den språkliga rytmen på menings- och styckenivå. Även om de narratologiska definitionerna spelar en underordnad roll i det följande, kan det finnas skäl att dröja något vid dem.

I *Aspects of the Novel* (1927) ägnade E.M. Forster ett kapitel åt ”Pattern and rhythm”. Han erkänner att bägge termerna är vaga och använder dem närmast metaforiskt (*pattern* fångar han genom att likna romaner vid ett timglas eller en kedja). Rytm ser han på två nivåer i en roman, som han försöker fånga med en musikalisk jämförelse:

Rhythm is sometimes quite easy. Beethoven's Fifth Symphony, for instance, starts with the rhythm 'diddidy dum', which we can all hear and tap to. But the symphony as a whole has also a rhythm – due mainly to the relation between its movements – which some people can hear but no one can tap to. This second sort of rhythm is difficult, and whether it is substantially the same as the first sort only a musician could tell us. What a literary man wants to say, though, is that the first kind of rhythm, the diddidy dum, can be found in certain novels and may give them beauty. And the other rhythm, the difficult one – the rhythm of the Fifth Symphony as a whole – I cannot quote you any parallels for that in fiction, yet it may be present.¹⁶

Med den första sortens rytm menar Forster inte språklig rytm, utan snarare användningen av små detaljer som återkommer i berättelser och därmed förbinder olika delar med varandra – något mindre påträngande än mönster eller symboler. Rytm i den meningen definierar han som ”repetition plus variation”.¹⁷ Rytm i den andra meningen vågar han sig inte på att definiera, men avser något i stil med relationen mellan delarna av en berättelse och hur den utvecklas. ”Expansion. That is the idea the novelist must cling to. Not completion. Not rounding off but opening out.”¹⁸

Forsters ganska diffusa resonemang om rytm gränsar till poetiken snarare än till narratologin, men under de följande decennierna har den berättartekniska forskningen försökt ringa in frågan om rytm i berättande prosa. Staffan Björck skilde mellan ”satsrytm” (som ibland kan anta ”metrisk rytm”) och vad han kallar ”den episka rytmen” eller berättartempot. Särskilt intresserade han sig för hur motivet (det skildrade) slår igenom i stil och berättartempo. ”Satsrytmen” beskrev han knappast med mer än metaforiska uttryck.¹⁹

Franz K. Stanzel menar å sin sida att rytmen i en berättelse främst alstras av berättartidpartierna (han räknar alltså bort dialogen). Hans intresse gäller tydligen nog narrativ rytm, *Erzählrhythmus*. Narrativ rytm skapas enligt honom dels av variationen mellan olika slags framställningsätt (rapportering, kommentar, beskrivning, etc.), dels genom växlingen mellan olika slags ”narrativa situationer” (vilket ungefär motsvarar scener). De berättelser som byter mellan olika slags framställningsätt och olika situationer har en starkt framträdande rytm.²⁰

För Gérard Genette var berättarrytmen en fråga om relationen mellan den berättade tiden och berättartiden, alltså omfattningen av den tid som skildras och den tid själva skildringen tar. Den narrativa rytmen skapas enligt honom av hur denna relation varierar över tid genom berättelsen, liksom av hur partier med summeringar alternerar med scenisk framställning. Genette påpekar också att detaljerade analyser av rytmen i romaner ofta försvåras av att den berättade tiden sällan specificeras ordentligt.²¹

Definitioner av det slag som diskuteras ovan avser alltså främst det som kan kallas berättartempo, snarare än den språkliga rytm som finns på satsernas nivå. Ett exempel på den sortens analys är Karl Berglunds datorstödda, manuellt kvantitativa analys av hur långt avståndet är i kriminalfiktions mellan när mördaren nämns första gången och när hen avslöjas (en relation som ger genren mycket av dess narrativa energi genom vad som kan kallas det uppskjutna avslöjandet).²² I vidare mening finns naturligtvis liknande observationer om narrativ rytm i alla analyser som inriktas på spänningarna mellan de aspekter på berättelser som kallas fabula/sujet eller story/discourse.

Ett exempel på en mer språkligt inriktad litterär analys som tangerar frågor om rytm är Tanya Clements analys av en av Gertrude Steins romaner, *The Making of Americans*. Clements analys är intressant för vårt projekt eftersom hon visar hur Steins roman kan

uppfattas som uppbyggd av två delar, med en tydligt signalerad brytpunkt. Med digitala verktyg kan Clement visa hur Stein arbetar med upprepningar av språkliga mönster genom hela romanen, men på olika sätt i de två delarna. En annan språklig men också innehållsligt viktig skillnad mellan de två delarna är pronomenfrekvens – det högfrekventa *I* i första delen ersätts av det högfrekventa *one* i andra delen.²³

Det verkar emellertid uppenbart att Gjerdmans observationer om Boyes stil gäller just den språkliga rytmen. De narratologiska aspekterna på berättarrytmen lämnas därför därhän i vår vidare undersökning. Istället har vi gått till den forskning som gäller rytm i svenskspråkig skönlitterär prosa.

Stilforskaren Nils Svanberg (som var samtida med Karin Boye i Uppsala) behandlade vid mitten av 1930-talet kortfattat frågan om rytm i litteratur. Det skedde i ett avsnitt betitlat ”Musikaliska stildrag”, där Svanberg främst uppehåller sig vid växlingen mellan obetonade och betonade stavelser. Han menar också att ”satsernas rytmbilder” hänger ihop med betoning, så att korta satser eller korta rytmiskt utpräglade enheter påverkar rytm bilden. Men han nämner även berättarrytmen, ”berättelsens eller skildringens tempo” och ”de tidslängder som olika motiv och teman upptaga”. En återkommande idé hos Svanberg är att rytmen framkommer bäst vid högläsning.²⁴

Den språkliga stilforskningen kring skönlitterärt berättande kom under de följande decennierna alltmer att ägna sig åt de mindre enheterna i texten vad gäller rytm. Ett exempel är Gösta Holm, som tydligt kopplar samman rytm och syntax i sina studier av prosastilar under olika epoker. Också Holm lyfter fram att språkliga enheters längd har inverkan på rytmen.²⁵

Det hittills mest omfattande försöket att fånga rytm i skönlitterär prosa har gjorts av Marianne Nordman, som har utarbetat en modell för analys av rytm i prosa. Hon bygger sin modell dels på kvantitativa aspekter, som variation av styckelängd, meningslängd och fraslängd, och dels på kvalitativa aspekter som variation av innehållsliga enheter, växling mellan betonad och obetonad stavelse, och stilistiska verkningsmedel som anaforer och allitteration.²⁶

Vissa delar av Nordmans modell återkommer i den analys Lisa Holm gör av rytmen i tio svenska samtida romaner.²⁷ Holm bygger sin analys på följande kriterier:

1. Meningslängd.
2. Fraslängd – med fras avses här grafisk fras, ord i mening avskilda med ”liter” skiljetecken, som komma, tankstreck, kolon och så vidare, det vill säga antal ord delat med antal (små) skiljetecken.
3. Andel meningar som innehåller minst en huvudsats.
4. Samlat antal skiljetecken.
5. Antal kolon.

6. Antal vidareindelande skiljetecken – komma, semikolon, tankstreck, parenteser.
7. Antal expressiva skiljetecken – frågetecken, utropstecken, tre punkter, tankstreck (när det markerar avbrott i replik).

I studien används utdrag ur varje roman på 2 500 ord, en begränsning som troligen främst gjorts eftersom materialet huvudsakligen behandlats för hand.

Det huvudsakliga resultatet av Holms analys är att hon tycker sig urskilja ett slags ”neutral romanrytm”, en rytm som bland annat utmärks av långa fraser och meningar, få ofullständiga meningar och få ”expressiva” skiljetecken, som frågetecken och utropstecken. Holm gör också en observation som är relevant för vår undersökning: andelen *hade* samvarierar med andra språkliga drag som till exempel meningslängd – romaner med hög genomsnittlig meningslängd har också hög andel *hade*.

Holms metod låter henne urskilja tydliga skillnader mellan de valda romanerna – men det ska samtidigt sägas att hon inte gör någon skillnad mellan berättarpartier och repliker. Även om kriterierna inte restlöst kan översättas till Gjerdmans metaforiska termer ”stram” och ”böjlig”, öppnar de en möjlighet att belysa stilen i en roman som helhet – och att jämföra den i exempelvis olika berättarpartier eller i repliker av olika personer.

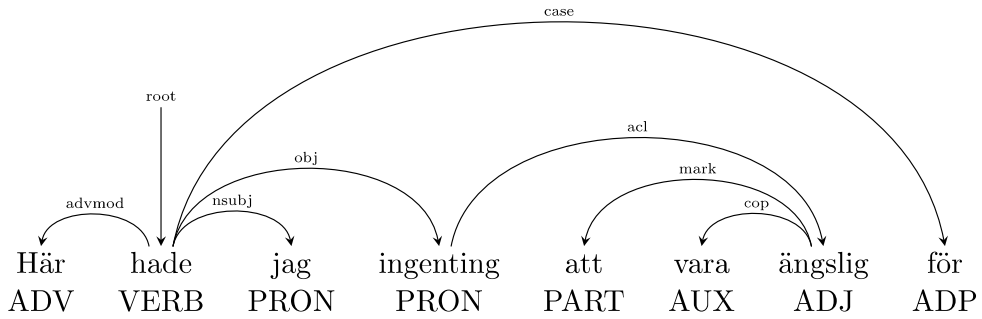
Som genomgången ovan visar finns det inte någon tydlig gemensam bild av vad rytm i litteratur innebär, men det går att urskilja vissa mönster. Flera forskare, bland annat Nils Svanberg och Gösta Holm, tar upp språkliga enheters längd som en viktig faktor för den uppfattade rytmen, och vi har därför valt att låta längd stå i centrum för vår analys, kompletterat med några andra indikatorer (se vidare avsnittet ”Indikatorer” nedan).

Terminologin gällande beskrivningen av berättande skiljer sig åt mellan narratologer och stilforskare, och dessutom mellan olika företrädare för de olika riktningarna. I det följande postulerar vi följande definitioner, som vi använder i urval och beskrivningar. En berättelse kan delas upp i nuhistoria (ibland: nuplan), förhistoria och efterhistoria (allt definierat efter den första och sista scenen). Själva framställningen består dels av berättarpartier, det vill säga berättarens egen framställning (hos till exempel Björck kallat relation), dels av olika inskottstexter – partier där inte berättaren för ordet i sin egenskap av berättare. De viktigaste av sådana inskottstexter är repliker, där olika personers tal återges (även berättarens i dennes egenskap av aktör). Sådana repliker markeras i tryckt text i regel genom olika typer av anföringstecken.²⁸

Bearbetning av texterna

Vi har fått tillgång till Litteraturbankens version av *Kallocain* i ett XML-format som följer layouten i originalupplagan (vilket alltså innebär att paginering och radnummer ska vara identiska). Vi har behandlat den råa textfilen i ett antal steg, för att kunna utföra våra analyser. I *Kallocain* finns en del som uppenbart inte hör till huvudberättelsens nuhistoria, det sista kapitlet, "Censors efterskrift" (284–285), skrivet av en fiktiv censor i en torr myndighetsprosa. Vi har exkluderat detta avsnitt från vår huvudanalys, men aktualiserar det ibland som jämförelsetext. Dessutom delar vi upp verket i två delar enligt Gjerdmans hypotes, dels delen fram till Leo Kalls omvändelse, dels delen efter denna (med gränsen vid s. 182, r. 10). I det följande kommer vi att dels tala om censors efterskrift, dels om romanens två delar (före/efter Gjerdmans gräns, men exklusive censors efterskrift).

