Universelle representasjoner av norske klokkeslett

Masteroppgave av Gisle Ytrestøl

februar 2006

Institutt for lingvistikk og litteraturvitenskap
Universitetet i Bergen

Universelle representasjoner av norske klokkeslett

Gisle Ytrestøl

Masteroppgave i datalingvistikk

Veileder: Helge Dyvik

februar 2006

Institutt for lingvistikk og litteraturvitenskap
Universitetet i Bergen

Sammendrag

Denne avhandlingen presenterer en komputasjonell grammatikk for norske klokkeslettsuttrykk. Syntaksen til klokkeslettsuttrykk skiller seg sterkt fra andre domener i den norske grammatikken og har ofte en idiomatisk form. Selv om norsk grammatikk generelt er godt kartlagt i den komputasjonelle grammatikken NorGram, er ikke dette domenet behandlet tidligere. Intensjonen er at mitt tillegg skal implementeres i NorGram, og i så måte forbedre denne grammatikken.

Den ferdige grammatikken bygger på en inngående studie av hvilke faktorer i setningen som virker inn på syntaksen til et klokkeslettsuttrykk. På bakgrunn av denne studien ble en minigrammatikk implementert på XLE-plattformen. Denne avhandlingen dokumenterer studien av klokkeslettsuttrykkenes syntaks, og den komputasjonelle grammatikken som analyserer disse uttrykkene vil bli presentert.

Abstract

This thesis presents a computational grammar for Norwegian time of day expressions. There is a firm distinction between the syntax of time expressions compared to other domains in Norwegian grammar, as expressions of time behave more idiomatically. Although a computational grammar for Norwegian exists, called NorGram, its grammar does not recognise time of day expressions. My intention is that this addition will be implemented in NorGram, thus enhancing the range of its grammatical coverage to include these expressions.

The finished grammar is based upon a study of the syntax of time expression. This study provides the background for the implementation of a miniature grammar on a XLE platform. This thesis documents the study of time of day expressions, and presents the computational grammar that analyses the expressions.

Forord

Denne avhandlingen inngår i forskningsprosjektet LOGON (Leksikon, Ordsemantikk, Grammatikk og Oversettelse for Norsk), som er et samarbeid mellom universitetene i Bergen, Oslo og Trondheim. Prosjektets målsetning er å bygge en demonstrator som oversetter norske setninger til engelsk innenfor et avgrenset domene. Min avhandling utgjør en masteroppgave i datalingvistikk ved Institutt for lingvistikk og litteraturvitenskap ved Universitetet i Bergen, og mitt bidrag til prosjektet er gjennom LOGON delfinansiert av Norges forskningsråd.

Det har vært lærerikt å arbeide på et større prosjekt, og jeg er svært glad for at jeg har fått muligheten til dette. Først og fremst vil jeg rette en takk til min veileder Helge Dyvik. Dyvik introduserte meg for oppgaven, og har vært til uvurderlig hjelp underveis i prosessen. Jeg vil også takke Paul Meurer, som har hjulpet meg med å legge grammatikken ut på nettet. Til slutt vil jeg takke Helge og Birgitte for deres innsats som korrekturlesere, og ikke minst Kari Mathilde for grundige gjennomlesninger, oppmuntring og støtte i løpet av året.

Bergen, februar 2006

Gisle Ytrestøl

Innhold

Kapittel 1

Introduksjon

1.1 Innledning

Det foreligger i dag gode verktøy for å parse norske setninger. NorGram-prosjektets (se avsnitt 2.5.1) målsetning var å "utvikle en gjenbrukbar komputasjonell grammatikk for norsk". Grammatikken som ble laget av NorGram brukes i dag i LOGON (se avsnitt 2.5.3). Et domene innen den norske grammatikken som ikke har fått en analyse, er klokkeslettsuttrykk. Min masteroppgave består i å lage adekvate analyser for disse klokkeslettsuttrykkene. Disse analysene skal enkelt kunne la seg implementere i NorGrams overordnede grammatikk, og benyttes i LOGON-prosjektet.

Innenfor LOGON skal de norske analysene representeres med c- og f-strukturer i tråd med LFG-formalismen, og analysene skal få en MRS-projeksjon. Min grammatikk følger denne konvensjonen, og hvordan utmatingen av en vellykket parse i min grammatikk ser ut, vil bli gjennomgått i detalj i kapittel 4. MRS-representasjonen er i hovedsak utviklet av min veileder Helge Dyvik på bakgrunn av de vurderingene jeg har gjort i forhold til hvordan klokkeslettsuttrykkene bør representeres i c- og f-strukturen.

Resultatet av masteroppgaven min består av en minigrammatikk som analyserer enkle setninger med et klokkeslettsuttrykk. Denne minigrammatikken er vedlagt i tillegg B-D. Grammatikken er skrevet på XLE-plattformen (se avsnitt 2.3). Selve grammatikken vil bli integrert i LOGONs overordnede grammatikk av Helge Dyvik, som er ansvarlig for analyse- og grammatikkdelen i LOGON-prosjektet. Den ferdige grammatikken er også tilgjengelig på nettet: **http://hilarion.aksis.uib.no:8030/logon/xle.xml**

1.2 Klokkeslettsuttrykkene

Et klokkeslett refererer til et tidspunkt på dagen. Det kan være en omtrentlig (1) eller en nøyaktig referanse (2), eller et klokkeslettsuttrykk hvor selve timen på døgnet er uuttrykt, og tidspunktet i seg selv er uspesifisert (3).

En konkret definisjon av hva jeg regner som et klokkeslettsuttrykk i denne oppgaven, finnes i avsnitt 4.1.3.

(1) Jeg kommer *litt over halv sju*.[1]

(2) Jeg kommer *halv sju*.

(3) Bussen går *ti over*.

I skriftlig representasjon av norske klokkeslettsuttrykk går det et skille mellom en bokstavelig (4) og en digital uttrykksform (5).

(4) klokka fem på halv elleve

(5) klokka 22.25

Jeg har bare konsentrert meg om å analysere den bokstavelige uttrykksformen av klokkeslettsuttrykk. Det er denne representasjonen som volder mest problem, i og med at den digitale følger en fast, relativt forutsigbar struktur, uansett hvilket klokkeslett som representeres. Imidlertid må det legges til at i skriftlig norsk er den digitale representasjonen av et klokkeslett vanligere enn en bokstavelig representasjon[2], og også disse representasjonene bør gis en analyse før man kan si at NorGram-grammatikken er i stand til å analysere norske klokkeslettsuttrykk på en tilfredstillende måte.

1.3 Grammatikaliteten til uttrykkene

Norsk grammatikk generelt er godt kartlagt, for eksempel i *Norsk referansegrammatikk* (Faarlund et al., 1997). Imidlertid har jeg ikke funnet noen grammatikker som inkluderer regler for syntaksen til bokstavelige klokkeslettsuttrykk som i eksempel (4). Innenfor dette domenet har jeg derfor selv laget reglene som definerer hvilke konstruksjoner som er grammatiske, og hvilke som ikke er det. I dette arbeidet har jeg brukt *Norsk aviskorpus* (NA, URL) og *Oslo-korpuset av taggede norske tekster* (OK, URL), og sett på forskjellige strukturer et klokkeslettsuttrykk kan inneha. Disse strukturene har jeg inkludert i min grammatikk, samtidig som jeg har arbeidet for å unngå uønsket overgenerering.

Jeg har jobbet ut ifra en deskriptiv framgangsmåte og latt forekomster av skriftlig uttrykkformer være retningsgivende for hvorvidt uttrykket skal tolkes som grammatisk. Fra dette har jeg skrevet generelle regler som beskriver de fenomenene jeg har observert. Men at en uttrykksform ikke finnes i korpus beviser ikke at uttrykksformen er uakseptabel. Jeg har derfor ved flere tilfeller brukt subjektivt skjønn for å vurdere hvilke uttrykk som skal anses som akseptable, og hvilke som skal forkastes.

[1] Klokkeslettsuttrykket er kursivert.

[2] Denne slutningen baserer jeg på korpusstudier.

1.3.1 Idiom?

Så langt det har vært mulig har jeg unngått å hardkode[3] ett og ett uttrykk. Jeg har heller skrevet generelle regler for hvilke ordklasser/undergrupper av ordklasser som kan stå i de forskjellige kontekstene. Så langt det er mulig, har jeg benyttet meg av eksisterende regler innen NorGram. Et slikt generaliseringsprinsipp er noe mer problematisk innenfor domenet klokkeslettsuttrykk enn det er i den øvrige norske grammatikken. Årsaken er at mange klokkeslettskonstellasjoner synes å være idiomatiske. I konstellasjoner som (6) brukes konvensjonelt bare tallordene *ett* og *tre* som kvantor for *kvarter*. Jeg har imidlertid åpnet for at alle kvantorer kan stå i denne posisjonen, og dermed ikke tolket *ett kvarter* og *tre kvarter* som rene idiom.

(6) klokka fem kvarter over ni

(7) klokka nitten over halv seks

Disse avgjørelsene fører til at grammatikken min analyserer en rekke uttrykk som vil oppfattes som unaturlige i de fleste sammenhenger, og ikke vil kunne gjenfinnes i et korpus. Å skille mellom hva som er ugrammatisk, hva som er nonsens, og hva som konvensjonelt ikke brukes i klokkeslettsuttrykk er vanskelig, og dette blir en subjektiv vurdering. Uttrykk som (6) og (7) analyseres som grammatiske på bakgrunn av at alle tallord tillates i posisjonen hvor *fem* og *nitten* står, selv om det konvensjonelt bare er noe få tallord som pleier å stå i denne posisjonen.

For å definere hva som grammatisk og hva som ikke er det, har jeg måttet foreta del valg. De fleste av disse valgene er gjort etter en vurdering opp mot egen språkfølelse, da det ikke foreligger litteratur som omhandler disse strukturene på norsk. Eksempler på klokkeslettsuttrykk jeg har definert som ugrammatiske er:

(8) *kvart etter åtte (i den intenderte betydningen *kvart over åtte*)

(9) *litt på åtte (i den intenderte betydningen *litt før åtte*)

I forhold til eksempel (6) og (7), foreligger det sterkere lingvistiske grunner for ikke å tillate (8) og (9). (8) er ugrammatisk fordi *kvart* ikke kan sies å representere *femten minutter* i denne konteksten. Slik grammatikken er definert, kan *kvart* bare tillegges denne betydningen internt i PP-TIME-frasen (se avsnitt 4.3), det vil si at preposisjonen må være *på* eller *over*. Mer kontroversiell er muligens regelen som definerer (9) som ugrammatisk. Selve regelen sier at gradsadverbial som *litt* ikke kan stå foran preposisjonen *på* i et klokkeslettsuttrykk (se avsnitt 3.3.1). Denne regelen støttes av korpusene jeg har brukt, hvor slike tilfeller ikke eksisterer. Rent tekstsøk gjennom *http://www.google.com* avslører imidlertid flere forekomster av strukturer til-

[3] Å skrive en spesifikk regel for hvert uttrykk.

svarende (9), noe som viser denne konstruksjonen likevel ikke er helt uvanlig. Forekomstene skriver seg imidlertid fra gjestebøker på nettet, blogger og lignende uformelle sammenhenger.

Innen rammene til LOGON-prosjektet har jeg ikke sett at det vil være tjenlig å tillate forekomster av svært muntlige former dersom dette kan føre til mange overgenereringer på andre områder. Imidlertid er det ikke noe problem å gå inn i koden ved en senere anledning for å forandre dette, for eksempel dersom NorGram skal brukes til å parse mer muntlig språk.

1.4 Presentasjon av oppgaven

Kapittel en

I dette første kapitlet introduseres problemstillingen for oppgaven. Jeg drøfter også hvilke kriterier jeg legger til grunn når jeg definerer et klokkeslettsuttrykk som grammatisk eller ugrammatisk. Videre i kapitlet presenteres temaene for de andre kapitlene.

Kapittel to

Her beskriver jeg den lingvistiske teorien jeg har støttet meg til i analysen av klokkeslettsuttrykkene. Denne teorien er LFG – Lexical Functional Grammar. LFG-grammatikken representeres på XLE-plattformen, og det blir gitt en kort innføring i hvordan XLE-koden leses. Jeg beskriver også MRS, som er en representasjonsmåte for å beskrive semantiske strukturer i et naturlig språk. Videre skriver jeg om de prosjektene som har relevans for oppgaven min. Mitt ferdige resultat skal anvendes av LOGON-prosjeket (se avsnitt 2.5.3). LOGON-prosjektet anvender eksisterende lingvistiske verktøy fra prosjektene NorGram og ParGram. Min analyse skal etter planen også integreres i grammatikken fra disse prosjektene, og jeg har derfor gitt en kort presentasjon av dem.

Kapittel tre

Kapittel tre er todelt. Den første delen omhandler klokkeuttrykk som tidsadverbialet, mens den siste delen omhandler preposisjonenes rolle i et klokkeslettsuttrykk. Avsnitt 3.2 er en teoretisk gjennomgang av klokkeslettsuttrykkets posisjon som en undergruppe av tidsadverbialene. I særlig grad vil jeg gå inn på hvordan klokkeslettsuttrykkets grammatiske funksjon i setningen er med å bestemme de syntaktiske restriksjoner uttrykket underlegges. Jeg vil vise til eksisterende lingvistiske teorier for tidsadverbial og tidsuttrykk, og baserer mine analyser på disse teoriene så langt det er mulig.

I del to av kapitlet, avsnitt 3.3, ser jeg på hvilken rolle preposisjonene spiller i et klokkeslettsuttrykk. Noen preposisjoner forandrer egenskaper innenfor det avgrensede domenet et klokkeslettsuttrykk utgjør. Andre preposisjoner tar med seg sine inherente egenskaper inn i klokkeslettsuttrykkene og styrer syntaksen til klokkeslettsuttrykket gjennom de restriksjonene de bringer inn i uttrykket. I dette avsnittet vil jeg gjennomgå hver enkelt preposisjon som kan stå i et klokkeslettsuttrykk, og ta for meg hvordan disse preposisjonene styrer syntaksen i uttrykket.

Kapittel fire

I dette kapitlet presenteres selve analysen av klokkeslettsuttrykkene. Den lingvistiske teoretiseringen jeg gjorde i kapittel tre, anvendes praktisk i implementeringen av klokkeslettsuttrykkene på XLE-plattformen. Jeg går i detalj gjennom utmatingen som gis i c- og f-struktur for en vellykket prising av et klokkeslettsuttrykk, og viser til XLE-reglene som utløser denne analysen. I avsnitt 2.4 presenterer jeg den grafiske MRS-representasjonen som genereres av XLE. Jeg kommer i liten grad til å diskutere den teoretiske begrunnelsen for hvorfor MRS-representasjonen ser ut som den gjør, siden det ikke er jeg selv som har hatt hovedansvaret for å utforme denne representasjonen.

Kapittel fem

I det avsluttende kapitlet evaluerer jeg det arbeidet jeg har gjort. Jeg vil her foreslå mulige utvidelser som kan gjøres for å optimalisere klokkesletts-grammatikken, samt drøfte mulige problemstillinger i tilknytning til implementeringen av min grammatikk i NorGram.

Kapittel 2

Teori og prosjekter

2.1 Innledning

I dette kapitlets første del (avsnitt 2.2) vil jeg beskrive det teoretiske grunnlaget jeg har basert denne oppgaven på. Siden min analyse skal integreres i større prosjekter, har jeg lagt til grunn den teorien som disse prosjektene benytter. Den lingvistiske teorien er LFG – leksikalsk-funksjonell grammatikk. I dette kapitlet vil jeg gi en presentasjon av de elementene i LFG som har vært viktigst for analysen jeg har gjort. Analysene av klokkeslettsuttrykkene er gjort på XLE, som er en plattform for grammatikkutvikling. I avsnitt 2.3 gir jeg en kortfattet beskrivelse av hvordan implementeringen fra LFG-analyse til XLE-notasjonen har foregått. Klokkeslettsanalysene er også representert i MRS-strukturer. I avsnitt 2.4 går jeg derfor igjennom de grunnleggende prinsippene til MRS.

De prosjektene som er relevante i forhold til denne oppgaven blir presentert i kapitlets siste del. Mine analyser av klokkeslettsuttrykkene er et tillegg til NorGrams analyse av norsk bokmål. NorGram-prosjektet er igjen en del av det internasjonale ParGram-prosjektet. I dag blir NorGrams analyse brukt innen LOGON, som er det prosjektet hvor mine analyser direkte vil bli anvendt. Disse tre prosjektene presenteres i avsnitt 2.5.

2.2 Leksikalsk-funksjonell grammatikk

I denne avhandlingen har jeg tilstrebet å gi norske klokkeslettsuttrykk en analyse tuftet på den lingvistiske teorien LFG – leksikalsk-funksjonell grammatikk. Teorien ble utviklet på slutten av 1970-årene, og pionerene i dette arbeidet var Bresnan og Kaplan (1982). Det fins en rekke bøker som gir en grundigere innføring i denne teorien, her vil jeg særlig trekke fram *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax* (Falk, 2001) og *Lexical-Functional Syntax* (Bresnan, 2001).

2.2.1 Hvorfor LFG?

LFG representerer et brudd med transformasjonsgrammatikken (TG), som fram til LFG-teoriens gjennombrudd var det dominerende syntaktiske rammeverket. TG bygger igjen på Noam Chomskys teorier rundt generativ grammatikk (Chomsky, 1957 og 1965). Om TG sier Van Valin (2001, s. 176) dette:

> TG was a constituent -structure-based theory [...] employing phrase-structure rules [...] and transformations, which are rules mapping one phrase-structure configuration into another.

TG gir to representasjoner av setningsstrukturen. Disse kalles dypstrukturen og overflatestrukturen. Dypstrukturen representeres som en hierarkisk trestruktur som viser relasjonene mellom ord og fraser innad i setningen. Frasestrukturregler bestemmer hvordan denne dypstrukturen ser ut. En setnings faktiske form slik den representeres i en muntlig eller skriftlig ytring, kalles overflatestrukturen. Avvikene mellom den dype strukturen og overflatestrukturen forklares ved hjelp av transformasjonsreglene. Den underliggende strukturen til en deklarativ og en interrogativ setning (for eksempel *Det regner i dag* og *Regner det i dag?*) er derfor grunnleggende lik, men siden transformasjonsreglene for å konstruere deklarative og interrogative setninger er forskjellige, vil de få ulik overflatestruktur.

Teorien rundt transformasjonsgrammatikken fikk stort gjennomslag, og ble den ledende syntaktiske teorien etter at den ble lansert. Også innen kognitive vitenskaper ble teorien fulgt tett:

> [M]any philosophers were attracted by the idea that deep structures might provide something very much like the 'logical form' of sentences needed for precise analysis of their role in inference. Likewise, psychologists hoped that the transformational derivations were a first approximation to the mental processes involved in the production and comprehension of utterances. (Bender et al., 2003, s. 529)

LFG-formalismen har derimot ikke noe skille mellom en dypstruktur og en overflatestruktur. Likevel er målsetningen den samme:

> LFG rejects the assumptions of transformational theory, not its goals. The basic argument for the LFG approach to syntax is simply that certain transformationalist assumptions are incompatible with the search for a theory of Universal Grammar. (Falk, 2001, s. 2)

LFGs frasestruktur følger den observerbare lineære sekvensen i en setning, og strukturen opererer følgelig ikke med tomme noder. Den sentrale ideen er at "relational concepts like 'subject' are of central importance and cannot be defined in terms of tree structures" (Bender et al., 2003, s. 538). Konseptene, i form av grammatiske relasjoner, skilles ut av setningen og representeres i 'attributt og verdi'-matriser.

2.2.2 LFGs setningsrepresentasjon

LFG har to former for setningsrepresentasjon. Disse to strukturene er *konstituent-strukturen* (constituent structure, i oppgaven referert til som c-struktur), og *funksjons-strukturen* (functional structure, forkortet f-struktur).

C-strukturen

C-strukturen følger de fleste konvensjonene til X' (uttales X-bar)-teorien. Målsetningen for denne teorien var "to constrain c-structure rules on the one hand and express cross-category generalizations on the other" (Falk, 2001, s. 34). Teorien skriver seg fra 1970 (Chomsky, *Remarks on Nominalization*), og *X'-Syntax: A Study of Phrase Structure* (Jackendoff, 1977) er det mest kjente verket om X'-teori. Analysen følger den lineære sekvensen til en ytring: en venstre-høyre-lesning for skriftlig tekst, og en temporal sekvens for muntlige ytringer. Ordene i setningen ordnes i frasekategorier som igjen gis en representasjon i et syntaktisk tre. Hvordan ordgrupper ordnes til en frase motiveres av hvilken mulighet ordene har til å interagere i forskjellige posisjoner i setningen (Falk, 2001, s. 35).

Det er fire hovedgrupper av leksikalske kategorier.[1] Dette er verb, pre- eller postposisjoner, nomen og adjektiv. Disse står som hoder i VP, PP, NP og AP. Disse leksikalske kategoriene er åpne klasser hvor nye ord kan komme til, og det er disse kategoriene som bidrar med meningsinnholdet i setningen.

(1) VP [spiste middag i går kveld]

(2) PP [i går kveld]

(3) NP [middag]

(4) AP [fri som fuglen]

2.2.3 Endosentrisitet og eksosentrisitet

Eksosentrisitet

Ikke alle konstruksjoner har en klassisk c-struktur fra hode til frasekategori. Disse konstruksjonene kalles eksosentriske (eng: *exocentric*), og introduserer den ikke-projiserte kategorien S (som står for 'sentence' eller 'small clause'). "Nonprojectivity means that S lacks a categorial head: its category

[1] Noen regner også med adverb, men forholdet mellom adverb og adjektiv er uklart (Falk, 2001, s. 34).

cannot be identified with any fixed category X°." (Bresnan, 2001, s. 110) I c-strukturen er døtrene til S flatt organisert under S:

> It can consist of a string of words, or a string of phrases with a single lexical item to serve as the functional head, or a NP (or DP) subject and predicate of any category. (Falk, 2001, s. 50)

I NorGram (se avsnitt 2.5.1) finner vi den eksosentriske kategorien S som datter til I'.[2]

Endosentrisitet

For endosentriske grammatikker går c-projeksjonen fra atom via underkategorier, og opp til toppkategorien:

(5) XP → YP, X'

(6) X' → X°, YP

Alle frasene inneholder ett ord (atomet) som er definert innenfor den samme kategorien. For verbfrasen (VP) er dette et verb og for nominalfrasen (NP) er dette nomenet. Dette ordet er ifølge X'-teorien hodet, og bare ett ord i hver frase har denne egenskapen. Dette prinsippet kalles endosentrisitet (Falk, 2001, s. 43).

X i (5) og (6) skal illustrere at alle frasekategorier har denne projeksjonen. XP er toppkategori til en hvilken som helst frase. Forgreningen denne toppkategorien dominerer står på høyre side av pilen. Toppkategorien i frasen kan skrives X". En ekspansjon av denne kategorien som ikke er atomisk, det vil si som ikke består av et ord, annoteres X', og det leksikale hodet, selve ordet, benevnes X.

(7) *Projeksjonen fra atom til toppkategori*

XP
SPEC X'
X COMP

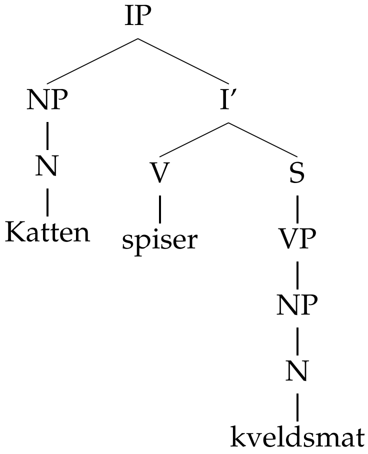
Spec og *Comp* er plasser reservert for hodets eventuelle spesifikator eller komplement. Dette er nye toppkategorier fra enten en leksikalsk eller funksjonell kategori. Hvilke projeksjoner som er mulige blir bestemt av frasestrukturreglene. Toppkategorien for hele setningen kalles IP. Bakgrunnen for dette er at det finitte verbet i setningen regnes som setningens hode. Verbformen har gitt kategorien I' navn, da I står for 'Inflection'. Ifølge konvensjonene for endosentrisitetsprinsippet skal toppkategorien da hete IP.

[2]Se for eksempel c-struktur (9).

Med de enkle frasestrukturreglene i (8), og forutsatt at vi har definert et leksikon som bestemmer ordklassene til ordene, kan vi ut fra X′-teorien gjøre følgende analyse:[3]

(8) IP → NP I′
I′ → V (S)
S → VP
VP → (V)(NP)
NP → N

(9) *Katten spiser kveldsmat*



2.2.4 Funksjonelle kategorier

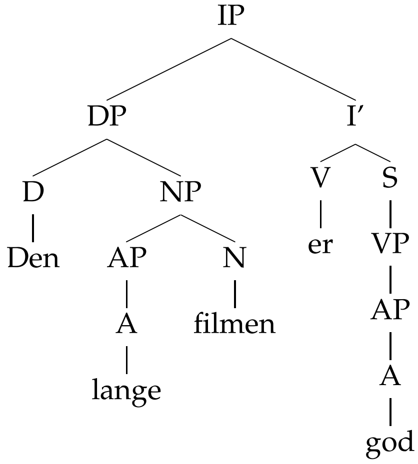
Til nå har jeg bare nevnt de leksikalske kategoriene. Men en viktig utvikling innen X′-teorien var teoretiseringen rundt lukkede klasser av grammatiske kategorier:

> These belong to a subdivision of ”functional” F° categories, which are generally closed classes of ”function words” distinguishable from the ”lexical words” which projects the levels of (lexical) categories seen above. [se eksempel (9)] (Bresnan, 2001, s. 99)

Endosentrisitetsprinsippet ble utvidet til også å gjelde disse klassene. Kategoriene bringer med seg grammatisk informasjon som tempus, bestemthet, modus og polaritet[4]. Dette er en lukket gruppe av ’funksjonsmarkører’(Bresnan, 2001, s. 99), som kan være bestemmelsesaffikser, determinativer, ordet *ikke* med flere. Bresnan (2001) opererer med de tre funksjonelle kategoriene C, I og D, som står for Complementizer, Inflection og Determiner.

[3] Parentes marker her at forekomsten er opsjonell.
[4] Hvorvidt det foreligger en negasjon i setningen.

(10) *Den lange filmen er god*



I (10) ser vi at den funksjonelle kategorien DP dominerer den leksikalske kategorien NP. *Den* bidrar med å markere bestemthet til den leksikalske kategorien. Dette ser vi ikke i c-strukturen, men må gå til f-projeksjonen.

F-strukturen

Mens c-strukturen ordner setningen i en hierarkisk trestruktur, representerer f-strukturen setningen i en matrise. Rent matematisk er f-strukturen et finitt sett av parene *attributt* og *verdi* (Bresnan, 2001, s. 47):

(11) Attributt-verdi-tabell

$$\begin{bmatrix} ATTRIBUTT_1 & verdi_1 \\ ATTRIBUTT_2 & verdi_2 \\ \ldots & \\ ATTRIBUTT_n & verdi_n \end{bmatrix}$$

Et attributt i en f-struktur kan være:

- en grammatisk funksjon, som SUBJ, OBJ, COMP og PREDLINK
- et grammatisk trekk, som NUM eller DEF
- PRED, som definerer den semantiske formen samt de andre grammatiske funksjonene (argumentene) på dette nivået

En verdi kan være:

- et symbol, for eksempel 'pl'
- en semantisk form (et uttrykk i anførselstegn, for eksempel 'katt' eller 'over<($\uparrow$ OBJ)>')

- en f-struktur
- et sett av f-strukturer

Hvert attributt har en unik verdi. Innen dette skjemaet er rekkefølgen på de forskjellige attributtene uvesentlig. I motsetning til c-strukturen, som gjenspeiler den lineære setningsoppbygningen ord for ord, kan f-strukturens representasjon reflektere flere ulike c-strukturer. Det viktige er at disse setningenes bestanddeler vil bli tolket på samme måte uavhengig av hvordan setningens fysiske struktur ser ut, fordi meningsinnholdet i disse tilfellene vil være tilnærmet likt. Setningene (12) og (13) kan gi den samme f-strukturen.[5]

(12) Bilen er blå.