Det första steget i analysen är att rensa XML-filen från all uppmärkning, så att enbart all text från första kapitlet och framåt kommer med. Texten delas sedan automatiskt upp i meningar, baserat på skiljetecken. Därefter tokeniseras texten, det vill säga delas upp i individuella ord och skiljetecken. Därefter applicerar vi automatiska verktyg från SwePipe,²⁹ som märker upp texten med ordklasstaggar och parsar den till en dependensrepresentation.³⁰ Ett exempel på hur denna uppmärkning kan se ut för en mening från *Kallocain* återfinns i figur 1. Där ser vi att varje ord har fått en ordklasstagg, exempelvis är "Här" adverb och "hade" verb. Ordklasstaggningen har utförts med verktyget *efselab*,³¹ som ingår i SwePipe. I samband med ordklasstaggningen tar vi även ut morfologiska särdrag, som att "hade" är aktivt och finit i indikativ och preteritum, och "vara" är aktivt och infinit. Dependensrepresentationen utgörs av ett dependens-träd, där orden förbinds med varandra och deras inbördes relationer (i huvudsak deras inbördes förhållanden som satsdelar) märks ut. Den representationen är markerad i figur 1 med hjälp av bågar mellan orden. Här kan vi exempelvis se att "jag" är subjekt till huvudordet "hade", och "ingenting" är objekt. Dependensanalysen har gjorts med verktyget *MaltParser* och baseras på uppmärkningsstandarden i projektet *Universal Dependencies*.³²



Figur 1: Exempel på en dependensanalys i form av ett dependensträd med ordklasstaggar.

Källa: mening ur *Kallocain*, s. 6.

En ambition i arbetet har varit att bedriva det med en hög grad av automatisering. Det hade varit fullt möjligt att gå in i texten med mer omfattande uppmärkning av olika typer av inslag i texten (t.ex. vem som talar), men vi har alltså valt att undersöka vilka resultat som kan nås utan sådana manuella insatser. Vi anser att detta arbets sätt är viktigt ur ett metodologiskt perspektiv, då det i framtida projekt kan låta oss analysera större mängder text än vad som är möjligt enbart med manuell analys. Även om det kan förekomma vissa fel i den automatiska analysen, så anser vi att den som helhet ger en bra överblick, eftersom den är baserad på relativt mycket text.³³ (Ett undantag från denna huvudmetod har dock gjorts i en delstudie som gäller rytmen i repliker i romanen.)

Repliker och berättarpartier

I vår undersökning är vi främst intresserade av hur rytmen förändras i Leo Kalls berättelse. Eftersom romanen är skriven i första person kan vi anta att berättelsen är skriven ur hans perspektiv, och återspeglar honom som person. Det finns dock även en stor mängd repliker i romanen, yttrade av många olika personer. Dessa personer kan inte antas förändra sin språkliga rytm på samma sätt som Leo Kall (vars egna repliker i och för sig kan tänkas ändra rytm under berättelsens gång). Att identifiera repliker och talare automatiskt är dessutom ett svårt problem, som inte kan lösas med hög tillförlitlighet. Vi har därför valt att i de flesta fall exkludera alla repliker ur vår undersökning, inklusive Leo Kalls repliker, och därmed enbart fokusera på berättarpartierna. Detta gör att vi kan vara säkra på att i stort sett all text vi analyserar är skriven ur Kalls perspektiv.³⁴

Repliker är i *Kallocain* tydligt definierade i ett nytt stycke som börjar med talstreck. Bortsett från eventuell anförings sats håller de i de allra flesta fall på till styckets slut. Det finns dock fall där en replik slutar mitt i ett stycke, och utan särskild markering

övergår i berättarparti. Dessa fall är dock relativt få, och då de är svåridentifierade automatiskt, har vi valt att alltid exkludera alla dialogstycken. I några fall har det medfört att vi har filtrerat bort en liten del av det totala berättarpartiet. Vi tror dock att detta har en perifer betydelse för studiens resultat.

Ett exempel på den sorts replik som ommarkerat övergår i berättarparti är:

– Hur känns det? frågade jag uppmuntrande, medan innehållet i sprutan minskades. Fortfarande från ordningsföreskrifterna visste jag ju, att man gärna borde fråga försökspersonen själv så mycket som möjligt, det gav honom en känsla av likställdhet och höjde honom på visst sätt över smärtorna. (53)

Indikatorer

För att operationalisera frågan om språklig rytm i skönlitterärt berättande har vi stannat för att huvudsakligen utgå från de kriterier hos Lisa Holm som har presenterats ovan. Vi har dock lagt till ett antal övriga indikatorer, som vi anser kan fördjupa beskrivningen av rytmiska element i texten. De sju första indikatorerna i listan nedan är baserade på Holm.

Vi har något modifierat ett par av Holms indikatorer. Hon har sorterat alla meningar i två grupper, de som innehåller minst en fullständig huvudsats och de som inte gör det. I vår automatiska analys har vi inte kunnat skilja mellan huvudsats och bisats, utan vårt mått *ejS* anger andel meningar som inte innehåller någon sats, det vill säga varken huvudsats eller bisats. Vi har också modifierat Holms indikatorer för skiljetecken, och istället för absolut antal (vilket Holm kunde använda eftersom alla textpartier hon analyserade hade samma längd), anger vi det istället som andel skiljetecken per 100 ord. Vi har också kompletterat undersökningen av skiljetecken genom att lägga till separata indikatorer för tankstreck och för tre punkter eller tre tankstreck (här kallat ellips) – användningen av dessa skiljetecken är ett påfallande inslag i *Kallocalain*. Holm använder inte ordlängd som mått, och här har vi också kompletterat hennes metod genom att lägga till ordlängd. I vår studie försöker vi ganska förutsättningslöst att pröva vilken relevans olika språkliga mått har, och därmed är också ordlängd av vikt. Våra nya indikatorer på ordnivå mäter dels genomsnittlig ordlängd baserat på såväl tecken som stavelser, dels andelen ord som är långa. Kategoriseringen av ”långa ord” utgår från läsbarhetsindex (LIX), som definierar långa ord som ord med minst 6 tecken.³⁵ Vi har även urskilt en grupp med extra långa ord med fler än 14 tecken, ett tillägg till LIX som gjorts av Mühlenbock och Kokkinakis.³⁶ Vi anger också andel ord med ett mycket högt antal stavelser, 5 eller fler. Ytterligare en skillnad mellan vår studie och Holms är att vi använder helt automatiska metoder. I vissa fall kan det leda till felanalyserade me-

ningar, men den metodiska vinsten är dock att automatiseringen tillåter oss att analysera längre texter, och alltså slippa begränsa oss till kortare utdrag ur texterna.

Baserat på Marianne Nordmans redan anförda arbete har vi även lagt till styckelängd som indikator, och vi följer henne genom att mäta den i antal ord per stycke. Slutligen ger vi även type/token ratio (TTR), som är ett mått på ordvariation. I Fjeldstads stilundersökning av Kalloccain lyfter han fram en motsättning mellan stereotypa uttryck och mer spontana, vilket är en motsättning som skulle kunna fångas av mått på ordvariation.³⁷ Detta mått är lägre ju oftare ord upprepas, och kan därmed anses vara ett indirekt mått på upprepningar.³⁸ Eftersom detta mått varierar med längden på texten det beräknas på, så har vi angett det som ett genomsnitt av TTR för textavsnitt med 2 000 ord.³⁹

Våra indikatorer summeras i tabell 1. Varje punkt inleds med en förkortning för varje indikator, som kommer att användas i tabeller och figurer.

Tabell 1. Indikatorer på rytm i svenskspråkig skönlitterär prosa

1. *MeningsL*: genomsnittlig meningslängd.
2. *FrasL*: genomsnittlig fraslängd (definierat grafiskt som antal ord mellan frasavgränsande skiljetecken).
3. *ejS*: Andel icke satsformade meningar (identifierat utifrån ordklassstaggar och dependensträd).
4. *Skilje*: Andel skiljetecken per 100 ord.
5. *Kolon*: Andel kolon per 100 ord.
6. *Vidare*: Andel vidareindelade skiljetecken per 100 ord (komma, semikolon, tankstreck (ej använt som replikmarkör) och parenteser).
7. *Expr*: Andel expressiva skiljetecken per 100 ord (frågetecken, utropstecken, tre punkter (ellips) samt tre tankstreck, vilket också används i en elliptisk funktion).
8. *Streck*: Andel tankstreck per 100 ord.
9. *Ellips*: Andel ellips (tre punkter eller tankstreck) per 100 ord.
10. *ordL/T*: Genomsnittlig ordlängd (i antal tecken).
11. *ordL/S*: Genomsnittlig ordlängd (i antal stavelser).
12. *Långa*: Andel långa ord (>6 tecken).
13. *Xlånga*: Andel extra långa ord (>= 14 tecken).
14. *StavS*: Andel ord med många stavelser (>= 5 stavelser).
15. *StyckeL*: Genomsnittligt antal ord per stycke.
16. *TTR*: Type/token ratio, det vill säga antal unika ord (types) i relation till antal löpord (tokens), givet som ett genomsnitt per 2 000 ord.

De flesta av våra indikatorer är enkla och kan relativt enkelt tas fram direkt ur texten.⁴⁰ Den indikator som kräver en djupare analys är *ejS*, där vi utgått från definitionen att en sats antingen ska ha ett subjekt och ett finit verb, eller ett verb i imperativ. Om det finns ett subjekt ser vi i dependensträdet, och verb, samt dess morfologi, får vi ut från ordklassstagningen.⁴¹

Analysmetod

Grundmetoden är alltså att tillämpa våra indikatorer på berättarpartierna i de två olika delarna av *Kallocain* och jämföra utfallet. Detta görs dels för hela texten, dels för avsnitt av texten på cirka 1 000 ord, för att kunna se trender i hur utfallet för indikatorerna varierar genom romantexten. Att avsnitten inte är exakt 1 000 ord vardera beror på att vi inte vill bryta mitt i en mening, eftersom flera av våra indikatorer är meningsbaserade. Vi inkluderar därför resterande ord i den mening som håller på vid 1 000 ord. Detta har resulterat i att avsnitten varierar i längd från 1 000 till 1 049 ord. Samma metod används i en mindre delundersökning som gäller replikerna i romanen.