(13) Blå er bilen.

$$\begin{bmatrix} \text{PRED} & \text{'være}\left\langle \left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{PREDLINK}\right)\right\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'bil'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{DEF} & + \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{'blå'} \\ \text{NUM} & \text{sg} \\ \text{DEF} & - \end{bmatrix} \end{bmatrix}$$

Det første attributtet vi støter på i f-strukturen er PRED. Alle leksikalske kategorier har en egen PRED. Som verdi til PRED står grunnformen til ordet. Verdien i seg selv er bare en merkelapp som identifiserer meningsinnholdet, og konvensjonelt brukes ordets leksikalske oppslagsform.[6]

I NorGram tar attributtet PRED på ytterste nivå i f-strukturen det finitte verbet som verdi. Dette betyr at setningens hode er det finitte verbet, en intuisjon som også reflekteres i NorGrams c-strukturer ved at V er hodet til IP (14).

(14)
```
   IP
   |
   I'
  /  \
 V    S
```
(**V** S)

[5] Med mindre grammatikken markerer hvilket ord som står som TOPIC/FOCUS i setningen.

[6] Dersom oppslagsordet i seg selv er flertydig, kan det gjøres ytterligere spesifiseringer. *Vær* som meteorologisk fenomen skilles for eksempel fra dyret *vær* ved at disse nomenene får formen 'vær-met' og 'vær-anim' som verdi for PRED i NorGram.

I (12) og (13) er det finitte verbet *er*, og verbets infinitivsform legger dermed beslag på verdi-plassen for attributtet PRED. Innen hake-parentesene følger verbets argumentstruktur (a-struktur), og disse argumentene er igjen representert som attributter på ytterste nivå i f-strukturen. Verbformen *er* gir informasjon også om tempus i setningen. Denne informasjonen vises ikke i c-struktur, men i f-strukturen representeres dette ved at attributtet TENSE får verdien 'pres'. Tempus-informasjonen hentes fra en morfologisk analyse av verbet.

2.2.5 Grammatiske funksjoner

Fra det finitte verbets a-struktur framkommer de predikatstyrte grammatiske funksjonene (*Governable grammatical functions*, se blant annet Dalrymple (2001, s. 11)). I (13) er dette SUBJ og PREDLINK.

Predikatet forutsetter at disse rollene blir fylt, ellers er setningen ugrammatisk. Modifikatorene er motstykket til de predikatstyrte grammatiske funksjonene. Modifikatorene opptrer uavhengig av argumentrestriksjonene predikatet pålegger, og det er prinsippielt sett ingen restriksjoner rundt antall modifikatorer.

Dalrymple (2001, s. 11) opererer med følgende todeling av hvilke roller som er predikatsstyrte, og hvilke som er modifikatorer:

(15) Predikatstyrte grammatiske funksjoner (PGF) og modifikatorer:

$$\underbrace{SUBJ/OBJ/XCOMP/COMP/OBJ_{\Theta}/OBL_{\Theta}}_{PGF} \quad \underbrace{ADJ/XADJ}_{Modifikatorer}$$

I min oppgave behandler jeg PREDLINK som en egen kategori av grammatiske funksjoner (noe tilsvarende gjøres i Butt et al., 1999, s. 70). Dalrymple bruker ikke funksjonen PREDLINK, og vil bruke XCOMP i dens sted.

Unikhet, kompletthet og koherens

Dersom en f-struktur skal være velformet må kravene til *unikhet, kompletthet* og *koherens* være tilfredsstilt. Disse kravene sikrer at leksikonoppslagene til ordene og frasestrukturreglene ikke disharmonerer med den faktiske setningsstrukturen.

Unikhetsprinsippet krever at hvert attributt må ha én unik verdi i f-strukturens attributt-verdi-matrise. Hvert leksikonoppslag inneholder en matrise, og hvordan disse matrisene føyer seg inn i hverandre i den helhetlige f-strukturen, bestemmes av ligningene assosiert med frasestrukturreglene. Både ordenes leksikonoppslag og annotasjonene på frasestrukturreglene legger videre restriksjoner på matrisene. Dersom en sammenføyning av alle matrisene fra leksikon til en f-struktur bryter med noen av unifikasjonsprinsippene (Falk, 2001, s. 17), er setningen ifølge grammatikken ikke velformet. Kompletthetskravet uttrykker Bresnan (2001, s. 63) slik:

> Completeness requires that every function designated by a PRED be present in the f-structure of that PRED.

Alle funksjoner som er til stede i et PREDs argument-struktur må forekomme på dette nivået i f-strukturen til PRED. For et PRED som i (16), må det foreligge et objekt (OBJ) på samme nivå.

(16) PRED 'med<(↑ OBJ)>'

Dette kravet blir tilfredsstilt i følgende f-struktur:

$$\begin{bmatrix} \text{PRED} & \text{'med}\left\langle \left(\uparrow \text{OBJ} \right) \right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'brød'} \end{bmatrix} \end{bmatrix}$$

I (17) brytes kompletthetsprinsippet fordi setningen ikke inneholder et objekt, noe verbets a-struktur (18) krever. Setningen er ugrammatisk fordi verbet krever en forekomst av en grammatisk funksjon som ikke foreligger.

(17) *Kvinnen lagde.

(18) 'lage<(↑ SUBJ)(↑ OBJ)>'

Videre lanserer Bresnan kravet om koherens:

> Coherence requires that every argument function in an f-structure be designated by a PRED. (ibid.)

Dette betyr at ingen (predikatstyrte) grammatiske funksjoner kan opptre innenfor en f-struktur med mindre de blir etterspurt i et PRED. Ut ifra denne restriksjonen vil (19) bli forkastet fordi verbets a-struktur (20) her bare krever et subjekt og ingen annen grammatisk funksjon:

(19) Mannen sover bilen.

(20) 'sove<(↑ SUBJ)>'

De eneste funksjonene som kan opptre uten å bli henvist av et PRED, er modifikatorene. Dette forutsetter imidlertid at frasestrukturreglene tillater forekomster av en modifikator i denne posisjonen. Innen f-strukturen benevnes modifikatorene med attributtet 'ADJ', som tar et sett av strukturer som verdi. ADJ er en forkortelse for 'adjunct'.

2.2.6 Frasestrukturreglene

Projeksjonen av f-strukturen blir uttrykt gjennom annotasjoner på frasestrukturreglene. Dersom vi går tilbake til de forenklede reglene i (9), mangler denne informasjonen. Vi kan se på hvordan f-strukturen for (9) ser ut:

(21) Katten spiser kveldsmat.

$$(21)\ \begin{bmatrix} \text{PRED} & \text{'spise}\left\langle \left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\ \text{TENSE} & \text{pres} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'katt'} \\ \text{DEF} & + \\ \text{NUM} & \text{sg} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kveldsmat'} \\ \text{DEF} & - \\ \text{NUM} & + \end{bmatrix} \end{bmatrix}$$

At *spiser*s PRED legger beslag på det ytterste nivået i f-struktur betyr at *spiser* er hode i IP. NP, og det leksikalske hodet *Katten* er en spesifikator til IP, og fyller i denne konteksten den grammatiske funksjonen SUBJ. Dette uttrykkes i reglene på følgende måte:

$$\begin{array}{ll} \text{IP} \rightarrow & \text{NP } (\uparrow \text{SUBJ}) = \downarrow \\ & \text{I}' \uparrow=\downarrow \end{array}$$

$$\text{NP} \rightarrow \quad \text{N} \uparrow=\downarrow$$

En datter overfører sin f-struktur til modernoden dersom annoteringen er ↑=↓. Dette leses "mine trekk skal unifiseres med min mors". Hodet i en projeksjon har alltid denne annoteringen. IPs spesifikator NP er imidlertid annotert slik:

NP (↑ SUBJ) = ↓

Dermed blir ikke NPs f-struktur unifisert med NPs modernode, som i dette tilfellet er IP. SUBJ introduseres som ny attributt i modernodens f-struktur, og SUBJs verdi er f-strukturen til NP. Siden NPs datternode N er hode i NP, vil Ns f-struktur fylle verdiplassen til SUBJ. Ns f-struktur hentes fra leksikon. Leksikonoppslaget for *katten* ser i dette tilfellet slik ut:

$$\begin{array}{lll} \text{katten} & \text{N} & (\uparrow \text{PRED})=\text{'katt'} \\ & & (\uparrow \text{DEF})=+ \\ & & (\uparrow \text{NUM})=\text{sg} \end{array}$$

Utsnittet av informasjonen dette leksikonoppslaget representerer, gjenfinnes som f-strukturen under SUBJ i (21). Resten av frasestrukturreglene og leksikonet som er nødvendig for å generere f-strukturen i (21) følger under:

Frasestrukturregler:

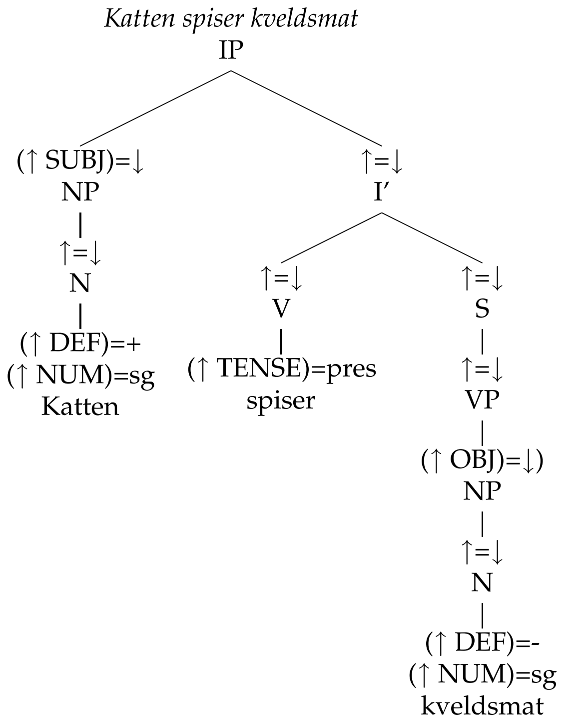
I′ →	V ↑=↓ S ↑=↓
S →	VP ↑=↓
VP →	(V ↑=↓) NP (↑ OBJ)=↓
NP →	N ↑=↓

Leksikon-oppslaget

spiser	V	(↑ PRED)='spise<(↑ SUBJ)(↑ OBJ)>' (↑ TENSE)=pres
kveldsmat	N	(↑ PRED)='kveldsmat' (↑ DEF)=- (↑ NUM)=sg

C-struktur (22) viser annoteringene fra frasestrukturreglene slik de framkommer i f-strukturen (21).

(22)



2.3 LFG innen XLE

Selve grammatikken som jeg har utarbeidet for klokkeslettsuttrykkene, er implementert på XLE-plattformen. XLE (Xerox Linguistic Environment) er:

> a computational environment that assists in writing and debugging Lexical Functional Grammars (LFGs) (Kaplan & Bresnan, 1982). It provides linguists with a facility for writing syntactic rules and lexical entries, and for testing and editing them. It also provides an interface with finite-state morphological analyzers. (Crouch et al., 2002-2005, URL)

Kompleksiteten til naturlige språk er det største hinderet for effektiv parsing innen maskinoversettelse (MO), noe som også gjenspeiler seg i XLE:

> [parsing and generation] with LFGs efficiently [is] difficult because the LFG formalism, like most unification-based grammar formalisms, is NP complete. This means that in the worst case the time that it takes to parse or generate with an LFG can be exponential in the length of the input. However, natural languages are mostly context-free equivalent, and one should be able to parse them in mostly cubic time. XLE is designed to automatically take advantage of context-freeness in the grammar of a

> natural language so that it typically parses in cubic time and generates in linear time. (Maxwell, 2002-2005, URL)

XLE tillater at koden skrives nært opp til en standard LFG-notasjon, og ved inkorporering av annoterte frasestrukturregler, et leksikon og en morfologisk analysator, vil programmet finne alle mulige analyser av en setning eller en frase. En implementering av en klassisk LFG-grammatikk vil gi et utmatningsresultat i form av:

- konstituent-strukturer (c-strukturer).

- kartet (the chart). Dette inneholder alle fullstendige eller ufullstendige deler av setningen som grammatikken tillater

- morfologien alle mulige morfologiske analyser av et leksikalsk atom.

- funksjons-strukturer (f-strukturer), og en visning av inkonsistente, ufullstendige eller inkoherente f-strukturer.

For LOGON vil det også genereres en semantisk representasjon (MRS) for analysen.

Siden utdraget fra grammatikken jeg har skrevet ligger vedlagt (tillegg B-D) kun som XLE-notasjon, vil jeg gi en rask innføring i hvordan XLE-koden skal leses. En grundigere manual for dette er tilgjengelig på http://www2.parc.com/istl/groups/nltt/xle/doc/walkthrough.html.

2.3.1 XLE-notasjonen

XLE-plattformen krever minimum tre brukerdefinerte komponenter for å fungere. Dette er leksikonet, frasestrukturreglene og en automat som analyserer morfologien.[7]

Frasestrukturreglene

XLE-notasjonen er lagt så nært opp til konvensjonell LFG-notasjon som mulig, men en del mer tegnsetting er lagt til for å klart definere hvor de forskjellige aspektene i reglene hører hjemme. Fra min grammatikk kan vi se på regelen som anvendes for et uttrykk som *klokka ti på halv ni* (23).

[7] Morfologi-automaten kan sløyfes dersom et fullformsleksikon benyttes.

(23) PP-TIME-regelen

```
PP-TIME -->
          (N-CLOCK: (^ CLOCK-OCCURRENCE)=c +_)
          TIME-MOD: (^ SPEC)=!
          @MRS-MINUTEMEASURE;
          P'-TIME: ^=!
                  (^ PTYPE TEMPFINALE)=c +.
```

Pilene fra konvensjonell LFG-notasjon er byttet ut med standard ascii-tegn:

$\rightarrow$ = `-->`
$\uparrow$ = `^`
$\downarrow$ = `!`

Reglene er tastet inn lineært. Modernoden står til venstre, og ekspansjonen på høyre side. Ekspansjonens *skjema* etterfølger ekspansjonskategorien. Skjemaet er informasjonen om hvilken grammatiske funksjon (for eksempel SUBJ(ect) eller ADJ(unct)) kategorien fyller, og hvilke restriksjoner som pålegges for at ekspansjonen (og dermed regelen) vil være grammatisk. Skjemaet kommer alltid etter et kolon, og alt som etterfølger kolonet leses som kategoriens skjema, helt fram til kolonet "lukkes", enten ved semikolon, parentes eller punktum.