Vi utför signifikanstestning av skillnaderna mellan delarna med två metoder. För de indikatorer som är medelvärden för längd (till exempel meningslängd och ordlängd) samt för TTR utför vi ett Mann-Whitney-test. För de indikatorer som är proportioner använder vi istället Chi-två-test (χ^2 -test). För delstudien där vi jämför kortare textavsnitt använder vi Mann-Whitney-test per textavsnitt. För samtliga signifikanstest använder vi signifikansnivån 5% ($\alpha < 0.05$).⁴²

För att ytterligare undersöka skillnaderna mellan de två delarna i *Kallocain* använder vi maskininlärning. I denna del använder vi oss av ännu kortare textavsnitt, om cirka 400 ord. Denna metod presenteras senare i uppsatsen. Här får det räcka med att säga att maskininlärningen dels låter oss undersöka om det är så stora skillnader mellan de två delarna att det automatiskt går att säga vilken del av boken ett textparti har hämtats ifrån, dels låter oss analysera vilka indikatorer som är viktigast för att kunna identifiera skillnaden.

Resultat – romanens delar

Tabell 2 nedan visar värden för våra indikatorer sett över de bägge delarna av *Kallocain*. Vi är främst intresserade av de två första delarna, före respektive efter Leo Kalls omvändelse. Först dock några ord om censors efterskrift.

Censors efterskrift är med enbart som referens, och saknar ett flertal indikatorer eftersom det är ett mycket kort textparti. Men vi kan lätt se att efterskriften är betydligt mer formellt skriven, med lång menings- och fraslängd, och därmed även färre skilje-

tecken (förutom tankstreck), längre ordlängd, och även en högre andel långa ord. Till det formella intrycket kan läggas avsaknaden av hjälpverbet *har* eller *hade* i bisats med tempus perfekt eller pluskvamperfekt:

Att den icke helt och hållet [har] förstörts beror på att just detta omoraliska innehåll av mera pålitliga forskare torde kunna användas som material, då det gäller att klarlägga mentaliteten hos de varelser, som bebo landet intill vårt. (284)

Ingen TTR (type/token ratio) anges för censorns efterskrift eftersom den beräknas per 2 000 ord, och efterskriften är kortare än så. Värdet skulle därför inte vara jämförbart.

Tabell 2: Resultat för rytmindikatorerna för de två delarna av *Kallocalain*, samt för censorns efterskrift.

Indikator	Första delen	Andra delen	Skillnad (%)	Censorns efterskrift
MeningsL	20,33	17,46	-14,1	51,00
FrasL*	8,24	7,75	-5,9	14,57
ejS	0,06	0,06	0	0
Skilje	12,30	13,10	7,3	6,86
Vidare	6,79	6,78	-0,01	5,50
Expr	0,32	0,39	21,9	0
Kolon	0,36	0,28	-22,2	0
Streck	0,48	0,59	22,9	0,92
Ellips	0,05	0,05	0	0
OrdL/T*	4,80	4,57	-5,0	5,45
OrdL/S*	1,72	1,64	-4,7	1,95
Långa*	0,20	0,17	-17,6	0,30
Xlånga*	0,014	0,007	-50,0	0,020
Stav5*	2,04	1,02	-50,0	4,90
StyckeL	87,68	73,99	-18,5	204,00
TTR*	0,43	0,39	-9,3	-

Indikatorer markerade med asterisk (*) har en statistiskt signifikant skillnad på 0.05-nivå mellan värdet i första och andra delen.

När det gäller jämförelsen mellan de två delarna av *Kallocalain* så finns det en del tydliga skillnader. Alla våra indikatorer för ordlängd uppvisar en signifikant skillnad mellan de två delarna, där den första delen har något längre ord, mätt både i stavelser och i tecken, men framförallt dubbelt så många mycket långa ord, enligt båda våra mått för

detta. Andelen mycket långa ord är dock låg i bägge delarna. Type/token ratio uppvisar också en signifikant skillnad, vilket tyder på något fler upprepningar i andra än första delen. Även fraslängden är signifikant längre i den första delen. Meningslängd samt flera av våra indikatorer för skiljetecken uppvisar en skillnad mellan delarna, som dock inte är signifikant. De flesta kategorier av skiljetecken förekommer dock relativt sällan, vilket gör det svårt att hitta signifikanta skillnader i ett sådant här relativt litet material. Notera dock att undersökningen trots det är betydligt större än tidigare relaterat arbete, där inga signifikanstester har gjorts. För några av de enskilda skiljetecknen är värdena dock relativt lika, liksom för andel icke satsformade meningar.

Utfallet av undersökningen kan alltså summeras som att tre indikatorer (nr 3, 6, 9) inte uppvisar någon skillnad mellan de bägge delarna, medan övriga tretton indikatorer uppvisar skillnader på mellan 4,7 och 50 procent, och dessa skillnader är signifikanta för 7 indikatorer (2, 10, 11, 12, 13, 14, 16). Ingen av våra indikatorer visar ett resultat som motsäger Gjerdmans hypotes. Notera dock att flera kategorier samvarierar, som fraslängd respektive vidareindelade skiljetecken, ordlängd i tecken respektive stavelser, och extra långa ord respektive ord med många stavelser. Andel icke satsformade meningar är helt lika i de två delarna, liksom förekomsten av ellips och vidareindelade skiljetecken. Vår tolkning av resultaten är att utfallet faktiskt tyder på att det finns en skillnad mellan de två delarna. Långa ord och meningar brukar kopplas till en mer formell stil, och här tyder resultatet på en mer formell stil i första delen. Men går det att beskriva skillnaden som att den andra delen har den ”böjligare, mera rörliga rytmen”? Vi kan fortfarande inte veta vad Gjerdman egentligen avsåg men vi tänker oss att ett textavsnitt med kortare språkliga enheter, alltså ord, fraser, meningar och stycken, kan uppfattas som ”mer rörligt”, med större möjligheter till varierad frasering, än ett (”stramare”) textavsnitt med längre språkliga enheter.

Framför allt meningar men även fraser är alltså längre i första delen. I gengäld innebär det att det finns något fler skiljetecken i den andra delen. För några av de enskilda skiljetecknen är värdena dock relativt lika. Vissa skillnader föreligger emellertid för kolon (vanligare i första delen) respektive för expressiva skiljetecken samt tankstreck (vanligare i andra delen). Tankstrecken stämmer överens med en mer rörlig rytm i andra delen – mer avbrott – medan de expressiva skiljetecknen kanske snarare är ett uttryck för de starkare känslor som blir synliga i romanens andra del.

I likhet med Lisa Holm har vi noterat intressanta skillnader när det gäller förekomsten av hjälpverbet *hade* i förhållande till de två delarna. Det visar sig att det även här finns en betydande skillnad med 0,87 *hade* per 100 ord i andra delen, jämfört med 0,55 i första delen, vilket stämmer väl överens med bilden av andra delen som ledigare och mindre formell. I den mer formella första delen finns många exempel på bisatser med struket *hade*: ”Bara några få minuter efter det jag [hade] släppt de båda ungdom-

marna med blicken och de för resten [hade] blivit skilda åt av otåliga kamrater, fastnade min uppmärksamhet i stället vid en mager medelålders kvinna” (s. 30), medan det i andra delen finns många exempel på motsatsen: ”Och om hon inte hade kommit, om jag inte träffade henne?” (s. 275).

I stort menar vi att analysen stöder Gjerdmans hypotes, även om mätmetoden är mycket grov. Framförallt är det intressant att ingen av indikatorerna motsäger Gjerdmans hypotes.

Resultat – textavsnitt

Figurerna 2–5 visar värden för textavsnitt om cirka 1000 ord för de två delarna (för en lista, se Bilaga).⁴³ Avsnitten är numrerade med löpnummer, skarven mellan de två delarna (det vill säga den scen där Kalls ”omvändelse” inträffar) är markerad med ett mellanrum i figurerna. I detta fall har TTR beräknats utifrån de cirka 1000 ord som finns i varje avsnitt. Graferna är uppdelade på fyra stycken, för att lättare kunna tolkas. När det gäller signifikans räknat per textavsnitt är det samma indikatorer som tidigare som uppvisar signifikanta skillnader, förutom andel skiljetecken, som här, beräknat på textavsnitt uppvisar en signifikant skillnad. Här ser vi tydligt att de flesta indikatorerna har mycket varierade värden genom romanen. Några indikatorer är dock mer stabila genom romanen än andra: främst ordlängd, fraslängd och andel vidareindelade skiljetecken.

Indikatorer som meningslängd, de flesta skiljetecken, andel icke satsformade meningar och andel långa ord varierar i mycket högre grad genom romanen. Slående är exempelvis hur mycket meningslängden varierar i olika partier. Fraslängden är mycket mer konstant, och ordlängden i sin tur ännu mer konstant.

Ett huvudresultat av denna mer finfördelade analys är att det inte är helt uppenbart att det utifrån dessa grafer skulle ha varit naturligt att dela upp romanen i två delar just enligt Gjerdmans förslag. Vissa värden förändras dock något där, kanske mest tydligt andelen extremt långa ord, men även ordlängd och fraslängd till viss del. För andra indikatorer som meningslängd, där variationen är mycket stor, ligger ändå medelvärdet signifikant lägre i andra delen (som även har framgått ovan av tabell 2).

Uppdelningen i textavsnitt gör det möjligt att urskilja avsnitt som utmärker sig genom en särskild språklig profil. Avsnitten 5, 12 och 20 skiljer ut sig genom långa meningar och få expressiva skiljetecken (och har dessutom få *hade*); avsnitten 12 och 20 avviker dessutom genom att rymma en hög andel extra långa ord (*xlånga*). De tre textställena återfinns alla i första delen, och de utmärks innehållsligt genom att Leo Kall återger händelseförlopp utan minsta tvekan – han är fortfarande alldeles säker och har inte börjat tvivla. Ett typiskt exempel från avsnitt 5:

Om något gruff uppstod, om något hemligt förehades, till exempel om någon av deltagarna försökte avlägsna sig sedan uppropet skett, var det ju till stor hjälp för ordförande och dörrvakter, som ju ofta kunde vara upptagna av någon praktisk detalj, att fyra polis-sekreterare hela tiden vakade över salen från en något så när avskild plats. (27–28)

Intressant att notera är att avsnitt 7 *inte* skiljer ut sig på samma sätt som 5, 12 och 20. Avsnitt 7 rymmer nämligen det formella brev från Propagandaministeriet där Leo Kall uppmanas att göra offentlig avbön för ett yttrande som har ansetts förgrifligt. Detta myndighetsbrev räcker dock tydligen inte för att ge avsnittet en påfallande stram rytm.

Dessa tre textställen har motpoler, främst i avsnitt 28 och 31, båda i romanens andra del, som utmärks av korta meningar och korta ord (och dessutom en hög andel av hjälpverbet *hade*). Avsnitt 31, som utspelas precis efter att Linda berättat sin syn på världen (som vi kallar FREESOBERSPEECH nedan) och där Leo försöker dra tillbaka sin angivelse av Rissen, har även den lägsta andelen unika ord av alla avsnitt. Avsnitt 28 är innehållsligt centralt: Leo Kall får idén att prova drogen kallocaïn på sin egen hustru, och utför detta efter att ha bundit henne. Händelseförloppet återberättas konkret och i enkla ord:

Åtta minuter beräknade jag som den säkra tiden för vätskan att verka. När de var gångna, löste jag näsduken. På hela hennes min såg jag, att sprutan hade tagit. (227)

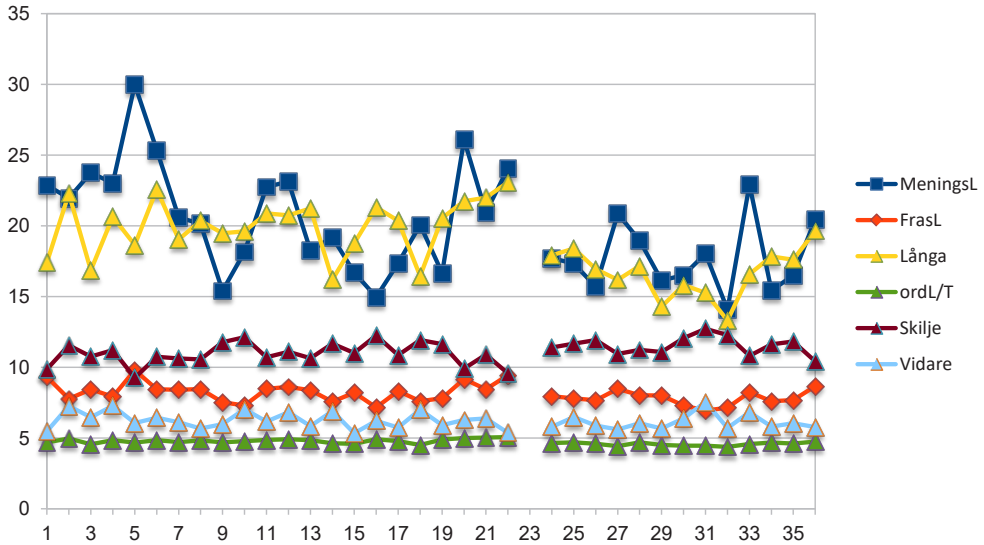
Av citatet framgår att det inte heller här finns något överflöd av expressiva skiljetecken. Just de expressiva skiljetecknen varierar stort genom romanen, och de tre textställen som ligger högst är ganska jämnt fördelade över romanen: avsnitt 8, 18 och 35. Det är inte helt enkelt att se något gemensamt hos avsnitten, förutom att skiljetecknen markerar Leo Kalls reflektioner:

Jag fick den föreställningen, att han satt där, alldeles utan kamp eller skam, och önskade att planet helt och hållet skulle varit av glas... En mycket träffande föreställning, tänkte jag, sådan var den mannen. Om jag kunde använda kallocaïn för mitt privatnöje – – – (143 f.)

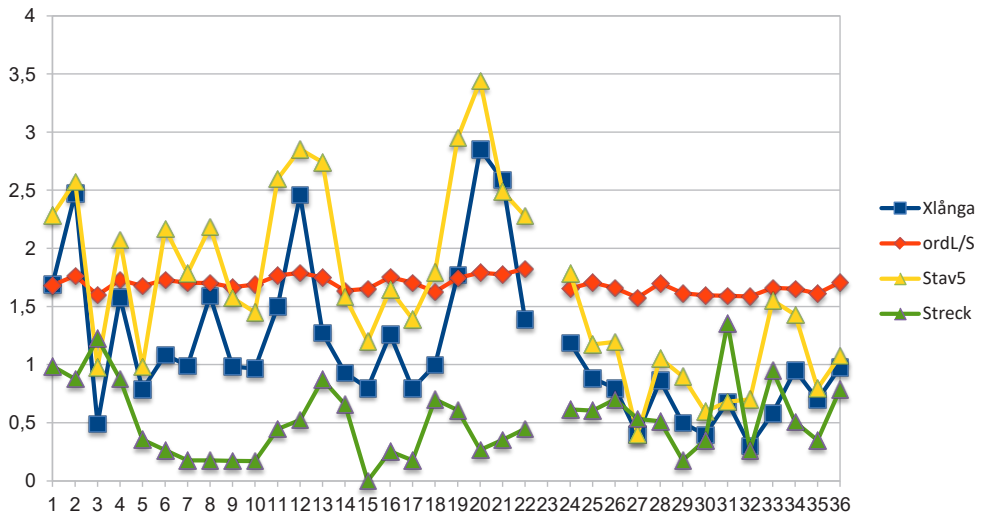
I romanen har egentligen det vi kallar expressiva skiljetecken inte någon tydligt expressiv funktion – snarare än känslor speglar de hur Leo Kall funderar och resonerar med sig själv, i första delen inom ramen för sin lojalitet mot staten och i andra delen som ett led i sin omvändelse. Andelen expressiva skiljetecken är allra störst i avsnitt 35, som är avsnittet där Leo Kall försöker fly och blir gripen av fallskärmstrupper. Inte ens i detta avsnitt signalerar skiljetecknen i första hand starka känslor utan också här förmedlas Leo Kalls intensiva diskussioner med sig själv: ”Ja – vad tänkte jag själv?” (277).

Utfallet av analysen av textavsnitten kan summeras som att den språkliga rytmen

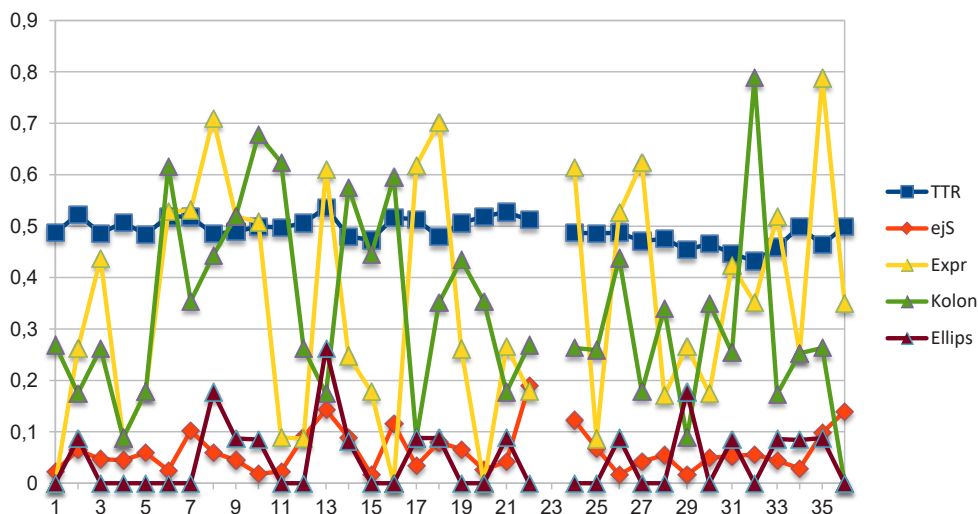
böljar romanen igenom, snarare än att den visar en distinkt och varaktig förändring på ett visst ställe. Några indikatorer visar ganska små skillnader mellan avsnitten, andra större. Men allt ska samtidigt ses mot bakgrund av det övergripande mönstret att tydliga skillnader faktiskt föreligger mellan de två delarna, sedda som helheter.



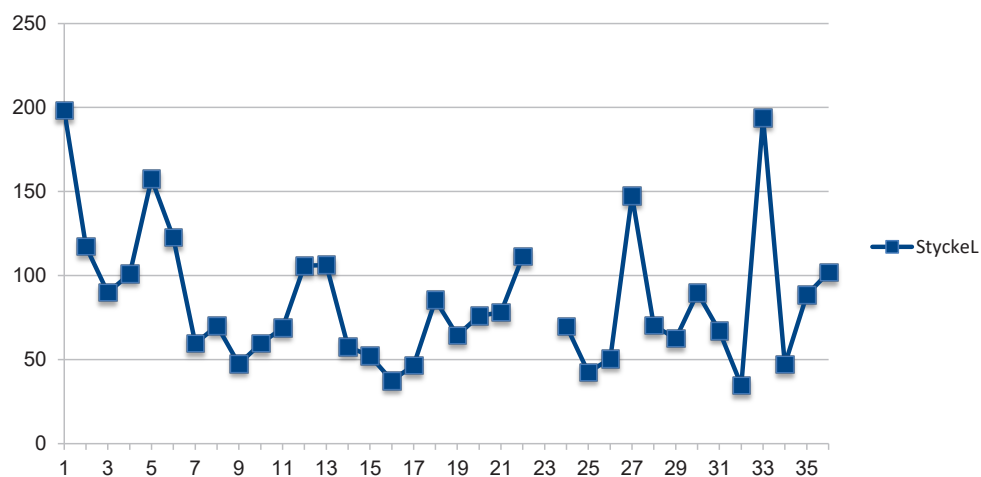
Figur 2: Trend för en delmängd av rytmindikatorerna i Kallocairn, för avsnitt med cirka 1 000 ord. Avsnitt 1–22 hör till första delen, avsnitt 23–35 hör till andra delen.



Figur 3: Trend för en delmängd av rytmindikatorerna i Kallocairn, för avsnitt med cirka 1 000 ord. Avsnitt 1–22 hör till första delen, avsnitt 23–35 hör till andra delen.



Figur 4: Trend för en delmängd av rytmindikatorerna i Kallocaïn, för avsnitt med cirka 1 000 ord. Avsnitt 1–22 hör till första delen, avsnitt 23–35 hör till andra delen.



Figur 5: Trend för en delmängd av rytmindikatorerna i Kallocaïn, för avsnitt med cirka 1 000 ord. Avsnitt 1–22 hör till första delen, avsnitt 23–35 hör till andra delen.

Rytm i repliker

Romanen igenom pågår en diskussion mellan företrädare för det som tidigare forskning har beskrivit som de två polerna, den förtryckande eller ”organiserade” gentemot den levande eller ”organiska”. Det finns gott om tillfällen när lojala representanter för staten (olika makthavare och övervakare) explicit ger andra personer tillsägelser för

bristande engagemang. Och det finns också flera ställen där personer uttrycker en uttalad protest eller en mer sökande tro på en ny slags frihet. I mellanläget finns också repliker som präglas antingen av ambivalens eller försök att dölja opposition bakom ytlig solidaritet med staten. Därutöver finns naturligtvis åtskillig vardagskonversation, utan egentlig ideologisk laddning.