Den første ekspansjonen i regel (23) er (N-CLOCK: (ˆ CLOCK-OCCURRENCE)=c +_). Ekspansjonen står i parentes, dette betyr at forekomsten er valgfri. Den grammatiske funksjonen er ikke definert. Dersom ingen funksjon defineres, leses det automatisk som ˆ=!, altså at ekspansjonens f-struktur unifiseres med moderkategorien på venstre side i regelen. Det eneste som står i N-CLOCKs skjema er restriksjonen (ˆ CLOCK-OCCURRENCE)=c +_. Dette leses som:

> Min mors f-struktur må inneholde følgende f-struktur:
>
> (24) [CLOCK-OCCURRENCE +_]

Denne restriksjonen begrenser hvilket nomen som kan velges fra leksikon. Bare bestemt form singularis av leksemet *klokke* tilfredsstiller dette kravet. At verdien +_ slutter med _(understrek), betyr at det er et instansiert symbol. Verdien kan da instansieres én, og bare én gang i f-strukturen:

> If instantiated symbols are involved, the truth value of a Boolean formula can further discriminate between the appearance of exactly one defining equation as opposed to more than one. (Crouch et al., 2002-2005, URL)

Tilbake i (23) ser vi at den eneste ekspansjonen på høyre side som definerer en grammatisk funksjon er TIME-MOD: (^ SPEC)=!;. Dette leses som: ”min f-struktur er verdi til attributtet SPEC i min mors f-struktur.”

Etter den siste ekspansjonens skjema avsluttes regelen med punktum. Hele ekspansjonen til kategorien på venstre side står mellom --> og punktum (.). En utførlig gjennomgang av XLEs notasjon er tilgjengelig på XLEs hjemmesider:
http://www2.parc.com/istl/groups/nltt/xle/doc/notations.html.

Leksikonet

(25) Leksikonoppslaget for *til*

```
til      P * @(SEMPREP til to)
               (^ PTYPE-CLOCK TEMP)=+
               (^ PTYPE-CLOCK SUBTEMP)=+.
```

Leksikonet i XLE definerer f-strukturen til lemmaet, og angir eventuelle restriksjoner ordet legger på setningens øvrige f-struktur. (25) viser leksikonoppslaget for *til*. Ordets kategori P følger umiddelbart etter ordets morfologiske form og definerer ordet som preposisjon. * markerer at ingen morfologisk analyse skal kobles inn, noe som ikke er nødvendig siden det ikke eksisterer bøyde former av *til*.[8] @(SEMPREP til to) kaller templatet @SEMPREP. Dette templatet inneholder en forhåndsdefinisjon av hvilke trekk som skal legges til for denne typen ord. Jeg gjennomgår bruken av templater i neste avsnitt.

Templatene

Bruken av templat er egentlig en snarvei som ikke er nødvendig for å anvende XLE-plattformen. Et templat inneholder bare en forhåndsdefinering av hvilke trekk eller restriksjoner som skal inn i f-strukturen. I en kompleks grammatikk blir imidlertid bruken av templater uunnværlig, siden et enkelt leksikonoppslag ville blitt svært plasskrevende og vanskelig å lese dersom ikke generelle templater ble brukt. I (25) kalles templatet @SEMPREP,

[8] Se avsnittet om morfologi i XLE, side 22.

og dette er et templat som anvendes av preposisjoner generelt. Templatet innfører de egenskapene preposisjonene har til felles i leksikonet. @SEMPREP ser slik ut i XLE-koden:

(26) @SEMPREP-templatet

```
SEMPREP (P S) =
          (^ PRED)='P<(^ OBJ)>'
          (^ PTYPE)=sem
          (^ PFORM)=P
          (s::M* relation)=S
          "</mrs>"%ARG2=m::(^ OBJ)
          @(MRS-SEMPREP P) "<mrs/>".
```

Her innføres trekkene `(^ PTYPE)=sem` og `(^ PFORM)=P` i leksikon. I tillegg defineres det at preposisjonen tar et objekt som argument. Dette markeres ved `(^ PRED)='P<(^ OBJ)>'`, hvor 'P' her er en variabel for preposisjonens lemma. De to siste linjene har konsekvenser for MRS-oppbygningen til preposisjonsfrasen. Den siste linjen kaller et nytt templat, @MRS-SEMPREP, som videre sørger for oppbygningen av MRSen.

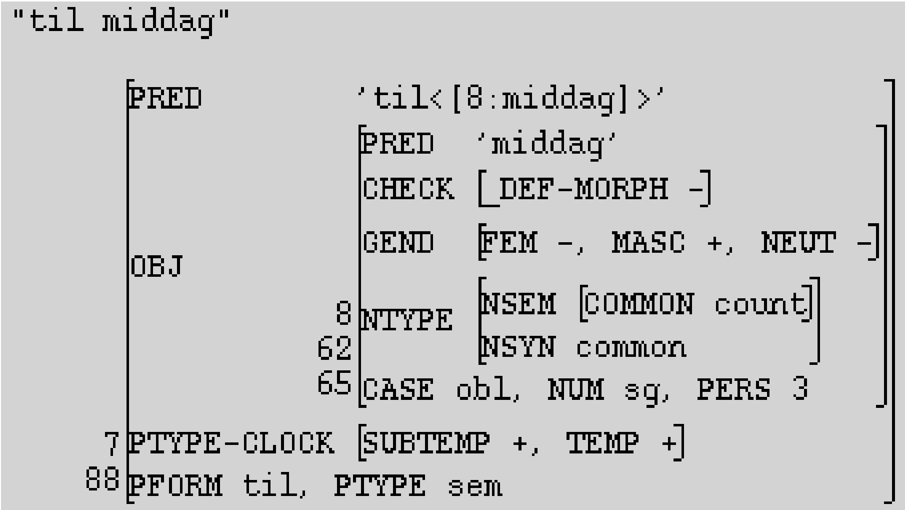
Figur 2.1 er display-utmatingen til XLE for f-strukturen til preposisjonsuttrykket *til middag*. Senere i oppgaven kommer jeg til å forenkle f-strukturene ved å utelate informasjon om nomenets bestemthet, numerus og genus og verbets tempus. Den eneste informasjonen jeg kommer til å vise, er de trekkene som er relevante for klokkeslettsuttrykket. Dette for å spare plass og gjøre f-strukturen mer oversiktlig.

Morfologi

For å slippe å legge inn samtlige former av et ord i leksikon kan det knyttes en morfologisk analysator til XLE. Denne analysatoren kan da kjenne igjen bøyde former av ordet, og bryter det ned til en grunnform samt en sekvens av abstrakte suffikser som annoterer bøyningskategoriene (Dyvik, 2003, s. 1). For ordet *studentene* gis følgende utmating fra NorGrams analysator, som også anvendes av LOGON:

student "+Noun" "+Masc" "+Def" "+Pl"

For videre informasjon om XLE henvises til XLE-dokumentasjonen (Crouch et al., 2002-2005).



Figur 2.1: F-struktur-utmatingen for *til middag* i XLE.

2.4 MRS

2.4.1 Minimal Recursion Semantics

Innen LOGON er MRSen selve *lingua franca* i applikasjonen. Det er denne representasjonen som oversettes fra norsk til engelsk, og denne overføringsdelen er frikoblet fra de andre komponentene (analysen og oppbygningen, se avsnitt 2.5.3). MRS er ikke i seg selv en lingvistisk teori, det er en representasjonsmåte, "a meta-level language for describing semantic structures in some underlying object language". (Copestake et al., 2003, s. 2) Konvensjonelle semantiske representasjoner av språk er problematiske for parsere fordi de krever mer informasjon enn en vanlig komputasjonell setningsanalyse kan framskaffe:

> In particular, they require that quantifier scope be specified. However, not only is it very difficult to resolve quantifier scope ambiguities, it is also often unnecessary in applications such as MT, since the resolution of quantifier scope usually has no effect on the translation. (Copestake et al., 2003, s. 2)

Prinsippet til MRS er å underrepresentere *scope*[9] dersom dette er redundant informasjon, men samtidig ha verktøy som er i stand til å oppdrive informasjonen dersom det likevel er nødvendig. Mellom beslektede språk er kvantorens rekkevidde ofte overflødig informasjon for utfallet av en oversettelse, siden flertydigheten i kildespråket gjerne lar seg direkte oversette

[9]Rekkevidden til en kvantor, fra nå *rekkevidde*.

til målspråket med det resultat at den samme flertydigheten bevares. Imidlertid kan små forskjeller i syntaksen til kilde- og målspråket gjøre at kvantorrekkevidden likevel blir avgjørende. I uttrykkene:

(27) Alle hvite hester er gamle.

og

(28) Alle gamle hester er hvite

vil den logiske formen være identisk dersom rekkevidden utelates:

(29) (alle(x), hest(x), gammel(x), hvit(x))

På grunn av tvetydigheten dette medfører, er (29) en uakseptabel representasjon av (27) og (28). Representasjonen for setning (27) må inneholde informasjon om kvantoren *alle*s rekkevidde, dens restriksjon og hvilket argument den binder. Dette kravet oppfylles i (30):

(30) (alle(x, hvit(x) $\wedge$ hest(x),gammel(x)))

I (30) tar kvantoren tre argumenter. Den binder argumentet x, restriksjonen *hvit(x)*, og rekkevidden $\wedge$ *hest(x),gammel(x)*. (30) kan bare representere setning (27). Representasjonen for (28) ville blitt slik (31):

(31) (alle(x, gammel(x) $\wedge$ hest(x),hvit(x)))

2.4.2 Elementærpredikasjonene

Basiselementene i MRS er EPene, elementærpredikasjonene (elementary predications). Copestake et al. (2003, s. 10-11) har denne definisjonen av EPen:

> An elementary predication contains exactly four components:
>
> - a handle which is the label of the EP
> - a relation
> - a list of zero or more ordinary variable arguments of the relation
> - a list of zero or more handles corresponding to scopal arguments of the relation
>
> This is written as handle:relation $(arg_1...arg_n, sc\text{-}arg_1...sc\text{-}arg_m)$.

En *EP-konjunksjon* defineres videre som: "a bag of EPs that have the same label." Dette gjelder for eksempel i nominalfraser hvor et adjektiv modifiserer nomenet. Både EPen for adjektivet og EPen for nomenet vil ha samme *etikett* (label) (se (33)).

2.4.3 MRS-representasjonen

En velformet MRS er en bag av EPer, samt tre andre komponenter:

> 1. The *top handle* of the MRS corresponds to a handle which will label the highest EP conjunction in all scope-resolved MRSs which can be derived from this MRS.
> 2. *Handle constraints* or *hcons* contains a (possibly empty) bag of constraints on the outscopes partial order. (ibid.)
> 3. The local top, or ltop, which will be the topmost label in an MRS. (Copestake et al., 2003, s. 17)

MRSen defineres som et firedelt sett <GT, LT,R, C> der GT tilsvarer topetiketten (*top label*), LT er den lokale toppen, R tilsvarer settet av EPer, og C er settet av håndtaksrestriksjoner.

2.4.4 Håndtaksrestriksjonene

Håndtakrestriksjonene er en type restriksjon som utøves på håndtakene i MRS. De brukes for å definere rekkevidden til kvantorene, og kalles en *qeq-restriksjon* (eller =q), som står for *equality modulo quantifiers*.

> A qeq constraint always relates a handle in an argument position to a label. The intuition is that if a handle argument h is qeq some label l, either that argument must be directly filled by l (i.e., h = l), or else one or more quantifiers 'float in' between h and l. In the latter case, the label of a quantifier fills the argument position and the body argument of that quantifier is filled either by l, or by the label of another quantifier, which in turn must have l directly or indirectly in its body. (Copestake et al., 2003, s. 14)

Et eksempel på en MRS-representasjon

En velformet MRS-representasjon av uttrykket *every dog probably chases some white cat* kan representeres slik:

(32) $h0$, $h5$,{ h1: every(x, $h2$, $h3$), $h4$:dog(x), $h5$: probably($h6$), $h7$:chase(x,y), $h8$:some(y,$h9$,$h10$), $h11$:white(y),$h11$:cat(y)},
{$h0$ =q $h5$, $h2$ =q $h4$, $h6$ =q $h7$, $h9$ =q $h11$}

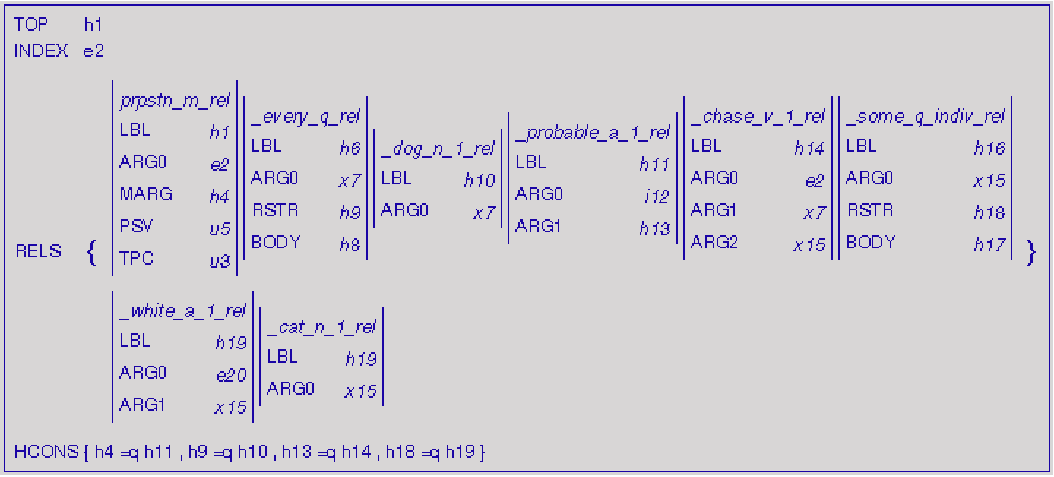
Her ser vi at nomenene er representert med EPer med et håndtak, et relasjonsnavn (navnet på nomenet) og en variabel. Verbet *chase* bruker variablene til å definere de grammatiske funksjonene de to nomenene har i setningen. Adjektivet *white* markerer at det korresponderer med nomenet *cat* både ved å ta samme variabel, og ved å inneha samme håndtak (33). Disse utgjør en EP-konjunksjon.

(33) $h11$:white(y), $h11$:cat(y)

Kvantorene benytter en liste av håndtak som markør for kvantorens rekkevidde (34).