I anslutning till huvudämnet för denna uppsats har vi frågat oss om det finns olika slags språklig rytm i de olika typerna av repliker. Eftersom gränserna dem emellan ibland är suddiga har vi stannat för att använda ett mer formellt kriterium och välja ut de scener där personer genomgår behandling med kallocaïn och i samband med det samtalar med förhørsledarna. Dessa scener har en relativt tydlig dramaturgi: den behandlade personen gör entré som en person som betar sig enligt samhällets normer, får sin injektion, börjar tala friare och i halvt monologartade utläggningar, för att sedan ganska snabbt komma ur kallocaïnuset. I det fåtal fall när de har fortsatta repliker börjar de visa ånger och förnyad lojalitet med samhället. Totalt finns åtta sådana förhör med repliker i romaner, från det första med medlem nr 135 i Frivilliga offerttjänsten, via förhören med personer kring den så kallade ökensekten och till de stora förhören med Linda Kall och Edo Rissen.

Vi har helt enkelt jämfört de förhörda repliker under kallocaïnuset [FREE-SPEECH] med samtliga övriga repliker i romanen [SPEECH].⁴⁴ Ett undantag har gjorts för Lindas repliker dagen efter kallocaïnuset, en lång berättelse om hennes syn på världen, som har förts till en egen grupp [FREESOBERSPEECH]. Rent formellt yttras replikerna inte under kallocaïnuset, men i romanens sammanhang är det uppenbart att de tillhör samma grupp. Som Linda uttrycker det vill hon fördjupa den förklaring av sin lära om inre frihet som hon började under ruset, snarare än att ta tillbaka något ("Efteråt har jag förstått, att antingen måste jag dö av skam, eller också måste jag fortsätta frivilligt." [241]). Vad som följer är både romanens längsta monolog och dess tematiska kärna; till denna monolog har vi också fört de repliker som Linda yttrar till Leo under den efterföljande natten.⁴⁵

Det är värt att notera att kallocaïnförhören med Linda Kall och Edo Rissen finns i romanens andra del, liksom Lindas långa monolog, de övriga förhören i romanens första del. Vi använder samma sexton indikatorer som tidigare. Kategorin FREESOBERSPEECH rymmer dock för få ord för att TTR ska kunna beräknas på ett jämförbart sätt. I detta fall har vi inte utfört några signifikanstest, eftersom de olika repliktyperna utgör en relativt liten datamängd.

Resultaten för denna delstudie återfinns i Tabell 3. Det visar sig att det finns en hel del skillnader mellan de tre olika grupperna. Inte särskilt förvånande är styckena (det vill säga replikerna i det här fallet) längre i FREESPEECH, och i ännu högre grad i FREESOBERSPEECH. För flera indikatorer ligger dock värdet för SPEECH mittemel-

lan det för FREESPEECH och FREESOBERSPEECH, som menings- och fraslängd, ingen huvudsats, skiljetecken och ellips. För de enskilda skiljetecknen, eller grupperna av dem, är dock oftast de två typerna av fritt tal mer lika, med högre andel tankstreck och ellips, och lägre andel av de flesta övriga. För indikatorerna som mäter ordlängd ser vi att orden är både längre i genomsnitt och att det finns en högre andel långa och extra långa ord i SPEECH. SPEECH har också en högre TTR än FREESPEECH. Totalt sett ger detta inga entydiga svar kring rytmen i de olika typerna av repliker, där endast ordlängd ger ett tydligt förväntat mönster, med längre ord i SPEECH än i de två kategorierna för fritt tal.

Tabell 3: Resultat för rytmindikatorerna i olika kategorier av repliker i Kallocain.

Indikator	SPEECH	FREESPEECH	FREESOBERSPEECH
MeningsL	13,70	10,78	15,66
FrasL	5,56	5,38	6,09
ejS	0,11	0,14	0,06
Skilje	18,34	18,66	16,74
Vidare	9,64	9,00	9,59
Expr	1,95	1,56	1,10
Kolon	0,25	0,23	0,18
Streck	0,72	1,06	1,15
Ellips	0,28	0,49	0,09
OrdL/T	4,53	4,03	4,14
OrdL/S	1,64	1,46	1,49
Långa	0,17	0,10	0,10
Xlånga	0,012	0,002	0,003
Stav5	1,70	0,51	0,59
StyckeL	41,27	80,73	143,38
TTR	0,39	0,28	–

Klassificering med hjälp av maskininlärning

I denna delstudie använder vi oss av maskininlärning, för att träna ett system som ska lära sig att identifiera om ett textavsnitt kommer från den första eller andra delen av *Kallocain*. Detta kan ses som ett klassiskt textklassificeringsproblem, där vi beskriver varje textavsnitt baserat på våra rytmindikatorer, och sedan klassificerar dem som till-

hörande antingen första eller andra delen av boken. Detta liknar andra textklassificeringsstudier, där man exempelvis försökt identifiera vem som är författaren till en text eller identifiera vilket kön en författare har.⁴⁶ I detta fall är avsikten att finna automatiska metoder för att bekräfta den skillnad mellan de två delarna vi har urskilt ovan.

Kategorierna som är målet med textklassificeringen kallas klasser. I vårt fall är frågan vilken del av *Kallocain* som ett textavsnitt kommer ifrån, och klasserna representerar därmed de två delarna av romanen. Uppgiften att klassificera texten löses i de allra flesta fall med hjälp av övervakad maskininlärning, vilket innebär att en maskininlärningsalgoritm med hjälp av en relativt stor mängd träningsdata lär sig vad som är typiskt för de olika klasserna.⁴⁷ Sedan ges algoritmen ny data i form av nya textavsnitt som inte ingår i träningsdata, kallad testdata, där algoritmen får försöka klassificera vilken klass varje textavsnitt tillhör. Med hjälp av resultaten kan forskaren utvärdera hur bra algoritmen fungerar. Huvudsyftet med denna delstudie är alltså att se om det finns så stora skillnader i den språkliga rytmen i de olika delarna av *Kallocain* att det är möjligt för en maskin att lära sig skillnaderna mellan dem enbart med hjälp av våra rytmindikatorer. Om vår klassificerare kan skilja textavsnitt från de två olika delarna åt i relativt hög grad, kan det ses som ytterligare stöd för Gjerdmans hypotes. Huvudsyftet med den här delundersökningen är alltså att i ännu högre grad bekräfta de tendenser vi redan sett.

Vi har valt att använda en linjär SVM (support vector machine) som vår maskininlärningsalgoritm. SVM är en standardalgoritm för maskininlärning, inklusive klassificering. Algoritmen bygger på att försöka separera de två klasserna, i vårt fall första respektive andra delen av *Kallocain*, med så stor marginal som möjligt.⁴⁸ Vi har två huvudanledningar till det: en sådan algoritm presterar i allmänhet bra för en mängd olika textklassificeringsproblem, och det går dessutom att tolka vad den gör, så att vi kan se vilka indikatorer som algoritmen finner vara typiska för de två olika delarna. I alla våra försök använder vi algoritmen som är implementerad i biblioteket LibLinear.⁴⁹

Som tidigare nämnts behöver SVM-algoritmen två typer av data, dels träningsdata, för att lära sig vilka särdrag som är viktiga för klassificeringen, dels testdata, för att kunna se hur väl algoritmen presterar. För att ge bra resultat behövs i allmänhet mycket träningsdata. I vårt fall har vi dock en ganska begränsad mängd data, eftersom romanen har en begränsad längd och en stor del text dessutom uteslutits i sin egenskap av repliker. Därför har vi fått göra en avvägning mellan att använda textavsnitt som är tillräckligt långa för att innehålla relevant information, och att sålla fram textavsnitt som är tillräckligt korta för att vi ska ha ett tillräckligt stort antal textavsnitt i vår tränings- respektive testdata. I preliminära experiment provade vi olika längd för textavsnitten, och kom fram till 400 ord per avsnitt som en bra kompromiss. Återigen bryter vi vid meningsavslut, så textavsnitten har en något varierande längd. Detta ger 54 textavsnitt från första delen och 32 textavsnitt från andra delen. För varje försök delar vi upp detta

material i 10 procent för testdata (det vill säga 9 avsnitt), och resterande 90 procent som träningsdata, vilket är ett standardförfarande.

Att göra detta bara en gång skulle kunna ge missvisande resultat, eftersom det beror på slumpen exakt vilka avsnitt som hamnar i tränings- respektive testdata, och därmed hur våra resultat ser ut. Vi utför därför vårt experiment 100 gånger, med olika slumpmässig fördelning av textavsnitten i testdata och träningsdata. I våra resultat redovisar vi dels det genomsnittliga resultatet, dels resultatet vid olika percentiler. Vi kommer även att analysera hur stor vikt de olika indikatorerna ges.

Eftersom våra textavsnitt är kortare än i föregående del, så kommer vissa indikatorer aldrig att förekomma, till exempel finns det inte kolon i alla avsnitt på 400 ord. Därför utesluter vi indikatorerna kolon, ellips och tankstreck (nr 5, 9 och 10) från denna analys. Vi beräknar TTR per 400 ord i denna del.

Eftersom den första delen är längre än de andra, så skulle ett program som enbart gissade att alla delar kommer från den första delen ge en korrekthet om 62,7 procent av fallen, då det finns 54 avsnitt i första delen, och 32 i den andra. För att vi ska kunna säga att datorn har lärt sig någonting behöver den alltså ge rätt svar i (betydligt) fler än 62,7 procent av fallen.

Resultat – maskininlärning

Tabell 4 sammanfattar resultaten av denna delstudie i form av en översikt av korrektheten för våra 100 experiment. P står för percentil, där den 10:e percentilen, P₁₀, är värdet där 10 procent av experimenten har ett sämre resultat, och 90 procent har ett bättre resultat. Jämfört med vårt riktmärke på 62,7 procent ligger medelvärdet betydligt högre: 72,1 procent. Den 25:e percentilen ligger under detta värde, medan medianen ligger över detta värde. Det innebär att i mellan tre fjärdedelar och hälften av försöken har datorn lyckats lära sig i alla fall något om vilken del ett textavsnitt hör till. Den 75:e och framförallt den 90:e percentilen ligger högt över 62,7 procent och visar att det definitivt är möjligt för datorn att lära sig något om de två delarna, med hjälp av våra indikatorer.

Variationen mellan experimenten är dock relativt hög, vilket kan ses på percentilerna, och även indikeras av standardavvikelsen på 14,6. Totalt sett kan alltså sägas att datorns analys i en majoritet av fallen överträffar ett slumpmässigt utfall när det gäller att avgöra till vilken del ett avsnitt hör (fortfarande enbart baserat på vår reducerade grupp rytmindikatorer). Variationen mellan de olika försöken är dock hög. Detta stöder våra slutsatser från tidigare, att det finns en relativt stor variation även inom de två olika delarna, vilket gör att resultaten varierar beroende på exakt vilka avsnitt vi har i vår tränings- respektive testdata.

Vi anser dock att denna analys ändå stöder tanken att det finns vissa skillnader mellan de två delarna. Och även om skillnaderna inte är så stora att det alltid går att avgöra vilken del ett mycket kort avsnitt (här: enbart 400 ord) hör till, så går det ändå ofta att göra det. Slutsatsen av detta är att maskininlärning gällande språklig rytm i skönlitterär prosa är möjlig utifrån de kriterier vi har arbetat med, samtidigt som delstudien även bekräftar att det är en stor variation mellan de olika textavsnitten inom en del. Men framförallt tycker vi att denna delstudie är viktig ur ett metodologiskt perspektiv för att den visar på möjligheterna att använda maskininlärning för studier av skönlitteratur.

Tabell 4: Korrekthet för maskininlärning av de två delarna av Kallocain.

Mått	Värde
Min	33,3
Max	100
Medel	72,1
STDev	14,6
P10	44,4
P25	61,1
P50 (median)	66,7
P75	77,8
P90	88,9

(P=percentil.)

För att vidare analysera de här resultaten så undersöker vi betydelsen av de vikter som SVM-algoritmen har tilldelat de olika indikatorerna. Detta är en automatiserad process som är en central del av träningen där tilldelningen av vikt för indikatorerna anpassas till träningsdatan. Vikterna används internt av maskininlärningsalgoritmen för att avgöra vilka indikatorer som är viktiga för att veta vilken del ett avsnitt hör till. En hög positiv vikt för en indikator innebär att ett högt värde på indikatorn tyder på att avsnittet hör till andra delen, en hög negativ vikt på en indikator innebär motsatsen, att ett högt värde på indikatorn gör det troligt att avsnittet hör till första delen, och ett värde runt noll innebär att indikatorn inte är av någon stor vikt. I analysen av detta använder vi termen ”bästa försök” för de försök som har högst korrekthet. Tabell 5 visar vikter för dels de tre bästa försöken, dels medelvärde och standardavvikelse för de fem respektive tio bästa försöken. Värden för samtliga försök är i ännu högre grad spretiga, beroende på variationen mellan textavsnitten.

De tre bästa försöken har alla mycket liknande vikter. Här ser vi att indikatorerna med störst negativ vikt, som alltså tydligast indikerar att ett avsnitt hör till den första delen är ord med fler än 5 stavelser och TTR. Indikatorn med högst positiv vikt är andel interpunktion, vilket kan anses vara något förvånande då den skillnaden inte är signifikant mellan de två delarna som helhet, dock för stycken uppdelade på 1000 ord. Överlag har de flesta indikatorer negativ vikt här, vilket återspeglar de i huvudsak högre värdena i första delen för flertalet indikatorer. Totalt sett stämmer detta i relativt hög grad överens med den tidigare analysen, angående vilka skillnader som finns mellan indikatorerna. Även fraslängd har dock en ganska hög positiv vikt, och skulle därmed vara typiskt för den andra delen, vilket inte stämmer riktigt lika bra överens med våra tidigare resultat. Resultaten för de fem bästa försöken, och i ännu högre grad för de tio bästa försöken, har dock en betydligt större variation, vilket kan ses i de höga standardavvikelseerna, som ofta är större än medelvärdet. Detta visar att klassificeringen i hög grad beror på vilka avsnitt som hamnar i test- respektive träningsdata, och tyder på att det finns svårklassificerade avsnitt. Något som inte är särskilt förvånande med tanke på de resultat som har redovisats ovan i figurerna 2–5, där man tydligt ser att flera indikatorer varierar i hög grad genom romanen.

Tabell 5: Vikter för de olika indikatorerna för maskininlärning, för försöken med högst korrekthet.

	Försök 1	Försök 2	Försök 3	Medel 5 bästa	Standard- avvikelse 5 bästa	Medel 10 bästa	Standard- avvikelse 10 bästa
Stav5	-0,42	-0,38	-0,42	-0,25	±0,37	-0,07	±0,42
TTR	-0,36	-0,37	-0,40	-0,22	±0,34	-0,09	±0,39
MeningsL	-0,26	-0,25	-0,31	-0,18	±0,23	-0,07	±0,27
Långa	-0,29	-0,30	-0,30	-0,17	±0,27	-0,06	±0,32
Xlånga	-0,24	-0,15	-0,18	-0,10	±0,17	-0,03	±0,20
ordL/T	-0,15	-0,19	-0,14	-0,09	±0,15	-0,04	±0,17
Vidare	-0,09	-0,09	-0,11	-0,07	±0,05	-0,04	±0,06
OrdL/S	-0,04	-0,16	-0,08	-0,05	±0,09	0,00	±0,10
ejS	0,10	0,11	0,11	0,05	0,11	0,00	0,12
StyckeL	0,12	0,13	0,20	0,08	0,15	0,04	0,15
Expr	0,13	0,14	0,18	0,10	0,14	0,04	0,15
FrasL	0,36	0,33	0,37	0,20	0,30	0,08	0,32
Skilje	0,60	0,63	0,60	0,37	0,54	0,12	0,63
Korrektthet	100	100	100	100		94,4	

Sammanfattning och utblick

Karin Boyes *Kallocain* (1940) är berättelsen om ett framtida totalitärt samhälle där huvudpersonen och berättaren Leo Kall genomgår en kris som får honom att överge sin tidigare lojalitet med den övervakningsstat han tidigare har tjänat entusiastiskt. Stil- och språkforskaren Olof Gjerdmann menade i en uppsats 1942 att denna åsiktsförändring var så stor att den efter en nyckelscen ledde till en friare och böjligare rytm i romanen (Gjerdmann anger stället exakt med sida och rad). Tanken är inte oförenlig med Boyes egna uttalanden om språklig rytm, som hon ville knyta till djupare skikt i en personlighet och en persons hela fysiska beteende. Tesen har också väckt visst gillande bland litteraturhistoriska forskare, som liksom Gjerdmann har knutit den till två övergripande stiltendenser i Boyes författarskap, en ”stram” och en ”friare”. Men den har aldrig prövats ordentligt.

I den här uppsatsen prövar vi Gjerdmanns hypotes för att i vidare mening utveckla metoder för att beskriva och analysera språklig rytm i skönlitteratur. Syftet är att verka metodutvecklande inom det fält som kan kallas digital humaniora, och som i detta fall utgörs av skärningsfältet mellan litteraturvetenskap, språkvetenskaplig stilforskning och digital filologi. Vi har utgått från de digitala versionerna av Boyes verk i Litteraturbanken och har huvudsakligen använt oss av datorstödda metoder, inklusive maskininläring. Analysen gäller den språkliga rytmen, inte den narrativa, eftersom Gjerdmann och andra tycks ha avsett den förra. Den kvantitativa analysen har i vår undersökning på strategiska punkter kompletterats med kvalitativ.

Eftersom Leo Kall är berättaren i romanen har vi fokuserat berättarpartierna (och alltså huvudsakligen uteslutit repliker och den Censorns efterskrift som avslutar romanen) för att se om språkrytmen ändras. Romanen har först delats i två delar, enligt Gjerdmanns hypotes, som har jämförts med varandra. Den språkliga rytmen fångas med 16 indikatorer, varav de 7 första har föreslagits av Lisa Holm och resten har tillagts för denna undersökning: det rör sig om sådant som andel skiljetecken och deras art, meningslängd, ordlängd med mera (tabell 1). Analysen i detta skede visar att flertalet indikatorer tyder på att det faktiskt finns en rytmskillnad mellan de två partierna, och att inga indikatorer motsäger Gjerdmanns tes att rytmen i den senare delen skulle vara friare (tabell 2).

I en andra del analyseras texten i mindre avsnitt om cirka 1 000 ord vardera. Analysen visar att den språkliga rytmen böljar romanen igenom, med temporära starka utslag mot både det ”strama” och det ”böjliga” i bägge delarna, orsakade av de händelser som skildras och Kalls värderingar av dem. Utfallet visar att det utifrån en sådan analys inte vore uppenbart att dra gränsen för en rytmförändring just där Gjerdmann gör det – men också att det finns en tydlig tendens att de extremt långa orden blir färre romanen igenom (figur 2–4).

Ännu en delundersökning har ägnats replikerna i romanen. Vi har frågat oss om rytmen i dem är friare i de scener där människor talar under kalloccainrus. Här ser vi stora skillnader mellan olika typer av repliker för ett flertal indikatorer. Vi ser dock inget tydligt mönster totalt sett. Vi kan dock se att såväl repliker som ord är kortare hos de som talar fritt, det vill säga under kalloccainrus eller i Lindas monolog, än i övriga repliker.

I en fjärde delundersökning har vi prövat maskininläring för att ytterligare fördjupa analysen (här dock med en något reducerad grupp rytmindikatorer). Även om materialet är litet, visar denna delstudie att maskininläring i hög grad är möjlig att använda för att urskilja vilken del ett textavsnitt hör till – detta med reservation för att variationen mellan de korta textavsnitten är stor inom de två delarna (tabell 4–5).

I en mer begränsad mening visar undersökningen alltså att Gjerdmans hypotes om förändringen i språklig rytm inte är obefogad, även om bilden är mer komplicerad på detaljnivå. I vidare mening har den hjälpt oss att utveckla en uppsättning indikatorer för att fånga språklig rytm i skönlitterär prosa. Som ett metodiskt huvudresultat av undersökningen kan även konstateras att det datorstödda tillvägagångssättet verkligen kan bidra till att ge en mer detaljerad bild av språklig rytm i skönlitterär prosa.

Detta öppnar naturligtvis för framtida, motsvarande undersökningar av andra verk och andra författarskap. Avslutningsvis diskuterar vi därför några möjligheter på detta område.

En är naturligtvis rytmen i Karin Boyes skönlitterära prosa, då främst hennes romaner. Boye publicerade elva böcker under sin livstid, fyra diktsamlingar, två novellsamlingar och fem romaner. Samtliga romaner skrevs strax innan de publicerades, medan novellsamlingarna i regel samlar upp sådant som tidigare publicerats i tidningar och tidskrifter.

Som det väsentligaste materialet för en större undersökning ser vi romanerna, *Astarte* (1931), *Merit vaknar* (1933), *Kris* (1934), *För lite* (1936) och *Kalloccain* (1940). De fyra första är alla berättade i tredje person med en ”allvetande” (eller åtminstone mångvetande) berättare vars relation tidvis är främst registrerande och analyserande, tidvis fokaliserar via olika personers medvetanden. Ibland begränsas denna fokalisering till en enda person. Ibland finns också inslag av fri indirekt anföring (*erlebte Rede*, *free indirect discourse*). *Kalloccain* är alltså unik bland Boyes romaner genom att vara berättad i första person.