(34) $h1$: every(x, $h2$, $h3$)

I listen av MRSens håndtaksrestriksjoner finner vi '$h2$ =q $h4$'. $h2$ knyttes her til etiketten $h4$, noe som markerer at *every* viser til nomenet *dog*. MRSer kan representeres som en trekkstruktur. Online-demonstratoren LinGO ERGs (CSLI LinGO Lab, URL) MRS-representasjon av (32) er gjengitt i figur 2.2.



Figur 2.2: LinGO ERGs MRS-representasjon av *Every dog probably chases some white cat*

2.5 LOGON, NorGram og ParGram

Min analyse av klokkeslettsuttrykk skal danne grunnlaget for LOGONs analyse av disse uttrykkene. LOGON er et akronym for "Leksikon, Ordsemantikk, Grammatikk og Oversettelse for Norsk". Dette er et forskningsprosjekt som arbeider med utviklingen av et program for maskinoversettelse av norske setninger til engelsk, og prosjektet er støttet av Norges forskningsråd. Initiativet er et samarbeid mellom universitetene i Bergen, Oslo og Trondheim, og knytter til seg forskere og samarbeidspartnere også eksternt fra disse institusjonene.

2.5.1 NorGram

LOGONs nåværende syntaktiske analyse bygger på resultatet av NorGram-prosjektet. NorGrams (Norsk komputasjonell grammatikk) formål var:

> [Å] utvikle en gjenbrukbar komputasjonell grammatikk for norsk bokmål og nynorsk med bred empirisk dekning og sunn teoretisk basis. (Dyvik, 1999, s. 1)

Bokmålsgrammatikken baserer seg på rammeverket leksikalsk-funksjonell grammatikk (LFG), ble skrevet i XLE (se avsnitt om LFG og XLE, 2.3) og har et leksikon bestående av 80 000 lemma. NorGram inngikk i et samarbeid med ParGram ved Xerox Palo Alto Research Center (PARC).

2.5.2 ParGram

ParGram er et samarbeidsprosjekt mellom forskere fra Palo Alto Research Center, Universitetet i Stuttgart, Fuji Xerox i Japan, universitetene i Essex og Oxford i England, samt Universitetet i Bergen.

> The ParGram project aims to test the LFG formalism for its universality and coverage limitations and to see how far parallelism can be maintained across languages. Where possible, the analyses produced by the grammars for similar constructions in each language are parallel. This has the computational advantage that the grammars can be used in similar applications and that machine translation (Frank, 1999) can be simplified. (Butt et al., 2002, s. 1)

En av målsetningene til ParGram-prosjektet er å teste hvor langt det er mulig å bevare en parallell representasjon av likeartede fenomener innen forskjellige språk. Også innen ParGram anvendes XLE som plattform. Tross de svært forskjellige c-strukturene språkene innen ParGram produserer, er målsetningen at setningsanalysens f-struktur skal være mest mulig lik for alle språkene. Ulikartede f-struktur-representasjoner av en ytring kan bare rettferdiggjøres dersom det foreligger sterke lingvistiske grunner til å gi dem forskjellige representasjoner. I prosjektrapporten konkluderes det med at:

> [t]he ParGram project has shown that it is possible to use a single grammar development platform and a unified methodology of grammar writing to develop large-scale grammars for typologically different languages. (Butt et al., 2002, s. 7)

2.5.3 LOGON

LOGON representerer den største satsingen på maskinoversettelse (MO) i Norge i dag. Oversettelsen går fra norsk (kildespråk) til engelsk (målspråk). LOGONs syntaktiske analyse baserer seg som nevnt på NorGram-prosjektet. NorGrams grammatiske koding i XLE ga opprinnelig en utmating i form av c- og f-strukturer. Denne grammatikken er i LOGON utvidet til også å generere MRSer (se avsnitt om MRS, 2.4), som benyttes av overføringsmodulen til LOGON.

Systemoppbygningen

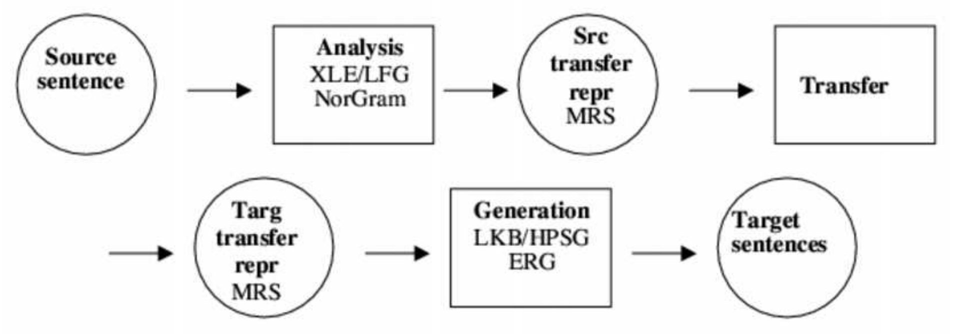
Verbmobil (Wahlster, 2000) var det siste store MO-prosjektet som baserte seg på dyp syntaktisk analyse med et solid lingvistisk teorigrunnlag. Trenden det siste tiåret går mot mer statistisk motiverte MO-strategier (Oepen et al., 2004, s. 2). LOGON prøver å forene disse to strategiene. For kildespråket (norsk) vil setningene få fullstendige LFG-analyser. Dersom parseren ikke er i stand til å finne en overordnet løsning for hele setningen, vil deler av setningen få hver sin analyse. Målsetningen er at disse delene skal kunne oversettes bit for bit til engelsk.

Selv relativt enkle norske setninger kan gi svært mange LFG-analyser. For eksempel gir setningen *Iglfjellområdet er også et flott område for skiturer* tolv forskjellige analyser. Her skal det kobles inn en statistisk metode for å beregne sannsynligheten for hvilke oversettelse som er riktig. Parseren skal ha en stokastisk rangeringsmekanisme for å sortere ut de mest sannsynlige analysene, men dette arbeidet er per dags dato ikke avsluttet.(LOGON Consortium, 2004)

LFG-analysene gir som i NorGram en c- og f-struktur-analyse av setningene. I tillegg dannes en MRS-representasjon av setningen. MRSen utgjør grunnlaget for oversettelse til en engelsk MRS-representasjon. Selve overføringsmodulen i LOGON baserer seg altså utelukkende på MRS. Med dette er LOGON "to our best knowledge [...] the first system to implement end-to-end MRS-based translation" (Oepen et al., 2004, s. 3). De engelske setningene bygges opp fra den engelske MRS-utmatingen ved hjelp av en HPSG-basert plattform.

I figur 2.3 illustreres de sentrale komponentene i overføringen fra norsk til engelsk. Den norske kilde-setningen analyseres av XLE-plattformen. XLE gir en MRS-utmating, som er den semantiske representasjonen av setningen. Denne semantiske representasjonen blir oversatt til engelsk i transferdelen. Oppbygningen til en engelsk setning går motsatt vei. Den semantiske representasjonen er innmating for LKB-generatoren, som baserer seg på HPSG-grammatikk. Fra denne generatoren bygges målsetningen opp.

Den engelske HPSG-grammatikken bygger på det eksisterende *LinGO English Resource Grammar* (Flickinger, 2000) og *LKB-generator* (Carroll et al.,



Figur 2.3: De sentrale komponentene i LOGON

1999). LinGO ERG ble opprinnelig bygd for Verbmobil, og er blitt tilpasset LOGONs materialdomene. Det arbeides fortsatt med å optimere oppbygningen av målspråket, et av de største problemene er at hver MRS i gjennomsnitt genererer 30 ulike mulige setninger i målspråket. Dette tallet må begrenses med statistiske metoder (Oepen et al., 2004, s. 9).

Materialet og den videre strategien

Den nåværende LOGON-prototypen baserer seg på et tekstmessig avgrenset domene. Domenet som er valgt er turistinformasjon og turistrelaterte publikasjoner. Materialtilfanget holdes nede fordi:

> [w]e are building a proof of concept demonstrator and not a real application:
>
> - Around 5000 lexemes from Norwegian should suffice. We expect 3000 sentences with an average length of 15 words to be suitable.
> - We shall get the text translated. We shall have two (three?) independent translations.
> - All translator are instructed to provide a direct, but idiomatic, translation without paraphrases. One of the translators is specially instructed to provide a translation as close as possible to the source text.(LOGON Consortium, 2004, s. 5)

LOGON tar altså ikke sikte på å nå en dekningsgrad som kan anvendes i en kommersiell applikasjon. Det er selve oversettelsesmetoden LOGON bygger på som skal etterprøves. Dekningsgraden LOGON når, vil være en pekepinn på hvor vellykket LOGONs strategi har vært, og vil danne grunnlag for videre forskning rundt MO-strategier knyttet til dyp syntaktisk analyse, semantisk overføring og statistisk utvelgelse av korrekte representasjoner.

2.6 Oppsummering

I dette kapitlet har jeg gitt en innføring i den lingvistiske teorien som ligger til grunn for min analyse av klokkeslettsuttrykk. Dette er LFG – leksikalskfunksjonell grammatikk. Jeg har også gitt en kort presentasjon av hvordan en grammatikk blir implementert på XLE-plattformen. Selve XLE-plattformen er for kompleks til at jeg har kunnet gi en detaljert beskrivelse av hvordan den fungerer, og min presentasjon er ment å være et hjelpemiddel for å kunne lese XLE-koden som klokkeslettsanalysen er skrevet i (vedlagt i tillegg B-D). XLE-dokumentasjonen (Crouch et al., 2002-2005) gir en god innføring i XLE-plattformen. *A Grammar Writer's Cookbook* (Butt et al., 1999) gir en videre introduksjon i en praktisk anvendelse av XLE for å implementere LFG-grammatikker.

En MRS-representasjon for klokkeslettsuttrykkene vil bli brukt i overføringsdelen mellom norsk og engelsk (se avsnitt 2.5.3). Jeg har derfor gjennomgått de grunnleggende prinsippene til MRS. For en mer utfyllende introduksjon til MRS anbefaler jeg *Minimal Recursion Semantics: An introduction* (Copestake et al., 2003).

De tre prosjektene som har vært relevante for denne oppgaven, blir presentert i siste del av kapitlet, fra avsnitt 2.5. Disse prosjektene har lagt premissene for hvordan min analyse skal utformes. I det neste kapittelet viser jeg til grammatiske aspekter som motiverer den analysen jeg presenterer for klokkeslettsuttrykkene. Her drar jeg inn litteratur som omhandler norsk grammatikk generelt, og kilder som går mer konkret på adverbial og tidsuttrykk.

Kapittel 3

Klokkeslettsuttrykkets syntaks

3.1 Innledning

I dette kapitlet skal jeg belyse hvilke syntaktiske restriksjoner klokkeslettsuttrykket pålegges fra resten av setningen. I første del av kapitlet, avsnitt 3.2, blir klokkeslettsuttrykket sammenlignet med andre adverbial, og det blir plassert innenfor *Norsk Referansegrammatikk*s (Faarlund et al., 1997) klassifisering av norske adverbial generelt. Videre belyser jeg syntaktiske egenskaper hos klokkeslettsuttrykkene som skiller dem fra andre tidsadverbial.

I andre del, avsnitt 3.3, ser jeg nærmere på de preposisjonsfrasene som utgjør et klokkeslettsuttrykk. Jeg går gjennom hvilke restriksjoner de forskjellige preposisjonene som kan stå som hode i et klokkeslettsuttrykk pålegger resten av setningen. Jeg ser både på preposisjonsfrasen PP-TIME (se avsnitt 4.3), som er spesiallaget for å analysere klokkeslettsuttrykk, og generelle preposisjonsfraser.

3.2 Tidsadverbialet

3.2.1 Hva er et tidsadverbial?

Tidsadverbialet er en undergruppe av adverbialene. Termen *tidsadverbial* brukes blant annet i *Norsk referansegrammatikk* (Faarlund et al., 1997) og *Verbfrasens struktur i svenskan* (Andersson, 1977). Adverbialgruppen lar seg lettest avgrense semantisk. Om den semantiske definisjonen av tidsadverbialene skriver Andersson (1977, s. 43-44):

> Det tycks vara möjligt att avgränsa tidsadverbialen semantiskt sett och formulera allmänna regler för deras användning. Emellertid utgör tidsadverbialen ingen koherent kategori. Den omständighet som de specificerar inbegriper alltid tid, men de kan modifisera verbfrasen i flera olika avseenden: de kan avgränsa

> en plats på tidslinjen, d.v.s. ge en tidfästning, de kan ange duration eller frekvens, eller handlingens ordningsnummer i en serie av likartade handlingar.

Adverbialene viser enten til et konkret tidsrom eller tidspunkt, eller til relativ tid[1]. Andersson (1977, s. 68) diskuterer dette, og bruker termene *deiktisk* og *ikke-deiktisk*[2] tidsreferanse. En ikke-deiktisk tidsreferanse viser til en konkret tid, og er uavhengig av utsagnstidspunktet eller det semantiske innholdet i setningen. Dette kan være tidsadverbialer som *9. april 1940*. Om de deiktiske adverbialene sier Andersson (1977, s. 74):

> De deiktiska adverbialen erhåller en referens först då de relateras till en deixispunkt, d.v.s. en prominent tidpunkt i den situationella eller språkliga kontexten.

Et enkelt klokkeslett vil komme inn under denne betegnelsen. Når jeg bruker uttrykket *tidspunkt* i denne oppgaven, legger jeg en definisjon fra Vannebo (1979, s. 77) til grunn:

> Et 'tidspunkt' vil kunne betraktes som en abstraksjon for angivelsen av ei mengde samtidige handlinger. Et slikt tidspunkt vil bl.a. kunne representeres i form av et klokkeslett (klokka åtte), av kalenderiske bestemmelser (i 1648, ved vårfrumess, i neste uke), av kosmiske begivenheter (ved neste nymåne, før frostnettene, ved vintersolverv) og av historiske begivenheter (Kristi fødsel, Romas grunnleggelse, OL i München).

Under det Vannebo definerer som historiske begivenheter kan i prinsippet alle nomen passe inn siden de alle i visse kontekster kan referere til hendelser. Dette blir diskutert blant annet i avsnitt 4.2.2. Selv om et klokkeslettsuttrykk tidfester klokkeslettet nøyaktig, vil ikke uttrykket i seg selv kunne tidfeste konkret hvilket punkt på tidslinjen tidsadverbialet refererer til. Den konkrete dagen klokkeslettet sikter mot vil ligge skjult i setningens kontekst eller ikke være uttrykt.

Et klokkeslett kan i seg selv være et tidsadverbial, eller det kan være del av et tidsadverbial der en preposisjon er hode. Andersson (1977, s. 46) deler tidsadverbialene opp i undergrupper og opererer med begrepet *tidpunktsadverbial* for tidsadverbial som svarer på spørsmålet 'når?'. Et enkelt klokkeslett der ingen preposisjoner tar hele klokkeslettet [3] som objekt, vil falle inn under denne kategorien. Men dersom klokkeslettet er en utfylling

[1] Faarlund et al. (1997, s. 804) definerer relativ tid som "eit tidspunkt i høve til eit anna tidspunkt".

[2] Orginalt *deiktiska* og *icke-deiktiska*.

[3] Med klokkeslett menes her klokkeslett uttrykt med kategoriene HOURP eller PP-TIME, se kapittel 4.

til en preposisjonsfrase, vil adverbialets egenskaper bli bestemt av preposisjonen. I setningen *Jeg har arbeidet siden halv tre*, vil tidsadverbialet *siden halv tre* være en durasjon. En durasjon beskriver en handling som strekker seg over tid og kan besvare spørsmål med 'hvor lenge?', i stedet for 'når?'.

Hvilke restriksjoner preposisjonene legger på klokkeslettsuttrykket og resten av setningen, blir behandlet i avsnitt 3.3. I dette avsnittet vil jeg bare behandle klokkeslettsuttrykk uten preposisjoner (1), eller klokkeslettsuttrykk med preposisjonene *over* eller *på* fra kategorien PP-TIME, som i (2).[4]

(1) Jeg kommer hjem **halv** tre.

(2) Middagen er klar klokka ti **på** halv ni

3.2.2 Tre adverbialtyper

Faarlund et al. (1997, s. 773) deler norske adverbial inn i tre grupper:

1. Bundne adverbial (Adverbial som står som utfylling til verbet.)

2. Frie adverbial (Adverbial som er adledd til verbet, men ikke utfylling.)

3. Setningsadverbial (Adverbial som står utenfor predikatet.)

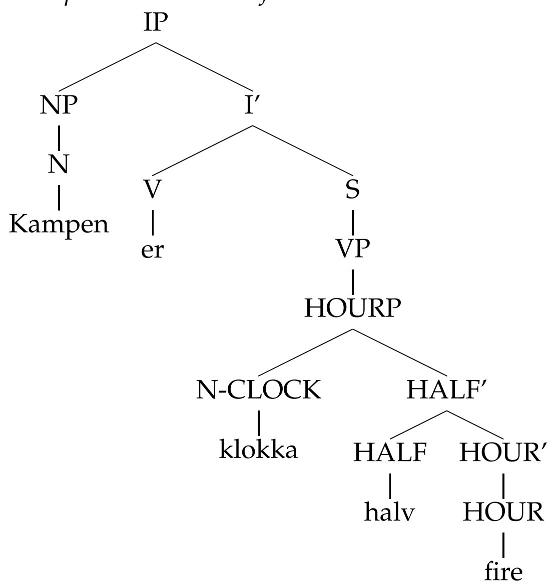
Bare de to første kategoriene er relevante for klokkeslettsuttrykk.

Bundne adverbial

I den første kategorien betegner bundne adverbial de adverbialene som er utfyllinger til verbet. Disse adverbialene er ifølge Faarlund et al. (1997, s. 774) "avhengige av valensen eller verdien åt verbet." Innenfor LFG-formalismen vil disse adverbene står innenfor argumentstrukturen til predikatet i setningen, og i f-struktur vil de stå som PREDLINK. I min grammatikk betyr dette i praksis at verbet er *være*. Alle tidspunktsadverbial vil her kunne stå som predikatsutfylling. Setning (3) og (4) gir to eksempler på klokkeslettsuttrykk som står som bundet adverbial i setningen.

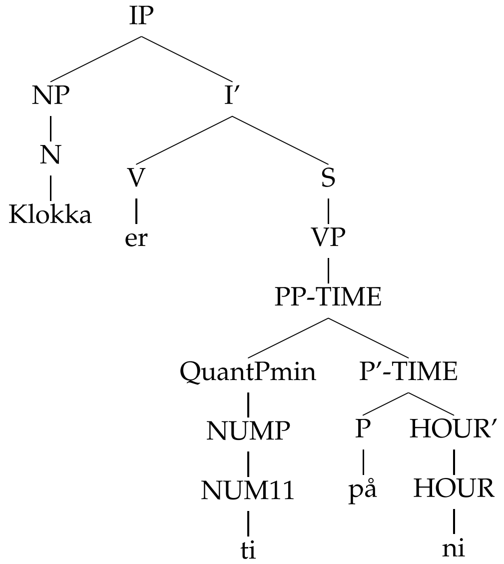
[4] Hodet i klokkeslettsuttrykket er uthevet.

(3) *Kampen er klokka halv fire*



(3)
$$\left[\begin{array}{ll}
\text{PRED} & \text{'være}\left\langle\left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{PREDLINK}\right)\right\rangle\text{'} \\
\text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{'kamp'} \\ \text{CLOCK-EVENT} & +\end{array}\right] \\
\text{PREDLINK} & \left[\begin{array}{ll}\text{PRED} & \text{'halv}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{'hour-4'}\end{array}\right] \\ \text{CLOCK-EXPRESSION} & + \\ \text{CLOCK-OCCURRENCE} & +_\end{array}\right]
\end{array}\right]$$

(4) *Klokka er ti på ni*



(4)
$$\left[\begin{array}{ll}
\text{PRED} & \text{'være}\left\langle\left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{PREDLINK}\right)\right\rangle\text{'} \\
\text{SUBJ} & \left[\text{PRED} \quad \text{'klokke'}\right] \\
\text{PREDLINK} & \left[\begin{array}{ll}
\text{PRED} & \text{'på}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\
\text{OBJ} & \left[\text{PRED} \quad \text{'hour-9'}\right] \\
\text{PTYPE-CLOCK} & \left[\text{TEMPFINALE} +\right] \\
\text{SPEC} & \left[\begin{array}{l} \text{PRED 'pro'} \\ \text{SPEC}\left[\text{NUMBER}\left[\text{PRED 'ti'}\right]\right] \end{array}\right] \\
\text{CLOCK-PREDLINK} & + \\
\text{CLOCK-EXPRESSION} & + \\
\text{CLOCK-OCCURRENCE} & -
\end{array}\right]
\end{array}\right]$$

Skal *være* ta et tidsadverbial med den grammatiske funksjonen PREDLINK, fordrer dette at subjektet viser til en konkret hendelse som forløper over tid dersom subjektet da ikke viser til betydningen *klokke*, og hele setningen uttrykker et tidspunkt (som i (4)). Dette er grunnen til at (6) som enkeltsetning er usemantisk. (6) kan imidlertid settes inn i en kontekst hvor setningen vil gi mening (7).

Dersom NorGrams grammatikk skulle blitt endret slik at (6) ville bli analysert som ugrammatisk, måtte alle nomen som kunne vise til en hendelse tagges slik at en restriksjon kunne legges inn, slik at verbet *være* krevde at subjektet må ha taggen 'hendelse' dersom PREDLINK er et tidsadverbial. En slik restriksjon ville ha analysert (6) som ugrammatisk såfremt *hus* ikke hadde denne taggen. Men siden så og si alle nomen kan gis en kontekst hvor de kan referere til en hendelse (som i (7)), vil ikke dette være ønskelig.

(5) Middagen er klokka åtte.

(6) ?Huset er klokka ni.

(7) Visningen av naustet er klokka halv ni, og huset er klokka ni.

Hvilke nomen på subjektsplass som kan forstås som en hendelse blir også diskutert i avsnitt 4.2.2. I de fleste tilfellene der subjektet ikke er *klokka/klokken*, vil *være* kunne erstattes med *begynne*. *Være* tolkes i (5) som *begynne* ut ifra kontekst og verdenskunnskap. Imidlertid kan det tenkes kontekster hvor *være* ikke opptrer synonymt med *begynne*. I (8) kan *er* tolkes som *begynne*, men det kan også tolkes som *holder på*.

(8) - Skal vi spise middag klokka sju?
- Er du gal? Fotballkampen er jo klokka sju!

Bruken av *være* i disse kontekstene vil sannsynligvis forekommer hyppigere muntlig enn skriftlig. Men det er på ingen måte grammatisk feil. I en automatisk parse må derfor tidsadverbialet analyseres som PREDLINK i f-struktur i disse tilfellene. Man kan godt tenke seg at andre verb kunne ta et klokkeslett som argument, men denne analysen blir ikke brukt verken i min eller i NorGrams grammatikk. Slik kunne en analyse av *finne sted* og *vare* sett ut dersom tidsadverbialet var inkludert i argumentstrukturen:

(9) Møtet fant sted klokka halv ti.

$$\begin{bmatrix} \text{PRED} & \text{'finne sted}\left\langle\left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{XCOMP}\right)\right\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{'møte'}\\ \text{CLOCK-EVENT} & +\end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'halv}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix}\text{PRED} & \text{'hour-10'}\end{bmatrix} \\ \text{CLOCK-EXPRESSION} & + \\ \text{CLOCK-OCCURRENCE} & +\_ \end{bmatrix}\end{bmatrix}$$

(10) Kampen har vart siden klokka åtte.

$$\begin{bmatrix} \text{PRED} & \text{'vare}\left\langle(\uparrow \text{SUBJ})(\uparrow \text{XCOMP})\right\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{'kamp'}\\ \text{CLOCK-EVENT} & +\end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'siden}\left\langle(\uparrow \text{OBJ})\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix}\text{PRED 'hour-8'}\end{bmatrix} \\ \text{CLOCK-EXPRESSION} & + \\ \text{CLOCK-OCCURRENCE} & +\_ \end{bmatrix}\end{bmatrix}$$

Slik min grammatikk er definert nå, vil disse tidsadverbialene blir kategorisert som ADJUNCT i f-strukturen, og verbene vil da bli analysert som intransitive.

Frie adverbial

De frie adverbialene står utenfor predikatets argumentstruktur og er en del av f-strukturens ADJUNCT. Disse adverbialene står som et motstykke til de bundne adverbialene, hvis forekomst er avhengig av predikatet. Et klokkeslettsuttrykk kan stort sett alltid føyes til en setning som et fritt adverbial, men resultatet kan bli at setningen vil falle unaturlig (se (11)). Setningen vil likevel ikke være ugrammatisk.

(11) Sokrates var en innflytelsesrik filosofi klokka ti over ni.

3.2.3 Syntaktiske egenskaper for klokkeslettsuttrykkene

Frie tidsadverbial

Klokkeslettsuttrykk som står som frie tidsadverbial følger stort sett de syntaktiske reglene til frie adverbial generelt, men det finnes noen begrensninger. I motsetning til en del andre frie tidsadverbial, som også kan stå i midtfeltet, kan klokkeslettsuttrykkene bare stå i forfeltet og sluttfeltet i setningen.

(12) Jeg har den siste tiden vært syk.

(13) *Jeg ble klokka halv åtte syk.

De frie tidsadverbialene kan heller ikke deles opp ved topikalisering. Denne regelen følger også klokkeslettsuttrykkene, og dette ser ut til å være et kjennemerke for setningsledd som står utenfor verbets argumentstruktur også i engelsk.

> [E]xtraction from adjuncts is generally ungrammatical. These restrictions have come to be known collectively as island constraints. (Falk, 2001, s. 162)

Ved topikalisering[5] av tidsadverbialet må hele frasen flyttes.

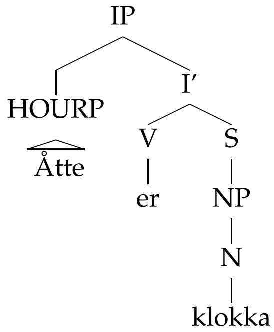
(14) Klokka ti minutter over halv ni kjørte mannen hjem.

(15) *Halv ni kjørte mannen hjem klokka ti minutter over.

Bundne tidsadverbial

Bundne tidsadverbial kan, i motsetning til de frie, splittes opp etter særskilte regler. Dersom klokkeslettsadverbialets hode i PREDLINK ikke er en preposisjon, kan hele klokkeslettsuttrykket topikaliseres. Denne flyttingen ligner flyttingen objektet til et bivalent verb kan gjøre (17).

(16) *Åtte er klokka*



(17) Middag spiser mannen.

Dersom hodet i et klokkeslettsuttrykk i PREDLINK er en preposisjon, vil denne preposisjonen enten være hode i PP-TIME eller PP. I dette tilfelle kan hele tidsadverbialet flyttes fram til forfeltet i setningen (som (19)), eller deler av det (20). Dersom bare deler av frasen flyttes, vil den framflyttede delen av tidsadverbialet være objektet til hodet i tidsadverbialet. I dette tilfellet oppstår en langdistanseavhengighet (Falk, 2001, s. 149-172). Tidsadverbialets hode vil da stå sist i setningen, og dersom flere adverbial skal føyes til, må disse følge etter det framflyttede objektet til tidsadverbialets hode (21).

(18) Klokka er ti minutter over halv ni.

(19) Ti minutter over halv ni er klokka.

(20) Halv ni er klokka ti minutter over.

[5] Falk (2001) gjør et skille mellom *topic* og *focus*, men denne diskusjonen tar jeg ikke for meg her. Det framflyttede elementet kaller jeg et topikalisert element.

(21) Halv ni om morgenen er klokka ti minutter over.

De syntaktiske reglene som gjelder for topikalisering av et klokkeslettsadverbial med en preposisjon som hode, sammenfaller med de reglene en preposisjonsfrase generelt er underlagt. Dette gjelder uavhengig av om preposisjonsfrasen står innad i argumentstrukturen til predikatet (som i (22)), eller om den står som ADJUNCT i setningen, som i (23).

(22) Bordet legger mannen avisa på.

(23) Kniv og gaffel spiser mannen med.

Sammenfallet av den syntaktiske strukturen til preposisjonsfraser generelt, og klokkeslettsfraser der en preposisjon forekommer, indikerer at disse klokkeslettsfrasene må analyseres slik at preposisjonen er hodet i disse frasene. Preposisjonene innad i klokkeslettsuttrykket er tema i følgende avsnitt.