Grundstrukturen i alla Boyes romaner är att de är berättade i huvudsakligen kronologisk ordning (om än med vissa återblickar). Det ger samtliga romaner ett tydligt urskiljbart nuplan. Samtidigt rymmer vissa romaner inskottstexter av liknande slag som censorns efterskrift i *Kalloccain*. Det viktigaste exemplet i Boyes romanproduktion är de partier i *Kris* som rubriceras ”dialoger” och illustrerar Malin Forsts problematik i pro-

salyrisk, symbolisk och dramatisk form. I en framtida analys bör enligt vår uppfattning sådana inslag analyseras för sig, skilda från den övriga romantexten.⁵⁰

Ett återkommande mönster i Boyes romaner – möjligen med undantag för *Astarte* – är att huvudpersonen genomgår en inre frigörelse, eller åtminstone en känslomässig upptining liknande den i *Kallocain*. I *Merit vaknar* överger huvudpersonen gradvis den falska idealiseringen av sin avlidne man, i *Kris* frigör sig Malin Forst från auktoriteternas försök att styra hennes liv och i *För lite* återknyter Harald Mårman till de verkliga skapande impulser som han har förträngt för brödskriveri. Därför finns det skäl att studera om också dessa romaner rymmer någon förskjutning i rytmen. Även om också dessa romaner rymmer partier som kan ses som nyckelscener är det vanskligt att som Gjerdman peka ut ett exakt ställe där en förändring sker. En möjlighet är istället att dela upp berättarpartierna i två eller flera lika stora delar för att kunna jämföra rytmen i de olika delarna inom och mellan romanerna.

Frågorna skulle då gälla rytm och rytmisk grundstruktur i prosaberättande inom ett författarskap. Den kan naturligtvis utvidgas till att omfatta fler författarskap, både med försök att fånga deras särart (och en eventuell utveckling över tid) och eventuella likheter mellan författargenerationer. Det material som nu byggs upp i Litteraturbanken utgör en utmärkt utgångspunkt för sådana storskaliga analyser av rytm i skönlitterär prosa.

Till sist vill vi peka på möjligheten att ytterligare utvidga antalet rytmindikatorer. En möjlighet är exempelvis att ha indikatorer som direkt representerar variation. Här är en bra utgångspunkt Marianne Nordmans analys av variation i fraslängd, meningslängd och styckelängd, som relativt enkelt bör gå att automatisera.⁵¹ Vi anser även att det skulle vara relevant att koppla språklig rytm till syntaktisk komplexitet, och här skulle den syntaktiska dependensanalysen kunna utnyttjas i högre grad, för en jämförelse mellan olika texter.

BILAGA. Textavsnitten nr 1–35 i *Kallocain*

Nr	Omfattning (sida:rad)	Inledning / avslutning
1.	5:1–9:21	”Den bok jag [---] vänta på Linda.”
2.	9:21–15:2	”Eftersom jag var [---] samband mellan dem.”
3.	15:2–22:7	”Jag ville [---] blekvioletta lågan.”
4.	22:8–27:16	”Vi ställde fram [---] egenskap av polissekreterare.”
5.	27:16–32:9	”Det betydde [---] hemlig och behärskad.”
6.	32:9–38:2	”Samma plåga [---] en serie pilskott.”
7.	38:2–45:12	”Jag kastade en [---] Linda själv.”
8.	45:12–51:18	”Jag började ana [---] hörde till yrket.”
9.	51:19–63:5	”Mannen sträckte [---] som skämdes.”
10.	63:5–73:2	”Mannen framför oss [---] slarv i sjukvården.”
11.	73:2–80:19	”Man kunde nog [---] lugn och ro.”
12.	80:12–85:20	”De tio efterkallade [---] radions ursäktstimme.”
13.	86:1–94:24	”Experimentet tog genast [---] sinnade medsoldater.”
14.	94:24–103:6	”Rissens ord [---] mena med läggdags.”
15.	103:6–117:8	”Barnen hade redan [---] uppoffrade lediga kvällar.”
16.	117:8–126:22	”Togo Baharas [---] obestämda framtidsplaner.”
17.	126:22–138:13	”Samtidigt medgav han [---] mening i dem.”
18.	138:13–144:13	”I varje fall [---] vilken den nu var.”
19.	144:13–156:20	”Lyckligtvis hölls [---] vägde vartenda ord.”
20.	156:23–164:11	”Just då tror jag inte [---] halvt mänsklig.”
21.	164:11–172:7	”Inte ens en stor [---] första grovgallring.”
22.	172:7–182:9	”En rad institutioner [---] allt motstånd.”
23.	182:10–189:4	”Kom ihåg att [---] på laboratorierummen.”
24.	189:5–201:7	”En gång kom vi att [---] jag verkligen hade.”
25.	201:8–211:16	”Det enda jag vågade [---] mitt eget kallocain.”
26.	211:16–216:10	”Det hände att jag [---] en gång i framtiden.”
27.	216:11–223:9	”Då jag kom [---] inte hindra vad som skedde.”
28.	223:9–231:2	”Två människor ensamma [---] utan att vända sig bort.”
29.	231:2–235:17	”Hennes mun [---] symbol för alltsammans.”
30.	235:17–242:1	”Och det äcklade mig [---] jag vet inte var.”
31.	242:1–256:12	”Nu: jag visste ingenting [---] honungslen artighet.”
32.	256:20–265:2	”Först senare slog mig [---] de utgående att stanna.”
33.	265:10–273:10	”Jag hade stigit ner [---] liten och hopkrupen.”
34.	273:10–278:3	”Tystnaden på takterrassen [---] överrösta larmet.”
35.	278:6–283:4	”Antingen de inte hörde [---] visste jag inte.”

NOTER

- 1 Uppsatsen har skrivits inom ett pågående samarbetsprojekt vid Uppsala universitet: "Från närläsning till fjärrläsning: digitala humaniora och nya former för textanalys". Projektet är tvåårigt (2017–2019) under ledning av Johan Svedjedal och bedrivs med finansiering från Områdesnämnden för humaniora och samhällsvetenskap vid Uppsala universitet. Forskargruppen består av Mats Dahllöf, Joakim Nivre och Sara Stymne (Institutionen för lingvistik och filologi), David Håkansson och Carin Östman (Institution för nordiska språk), samt Karl Berglund och Johan Svedjedal (Litteraturvetenskapliga institutionen). Vi tackar de övriga inom projektgruppen för värdefulla synpunkter under arbetets gång, liksom de två anonyma referenter som har läst uppsatsen för *Sammlaren*.
- 2 Om kursen, Olof Gjerdmán, "Rytm och röst", *Karin Boye. Minnen och studier*, Utgivna av Margit Abenius och Olof Lagercrantz, Stockholm 1942, s. 143–160, särsk. s. 145.
- 3 Gjerdmán 1942, s. 155.
- 4 Gjerdmán 1942, s. 183.
- 5 Till de viktigaste bidragen om *Kalloccain* hör Karin Tivenius & Jan C. Bouman, "Kalloccain: Karin Boyes mardröms-autobiografi. En litterär-psykologisk studie", *Edda*, 61, 1961, s. 177–205 (dock mycket metodiskt tveksam i sitt psykologiserande); Anton Fjeldstad, "Stilen i Karin Boyes roman 'Kalloccain'", *Edda*, 72, 1972, s. 101–116; Anton Fjeldstad, "Allusion och myte i Karin Boyes *Kalloccain*", *Tidskrift för litteraturvetenskap*, 1972/1973:4, s. 240–248; Helena Forsås-Scott, "Död och liv i Karin Boyes *Kalloccain*", *Edda*, 84, 1984:4, s. 202–218; Åke Sandahl, "Kalloccain och Vi – två dystopier", *Karin Boyes liv och diktning*, XI. *Minnesskrift*, Huddinge 2006, s. 39–47; och partierna om *Kalloccain* i Barbro Gustafsson Rosenqvist, *Att skapa en ny värld. Samhällssyn, kvinnosyn och djuppsykologi hos Karin Boye*, Stockholm 1999.
- 6 Staffan Björck, *Romanens formvärld. Studier i prosaberättarens teknik*, Stockholm 1970 [1953], s. 39.
- 7 Litteraturbanken, litteraturbanken.se/start.
- 8 Margit Abenius, *Drabbad av renhet. En bok om Karin Boyes liv och diktning*, Stockholm 1950, s. 349.
- 9 Fjeldstad 1972. Termerna "organiserat" och "organiskt", s. 105–106; om stereotypa och spontana uttryck, s. 108–114.
- 10 Uppgifterna nedan bygger på sökning av orddelen *-rytm-* i versionerna av Karin Boyes elva livstidspublicerade böcker i Litteraturbanken; n=27. De aktuella böckerna är (samtliga med förlagsort Stockholm): *Moln. Dikter* (1922); *Gömnda land. Dikter* (1924); *Härdarna. Dikter* (1927); *Astarte. Roman* (1931); *Merit vaknar* (1933); *Uppgörelser. Skisser och noveller* (1934); *Kris* (1934); *För trädets skull. Dikter* (1935); *För lite. Roman* (1936); *Ur funktion. Noveller* (1940); *Kalloccain. Roman från 2000-talet* (1940).
- 11 Karin Boye, *Härdarna. Dikter*, Stockholm 1927, s. 22.
- 12 Om förälskelsen och den aktuella dikten, se t.ex. Johan Svedjedal, *Den nya dagen gryr. Karin Boyes författarliv*, Stockholm 2017, s. 206.
- 13 "Lägg livet i kistor!", ur novellen "För tung", Karin Boye, *Uppgörelser. Skisser och noveller*,