3.3 Preposisjonene

I dette avsnittet vil jeg ta for meg de forskjellige preposisjonene som jeg har behandlet i avhandlingen. Disse preposisjonene står som hode i en PP-frase, eller som hode i PP-TIME. Felles for preposisjonene er at de enten kan ta et klokkeslettsuttrykk som utfylling eller selv være hode i et klokkeslettsuttrykk. Utfyllingen kalles *objektet* til preposisjonen, og jeg vil holde meg til denne benevnelsen.

3.3.1 *Over* og *på*

Preposisjonene *over* og *på* er de preposisjonene som i størst grad blir brukt internt i klokkeuttrykk. De kan kalles klokkefrasens grunnpreposisjoner. Dette er de to eneste preposisjonene som kan stå som hode i PP-TIME (om PP-TIMEs frasestrukturregler i XLE, se avsnitt 4.3). Internt i PP-TIME-frasen kan preposisjonene semantisk leses som *mer enn* eller *mindre enn*, eller *senere enn* og *tidligere enn*.

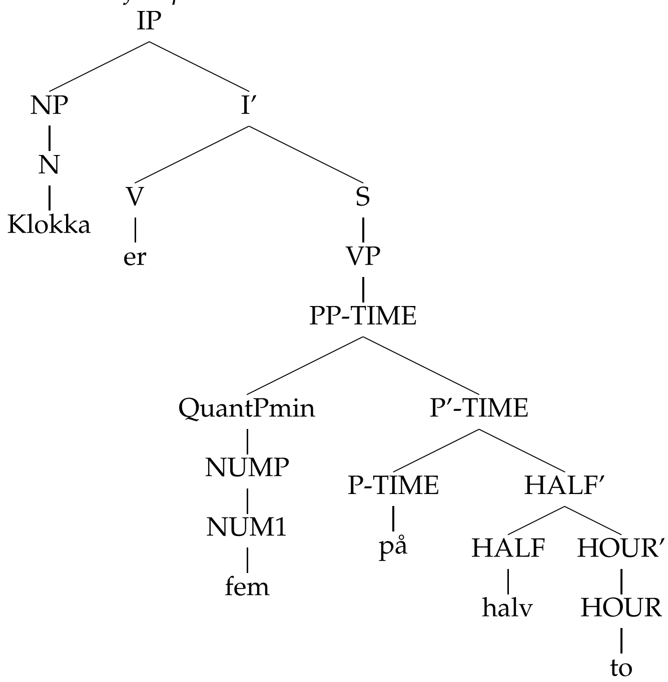
Over

For preposisjonen *over* finner vi igjen betydningen *mer enn* også i uttrykk som ikke er relatert til klokkeslett. Under en av oppføringene til *over* i *Bokmålsordboka* (Landrø og Wangensteen, URL) finner vi: "mer enn *det var o- hundre deltakere i rennet | klokka er o- ti | by o- noen*". *Over* blir også brukt i situasjoner som "Det kommer til å regne over helga", eller "Jeg er tilbake over sommeren". Her kan det diskuteres hvorvidt *over* skal forstås som en lokativ preposisjon som brukes i overført betydning. Tidsrommet det refereres til kan intuitivt oppfattes som noe som ligger bakenfor *helga* eller *sommeren*. En annen tolking er å lese *over* i denne sammenhengen som *senere enn*, en betydning preposisjonen kan tillegges også når den står i et klokkeslettsuttrykk.

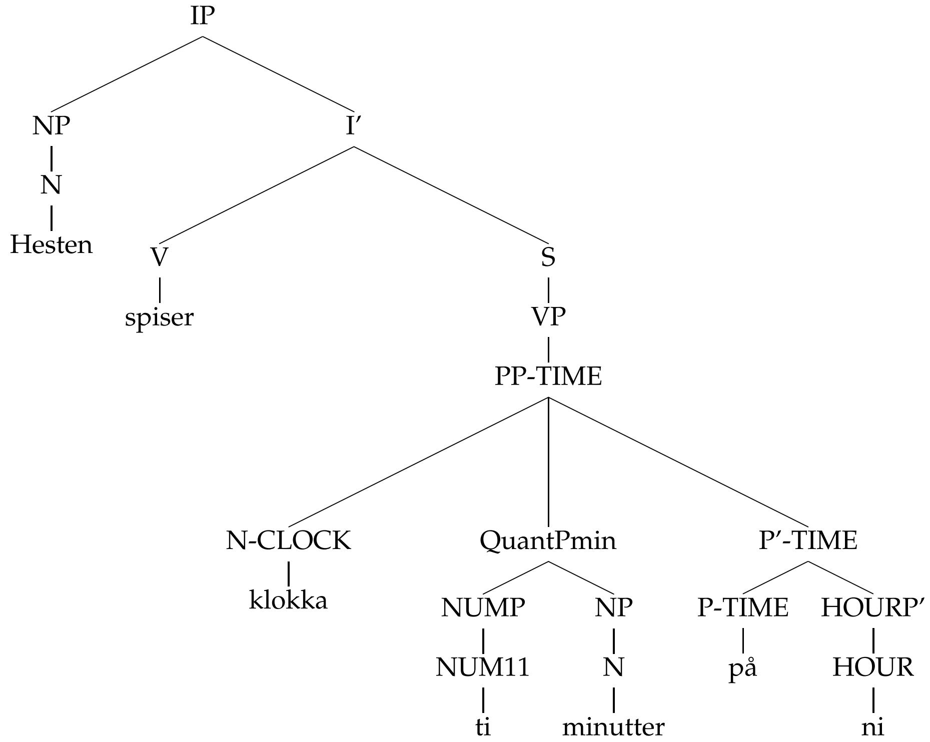
På

Preposisjonen *på* har en mer begrenset distribusjon enn *over* i tidsuttrykk. *På* blir brukt i tidsuttrykk av ulik art, for eksempel i uttrykk som "på slaget fem", "komme på timen" og "på lenge". Men dersom *på* skal være et synonym til *mindre enn/tidligere enn*, må preposisjonen stå i en helt bestemt kontekst. I min grammatikk er denne konteksten avgrenset til PP-TIME. *På* fungerer bare som preposisjon i et klokkeslettsuttrykk dersom *på* står som hode i PP-TIME, som i (24) og (25):

(24) *Klokka er fem på halv to*



(25) *Hesten spiser klokka ti minutter på ni*



I (24) står et tallord alene som spesifikator for PP-TIME. I PP-TIME blir dette tallordet forstått som en kvantor i en kvantorfrase der hodet i frasen er undertrykt, men representerer nomenet *minutter*. Slike konstruksjoner finnes ikke i en vanlig PP-frase.

Forskjellene mellom *over* og *på*

Over kan brukes i kontekster hvor *på* vil være ugrammatisk, og hvor *over*s antonym vil bli *før* i stedet for *på*. Disse forekomstene finner vi i tidsreferanser hvor preposisjonen er hode i en PP som tar et klokkeslettsuttrykk som objekt. Eksemplene nedenfor illustrerer dette. (27) er ugrammatisk fordi *på* ikke kan stå i denne posisjonen. Preposisjon *før* må her stå som *over*s antonym (som i setning (28)).

(26) Klokka er like over åtte.

(27) *Klokka er like på åtte.

(28) Klokka er like før åtte.

På kan ikke stå som hode i PP i den underforståtte betydningen *mindre enn*. Denne egenskapen har preposisjonen bare internt i PP-TIME-frasen. Årsaken er at en av grunnbetydningene til *over* er *mer enn*, mens *på* bare i helt særskilte klokkeuttrykk betyr *mindre enn*. Situasjonen er den samme hvis gradsadverbialet faller bort:

(29) Klokka var over åtte.

(30) Klokken var før 24.00.[6]

(31) *Klokka var på tolv.

Alle disse tidsuttrykkene er vanlige PP-fraser som tar klokkeslettet som objekt. I disse PP-frasene kommer skillet i bruken av *på* og *over* tydelig fram. Mens *over* i denne sammenhengen kan brukes i betydningen *senere enn*, er ikke *på* et gyldig antonym.

I (29) og (30) opptrer hodet i PP uten spesifikator. *Over* kan bare fungere som hode i denne konteksten dersom PPs grammatiske funksjon er PREDLINK, og hele setningen uttrykker et klokkeslett.[7] I (32) vil ikke preposisjonsfrasen være et velformet klokkeslettsuttrykk, siden subjektet viser til en hendelse:

(32) *Kampen er over åtte.

Over åtte vil ikke i (32) bli tolket som en tidsreferanse. Imidlertid kan det settes inn en spesifikator, for eksempel et gradadverbial, foran preposisjon. I så fall oppfattes preposisjonsfrasen som et klokkeslettsuttrykk:

(33) Kampen er like over åtte.

De samme restriksjonene gjelder dersom den grammatiske funksjonen er ADJUNCT. *Over* kan heller ikke her opptre uten spesifikator innenfor PP, dersom *over* i seg selv er en del av et klokkeslettsuttrykk, eller tar et klokkeslettsuttrykk som objekt. Det er interessant å merke seg at denne restriksjonen ikke inntreffer i alle preposisjonsfraser med temporal referanse:

(34) Vi kommer igjen over jul.

(35) *Vi spiser middag over fire.

Årsaken til denne distinksjonen kan være at *over* i (34) forstås som en lokativ preposisjon brukt i overført betydning, se avsnitt om *over*, side 39.

3.3.2 Andre preposisjoner

Faarlund et al. (1997, s. 1069) behandler en del ord som innleder tidsleddsetninger. Disse ordene er *etter, fra, før, innen, siden* og *(inn)til*. Ordene opp-

[6] Hentet fra *Dagens Næringsliv*, http://www.dn.no/arkiv/article32524.ece, 18.12.2005. En slik ordlyd forekommer sjelden på norsk, og denne tidsrelasjonen blir oftest uttrykt på en annen måte.

[7] Dersom subjektet refererer til en hendelse, kan ikke *over* innlede slike klokkeslettsuttrykk. Disse tilfellene er drøftet i avsnitt 4.2.2

trer både som subjunksjoner og preposisjoner, avhengig av deres grammatiske funksjon i setningen. Som subjunksjoner innleder disse ordene leddsetninger:

(36) Siden du ikke kom, dro jeg hjem.

Disse subjunksjonene/preposisjonene kan også stå som hode i PP. Ofte viser preposisjonens objekt da til et tidspunkt eller en hendelse som kan tidfestes, og for preposisjonene *siden* og *før* er dette et krav. Objektet kan være et klokkeslett, eller utfyllingen kan være en nomenfrase som viser til en bestemt tid eller hendelse:

(37) Jeg har vasket opp siden frokost.

(38) Fra høsten (av) blir det andre boller.

(39) Jeg skal slappe av etter jul.

I disse eksemplene står objektet til preposisjonen som en tidsreferanse og kunne i alle tilfellene vært byttet ut med et klokkeslettsuttrykk. De ulike preposisjonene krever også at verbet står i bestemte tempus, men dette kommer jeg tilbake til i behandlingen av hver enkelt preposisjon. I min minigrammatikk har jeg ikke lagt inn restriksjoner for hvilke objekter disse preposisjonene kan knytte til seg, så lenge objektet ikke er et klokkeslettsuttrykk.

3.3.3 Preposisjonenes syntaktiske krav til setningen

I klokkeslettsuttrykk stiller preposisjonene ulike krav til resten av setningens syntaks. Tabell 3.1 framstiller hvilke av preposisjonene som tar spesifikator, og hvilke av preposisjonene som også fungerer som subjunksjon. Med spesifikator menes her et ledd som kommer umiddelbart før hodet i preposisjonsfrasen og som semantisk modifiserer tidsuttrykket. Dette kan være et gradsadverbial (som *like* og *rett*), eller kvantorfraser eller nomenfraser med temporal referanse (for eksempel *ti minutter* eller *sekunder*).'± Subjunksjon' indikerer om preposisjonen også kan stå som subjunksjon, i tillegg til å fungere som hode i PP/PP-TIME. Kategorien '±Tar PREDLINK' viser hvorvidt preposisjonen kan stå som hode i PREDLINK, eller om preposisjonen bare kan forekomme i den grammatiske funksjonen ADJUNCT.

	Subjunksjon	Tar spesifikator	Krever spesifikator	Tar PREDLINK
Innen, siden, fra, til	+	-	-	+[8]
etter, før	+	+	-	+
over	-	+	-[9]	+
på	-	+	+	+

Tabell 3.1: Preposisjonens egenskaper

(40) *Hesten spiser like fra halv åtte.

(41) Hesten spiser før halv åtte.

(42) *Hesten spiser på halv åtte.

(43) Kampen er til ni.

Eksempel (40) er ugrammatisk siden *fra* ikke kan ta denne typen spesifikator, og eksempel (42) er ugrammatisk fordi *på* krever spesifikator. *Før* stiller ingen krav til at en spesifikator må stå foran preposisjonen, og (41) er derfor grammatisk. (43) er grammatisk siden ingen spesifikator forekommer her, og subjektet refererer til en hendelse.

Innen, siden, fra og *til*

Det er naturlig å dele preposisjonene inn i tre klasser i forhold til deres syntaktiske krav til resten av setningen. *Innen, siden, fra* og *til* opptrer likt med hensyn til de syntaktiske kriteriene i tabell 3.1. I en preposisjonsfrase som tar et klokkeslettsuttrykk som objekt, kan ikke disse preposisjonene ta en spesifikator som modifiserer klokkeslettsuttrykket.[10] Denne restriksjonen skiller disse preposisjonene fra preposisjonene *etter* og *før*:

(44) Vi spiser middag ti minutter etter åtte.

(45) *Vi spiser middag ti minutter siden åtte.

På den endimensjonale tidslinjen kan vi si at *innen* og *til* refererer til tidsrommet fram til tidspunktet[11] som er definert i preposisjonsfrasens objekt. *Siden* og *fra* tar derimot utgangspunkt i tidspunktet i preposisjonsfrasens

[8] Kan bare stå som hode i PREDLINK når subjektet viser til en hendelse.

[9] Spesifikatoren kan bare utelates når preposisjonsfrasens grammatiske funksjon er PREDLINK, og hele setningen uttrykker et tidspunkt, det vil si at subjektet ikke refererer til en hendelse.

[10] *Siden* tillater at *helt* eller *like* står foran denne preposisjonen i uttrykk som *Hesten har spist helt/like siden halv åtte*. Men *like* og *helt* er ikke her å betrakte som semantiske modifiseringer av klokkeslettet, men snarere en forsterkning.