- Stockholm 1934, s. 127; ”en linje inifrån”, ur novellen ”Spårvagnen och paradiset”, Karin Boye, *Ur funktion. Noveller*, Stockholm 1940, s. 263.
- 14 Karin Boye, ”Språket bortom logiken”, *Spektrum* [2:6] [hösten], 1932, s. 5–11, s. 6. Se även, ”Tendens och verkan”, *Dikten, diktaren och samhället*, Stockholm 1935, s. 43–53, samt ”Diktning och moral”, *Karavan*, 1935:3, s. 106–110.
- 15 Jfr Svedjedal 2017, s. 548.
- 16 E.M. Forster, *Aspects of the Novel*. Edited by Oliver Stallybrass, Harmondsworth 1979 [1927], s. 146.
- 17 Forster 1979, s. 149.
- 18 Forster 1979, s. 149.
- 19 Björck 1970, s. 38–40; jfr t.ex. s. 224–226.
- 20 Franz K. Stanzel, *Theorie des Erzählens*, 2. verbesserte Auflage (Uni-Taschenbücher 904), Göttingen 1982 [1979], s. 98–103.
- 21 Gérard Genette, *Narrative Discourse. An Essay in Method*. Translated by Jane E. Lewin. Foreword by Jonathan Culler, Ithaca, New York 1980 [franska orig. 1972 som en del av författarens *Figures III*], främst s. 87–88, 97, 109–110.
- 22 Karl Berglund, ”Killer plotting. Typologisk intriganalys utifrån fjärrläsningar av 113 samtida svenska kriminalromaner”, *Tidskrift för litteraturvetenskap*, 2017:3/ 4, s. 41–68.
- 23 Tanya Clement, ”The story of one: Narrative and Composition in Gertrude Stein’s *The Making of Americans*”, *Texas Studies in Literature and Language*, 54, 2012:3, s. 426–448.
- 24 Nils Svanberg, *Svensk stilistik. En handbok för stilhistoriska studier*, Stockholm 1936, s. 145, 144.
- 25 Gösta Holm, *Epoker och prosastilar. Några linjer och punkter i den svenska prosans stilhistoria* (Lundastudier i nordisk språkvetenskap. Serie A, 17), Lund 1967, t.ex. s. 105, 139, 203.
- 26 Marianne Nordman, ”Rytm i prosa. En analysmodell och dess tillämpning”, i *Bidrag till en nordisk metrik*, vol. 1 (Skrifter utgivna av Centrum för Metrisk Studier 9), Göteborg 1999, s. 157–253.
- 27 Lisa Holm, ”Rytm i romanprosa. En studie av rytmiska signalement i tio samtida svenska romaner”, i *Det skönlitterära språket. Tolv texter om stil*, red. Carin Östman, Stockholm 2015, s. 215–235.
- 28 För en utförligare diskussion av denna terminologi, se Johan Svedjedal, *Almqvist – berättaren på bokmarknaden. Berättartekniska och litteratursociologiska studier i C. J. L. Almqvists prosafiktion kring 1840* (Skrifter utgivna av Avdelningen för litteratursociologi vid Litteraturvetenskapliga institutionen i Uppsala 21), Uppsala 1987, s. 41–47. Termen ”inskottstext” är en svensk översättning av Susan Sniader Lansers ”inserted text” (Susan Sniader Lanser, *The Narrative Act. Point of View in Prose Fiction*, Princeton 1987, s. 134–135).
- 29 Swedish annotation pipeline, hämtad från github.com/robertostling/efselab (8.6.2018).
- 30 Med parsning avses att automatiskt göra en grammatisk analys.
- 31 Robert Östling, ”Part of Speech Tagging: Shallow or Deep Learning?”, *North European Journal of Language Technology*, 5, 2018:1, s. 1–15.
- 32 Universal Dependencies, se universaldependencies.org; om MaltParser, se Joakim Nivre; Johan Hall & Jens Nilsson, ”MaltParser. A Data-Driven Parser-Generator for Depen-

- dency Parsing”, i *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC2006)*, 2006, s. 2216–2219.
- 33 Eftersom vi inte har någon manuellt uppmärkt text från Kallocain går det inte att säga exakt hur hög andelen fel är. I allmänhet har dock ordklasstaggare mycket låg andel fel, för *efselab* som används i studien rapporteras en andel fel på 3,7–5,0% för två olika svenska korpusar (Östling 2018, s. 1–15). För dependensträd är felnivån något högre; i egna experiment där vi testat MaltParser för de tre svenska korpusarna i Universal Dependencies, har vi fått en korrekthet på 78,0–83,5% för själva träden (mättet UAS) och 70,3–80,2% för träden inklusive etiketter (mättet LAS). Vi har även använt dependensinformationen i en begränsad grad här, främst för att hitta subjekt, för vilket sådana verktyg i regel gör ännu färre fel.
- 34 Ett undantag är det stränga myndighetsbrev som tillställs Leo Kall med en uppmaning till honom att göra offentlig avbön efter ett uttalande som kan uppfattas som illojalt (s. 38–39).
- 35 Om LIX, se Carl-Hugo Björnsson, *Läsbarhet*, Stockholm 1968.
- 36 K. Mühlenbock & S.J. Kokkinakis, ”LIX 68 revisited – an extended readability”, i *Proceedings of the Corpus Linguistics Conference*, Liverpool 2009.
- 37 Fjeldstad 1972.
- 38 Att direkt mäta andelen upprepningar är inte helt lätt, eftersom det kräver en bestämd definition av när ett ord räknas som upprepat, hur nära det upprepade ordet det ska stå, vilka typer av ord som ska räknas (upprepningar av funktionsord torde vara mindre intressant än upprepningar av innehållsord), med mera. För en diskussion av vissa problem med hur upprepningar i form av epifora och epanafora kan definieras, se Marie Dubremetz, *Detecting Rhetorical Figures Based on Repetition of Words: Chiasmus, Epanaphora, Epiphora*, Uppsala 2018.
- 39 Avsnittslängden på 2000 ord är godtyckligt vald. Andra avsnittslängder ger liknande resultat. Så länge avsnitten som TTR beräknas på är lika långa anses de jämförbara, däremot inte om vi hade jämfört textavsnitt av olika längd, vilket skulle skett ifall vi beräknat TTR på all text i respektive del.
- 40 Det ska observeras att våra indikatorer alltså inte rymmer kategorin betonad/obetonad stavelse.
- 41 Vår metod har gjort att vi inte med säkerhet har kunnat skilja mellan huvudsats och bisats, och vi frångår alltså något Holm för denna indikator. Där hon kunde sortera ut meningar med minst en huvudsats har vi fått nöja oss med att sortera ut meningar med minst en sats (oberoende av om det är huvud- eller bisats).
- 42 Mann-Whitney-testet är ett icke-parametriskt test som kan användas för att se om värden från en population är signifikant högre eller lägre än värden ur en annan population. I det här fallet är populationerna de två delarna av *Kallocain*. Till skillnad från de flesta andra signifikanstest för att undersöka detta finns inget antagande om att värdena är normalfördelade, vilket vi i det här fallet inte vet om de är. Chi-två-test kan användas för att undersöka proportioner, för att se om de förväntade frekvenserna för ett fenomen skiljer sig från de observerade. För att chi-två-testet ska vara korrekt bör frekvensen för varje observation

- (till exempel antal kolon i respektive del) vara som lägst 5, vilket är fallet för alla våra indikatorer.
- 43 Eftersom avsnitten ska vara av ungefär samma längd kommer romanens sista rader inte med i analysen, eftersom detta återstående avsnitt är för kort (s. 283, rad 5–22).
- 44 Replikerna under kallocairus finns på s. 54–62, 91–94, 120–123, 127–130, 132–136, 137, 227–230 och 260–264.
- 45 Lindas repliker finns på s. 241–249 och 252–253.
- 46 Identifikation av författaren till en text, se t.ex. Matthew L. Jockers & Daniela M. Witten, "A comparative study of machine learning methods for authorship attribution", *Literary and Linguistic Computing*, 25, 2010:2, s. 215–223. Identifikation av en författares kön, se t.ex. Moshe Koppel, Shlomo Argamon & Anat Rachel Shimoni, "Automatically Categorizing Written Texts by Author Gender", *Literary and Linguistic Computing*, 17, 2002:4, s. 401–412.
- 47 Med träningsdata avses att klassen är känd på förhand.
- 48 För fler detaljer om hur algoritmen fungerar, se exempelvis T. Hastie, R. Tibshirani & J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2 ed., New York 2009.
- 49 R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang & C.-J. Lin, "LIBLINEAR: A library for large linear classification", *Journal of Machine Learning Research* 9, 2008, s. 1871–1874.
- 50 De aktuella partierna är (allt i Boye, *Kris* [1934]): "Dialog I: Om sunda ideal" (s. 15–28); "Vilken ond hövding..." (s. 29–30); "Makternas schackklubb" (s. 31–41); "Ur ett brev från Dagny Ritzelius" (s. 81–86); "Ur en dagbok" (s. 87–89); "Dialog II. Om de fromma ordens betydelse" (s. 147–163); "Jaså. Tror ni jag tänker..." (s. 183–193); "På vilken undrets tröskel..." (s. 194–196); [Vit och svart] (s. 239–240); "Dialog III. Om Malin Forst" (s. 241–251). Det ska erkännas att gränserna mellan olika typer av material inte alltid är enkla att dra, främst eftersom nuplanet ibland rymmer korta insprängda partier av utpräglad prosalyrisk karaktär. Om den litterära metoden i dessa avsnitt, se främst Gunilla Domelöf, *I oss är en mångfald levande. Karin Boye som kritiker och prosamodernist* (Acta Universitatis Umensis. Umeå Studies in the Humanities 77), Umeå 1986, s. 231–261.
- 51 Nordman 1999.

ABSTRACT

Sara Stymne, Department of Linguistics and Philology, Uppsala University
Johan Svedjedal, Department of Literature, Uppsala University
Carin Östman, Department of Scandinavian Languages, Uppsala University

Linguistic Rhythm in Narrative Prose: the case of Karin Boye's *Kallockain* (Språklig rytm i skönlitterär prosa. En fallstudie i Karin Boyes *Kallockain*)

The concept of rhythm in prose is ambiguous, and there is no consensus on how to define it. In this work, we focus on linguistic rhythm, at word, sentence and paragraph levels. We adopt and slightly extend rhythm indicators used in previous research, and show that these can be calculated fully automatically, on a much larger scale than previously done.

We adopt the Swedish poet and novelist Karin Boye's (1900–41) novel *Kallockain* (1940), as a case study. It is an icily dystopian depiction of a totalitarian future, where the protagonist Leo Kall first embraces this system, but for various reasons later rebels against it. The peripety comes when he gives a public speech, questioning the State. It has been pointed out that the novel from precisely this point on is characterized by a much freer rhythm, and that Boye as an author had considerable interest in questions of linguistic rhythm. This paper sets out to test this hypothesis by applying sixteen indicators of linguistic rhythm in narrative prose (such as word length, sentence length, ratio of punctuation, etc.).

We first note that we can expect differences between narrative and dialogue and limit most of our study to the first-person narrative. We find that there are significant differences mainly between phrase and word lengths in the parts before and after Leo Kall's conversion. In a further investigation we note that there is also great variation among indicators within each part of the novel. We also show that machine learning can be used to differentiate small segments from each part of the novel, with higher accuracy than a random classifier. Finally, we undertake a small study of dialogue, which, however is mainly inconclusive. In summary we find some support for the claim that there is a rhythm break in *Kallockain*. We also believe that our study is important from a methodological point of view, since it provides a method for large-scale studies of prose rhythm in the future.

Keywords: Karin Boye, Literary stylistics, Linguistic rhythm in narrative prose, Stylometry, Digital Humanities