[11] Se definisjonen på *tidspunkt*, kapittel 3.2.1.

objekt, og tiden etter dette. Alle disse preposisjonene griper fast i et konkret punkt på tidslinjen, og orienterer ut ifra dette. Objektet til preposisjonen definerer dette utgangspunktet, og det tidsspennet uttrykket definerer per se, er tidsrommet fra objektets tidspunkt og i uendelig retning framover eller bakover på tidslinjen.

Konteksten for øvrig modererer som oftest tidsrommet fra å definere uendelighet til å referere til et begrenset tidsrom. Det tas for eksempel ofte utgangspunkt i handlingstiden[12] i ytringen og fram til tidspunktet som defineres i PPs objekt når preposisjonen er *innen* eller *til*, som i (46) og (47):

(46) Flyet kommer innen halv åtte.

(47) Du må være tilbake til klokka fire.

For *siden* ligger handlingstidspunktet alltid bak i tid, siden verbet må stå i perfektum (se også avsnittet om *siden*, side 50). Tidsrommet som defineres i setninger med *siden* som hode i et klokkeslettsuttrykk, vil derfor avsluttes i utsagnsøyeblikket[13] dersom det finitte verbet står i presens perfektum (48). Står verbet i preteritum perfektum vil tidsrommet være avsluttet i fortiden (49).

(48) Han har vært her siden klokka ni.

(49) Han hadde vært siden klokka ni, men ble lei og gikk hjem.

Preposisjonen *fra* kan derimot potensielt stå i tidsuttrykk som definerer et tidsrom som går mot uendelig. Det er bare verdenskunnskap som gjør oss i stand til å avgjøre at tidsrommet tidsadverbialet refererer til, likevel er begrenset (50).

(50) Fotballkampen er fra halv åtte.

Terminative og punktuelle tidsrelasjoner

Tidsrelasjonene som defineres av preposisjonene *innen, siden, fra* og *til* er terminative. En terminativ tid står i motsetning til en punktuell tidsrelasjon. Punktuell tid defineres som en tidsrelasjon som viser til et avgrenset tidsrom på tidslinjen. En terminativ tidsrelasjon refererer derimot til handlinger som finner sted før eller etter et visst tidspunkt på tidslinjen (Faarlund et al., 1997, s. 547). Terminative tidsrelasjoner kan også uttrykkes med preposisjonene *før* og *etter* dersom de står uten spesifikator i PP (51).

(51) Etter oss kommer syndfloden.

[12]Punktet på tidslinjen hvor handlingen finner sted med utgangspunkt i verbets tempus, se Faarlund et al. (1997, s. 544).

[13]Punktet på tidslinjen hvor selve ytringen finner sted, se Faarlund et al. (1997, s. 544). I setninger i nåtid, vil utsagnsøyeblikket falle sammen med handlingstiden.

Derimot definerer *etter* og *før* en punktuell tidsrelasjon dersom de får en temporal spesifikator (som i eksempel (44)). Denne muligheten har ikke *innen, siden, fra* og *til*, og preposisjonsfraser med en spesifikator foran disse preposisjonene er ugrammatiske dersom preposisjonsfrasen er et klokkeslettsuttrykk (52), eller et tidsuttrykk generelt (53):

(52) *Vi spiser middag en stund fra klokka åtte.

(53) *De skal være her noen uker til jul.

På grunn av disse preposisjonenes orientering ut ifra et definert tidspunkt, kan de heller ikke stå som hode i et klokkeslettsuttrykk der klokkeslettsuttrykkets grammatiske funksjon er PREDLINK, så lenge intensjonen er at hele setningen skal definere et tidspunkt:

(54) Konserten er til klokka ni.

(55) *Klokka/Den er fra halv ni.

(56) Konserten har vært siden ni.

(57) Konserten varer til halv ni.

Jeg vil anta at tilfeller hvor disse preposisjonene står som hode i PREDLINK, vil være mer frekvente i muntlige enn i skriftlige sammenhenger. Mitt inntrykk av setning (54) og (56) er at de føles mer uformelle enn setning (57), hvor verbet *være* byttes ut.

På **og** ***over***

I den andre yttergrensen finner vi *på* og *over*. Disse preposisjonene er de eneste som kan stå som hode i PP-TIME. Fraseregelen til PP-TIME krever at en spesifikator er til stede, noe som skiller PP-TIME grammatisk fra PP. Men som nevnt i avsnittet 'Forskjellene mellom *over* og *på*', side 41, kan *over* også stå som hode i PP, under visse omstendigheter også uten spesifikator, se tabell 3.1. Dersom *over* står uten spesifikator, uttrykkes i utgangspunktet en terminativ tidsrelasjon (som i eksempel (58)).

(58) Klokka er over åtte.

Dersom (58) er et svar på spørsmålet *'Hvor mye er klokka?'*, uttrykkes gjerne implisitt en punktuell tidsrelasjon hvor spesifikatoren er undertrykket, og svaret skal gjerne tolkes synonymt med (58).

(59) Klokka er like over åtte.

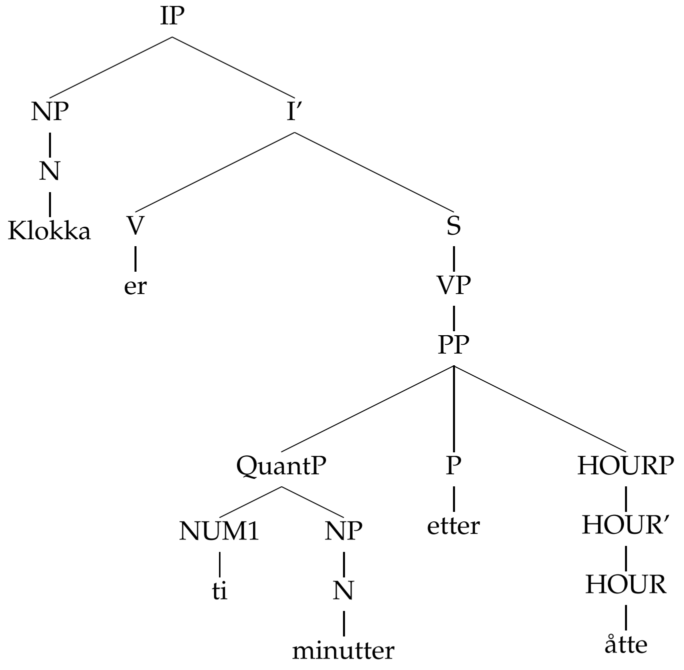
Imidlertid trenger ikke (58) bare å referere til tidsrommet like etter åtte, og (58) er i utgangspunktet en terminativ tidsrelasjon. I (60) er det ingen tvil om at det er en terminativ tidsrelasjon som beskrives.

(60) Det er ikke lov å ta pause før klokka er over åtte.

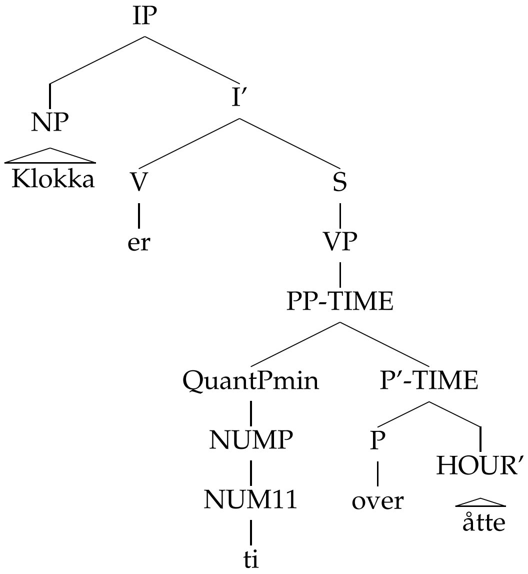
Etter og *før*

Mellom disse to yttergrensene står *etter* og *før*. Disse preposisjonene kan bare stå som hode i PP, men de kan stå både med og uten spesifikator. Semantisk sett ligger disse preposisjonene svært nært *over* og *på*. *Etter* og *før* kan ta mange av de samme spesifikatorene som *over* og *på*, og meningsinnholdet vil bli det samme. Men den syntaktiske analysen vil være forskjellig, ettersom preposisjonene er hoder i ulike fraser:

(61) *Klokka er ti minutter etter åtte.*

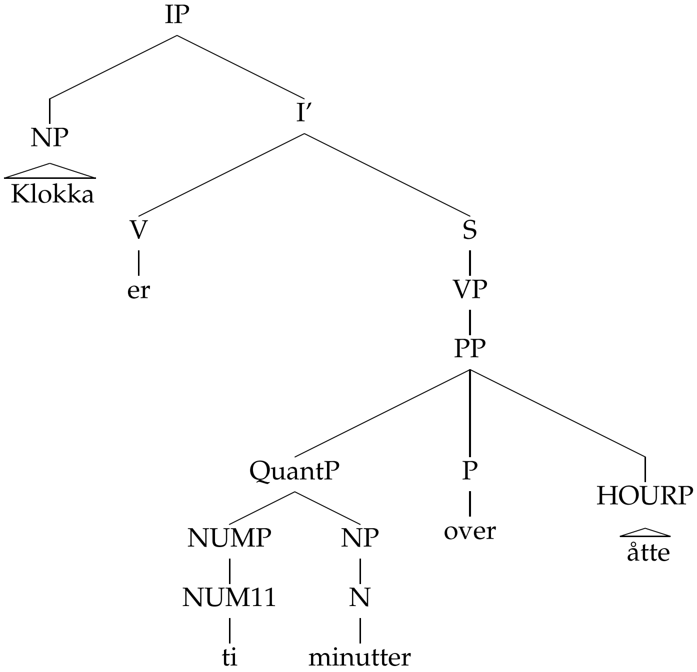


(62) *Klokka er ti over åtte*



I (61) ser vi at *etter* er hode i PP, mens i (62) er *over* hode i PP-TIME. I (62) er dette den eneste mulige analysen. Derimot vil setningen *Klokka er ti minutter over åtte* ha to analyser. Her kan *over* analyseres både som hode i PP-TIME eller i PP:

(63) *Klokka er ti minutter over åtte*



Denne doble analysen er en konsekvens av at *over* kan opptre både innenfor PP-TIME, og i klokkeslettsuttrykk i PP. Dette er i utgangspunktet ikke en ønskelig situasjon, og fenomenet drøftes i avsnitt 4.4.2.

Som nevnt i avsnittet om *Innen, siden, fra* og *til* (side 44), kan *etter* og *før* orientere ut fra et definert tidspunkt, og potensielt uendelig langt framover eller bakover på den endimensjonale tidslinjen. *Etter* og *før* har disse egenskapene dersom det ikke står en spesifikator foran preposisjonen i PP, som i (64):

(64) Vi ser fotballkampen etter klokka ni.

Vi kan si at tidsrelasjonen som *før* og *etter* definerer, er terminativ i klokkeslettsuttrykk dersom det ikke står en spesifikator foran preposisjonen, som i (64). Derimot er tidsrelasjonen punktuell i setningen (61). Spesifikatoren orienterer tidsrelasjonen til et definert tidsrom på tidslinjen.

I en terminativ orientering fra ett punkt og til uendelig, skulle man intuitivt tro at *etter* og *før* i konstellasjoner som (64) ikke kunne stå i PRED-LINK dersom hele setningen refererer til et klokkeslett, noe som *innen, siden, fra* og *til* ikke kan:

(65) *Klokka er siden halv ni.

Denne hypotesen støttes av både Oslo-korpuset (OK, URL) og Norsk aviskorpus (NA, URL) hvor jeg ikke finner noen slike forekomster. Rent tekstsøk ved *www.google.com* avslører imidlertid at disse konstellasjonene forekommer, men de er ikke særlig utbredt. Jeg har tidligere nevnt setningen fra Dagens Næringsliv, fra eksempel (30):

(66) Klokken var før 24.00.

De spredte forekomstene av disse konstruksjonene gjør at jeg tillater dette i min grammatikk. Intensjonen ved uttrykket (66) er sannsynligvis å uttrykke at det snart var midnatt, og dette blir vanligvis uttrykt ved å skyte inn en spesifikator, som i (67):

(67) Klokka var like før 24.00.

Dermed orienterer ikke klokkeslettsuttrykket ut fra et punkt og mot uendelig. Spesifikatoren *like* markerer at det er snakk om en tidsperiode som står nær tidspunktet i preposisjonens objekt.

3.3.4 Krav til verbet i setningen

I dette avsnittet belyser jeg hvilke generelle, syntaktiske restriksjoner de ulike preposisjonene legger på verbet i setningen. Disse restriksjonene kan implementeres i min grammatikk dersom dette er ønskelig.

Siden

Dersom *siden* er hode i en preposisjonsfrase som ender i et klokkeslettsuttrykk, er verbformen alltid en perfektum.

(68) Han har ikke vært her siden klokka ni.

(69) Han hadde ikke spist siden ti på to.

(70) Han ville ikke ha vært her siden halv åtte i dag morges, om han ikke hadde stukket innom i tre-tiden.

Fra og *til*

Dersom *fra* innleder en tidsadverbial og verbet i setningen står i perfektum, beskrives i utgangspunktet den samme tidsrelasjonen som *siden* innleder. Dette er en terminativ tidsrelasjon som strekker seg fra tidspunktet som defineres i adverbialet og framover. Korpussøk og rene tekstsøk på internett tyder på at *siden* er langt mer frekvent som innleder av disse tidsadverbialene enn det *fra* er dersom verbtiden er perfektum. Av de to setningen (71) og (72), er også (72) den klart mest vanlige. (71) kan imidlertid ikke sies være en uheldig konstruksjon i seg selv.

(71) Han har vært her fra åtte.

(72) Han har vært her siden åtte.

Verbformer i perfektum er vanlig dersom tidsadverbialet innledes av *til*. *Til* stiller ikke konkrete krav til hvilke verbtider som er akseptable i setningen, men det stiller derimot krav til verbalhandlingens aksjonsart. Disse begrensningene er felles for *siden*, *fra* og *til* og behandles i neste avsnitt.

Restriksjoner knyttet til verbalhandlingens aksjonsart

Vendler (1967) identifiserte fire ulike aksjonsarter i engelske verbalfraser, og denne inndelingen er også relevant for norsk. Taksonomien deler aksjonsartene inn i kategorier ut ifra hvordan tidsforløpet til en situasjon skal tolkes. Denne inndelingen er (Saeed, 1997, s.123):

1. **States** - Tilstander
   Engelsk: *love, believe, know*
   Norsk: *elske, tro, vite*

2. **Activities** - Aktiviteter
   Engelsk: *run, push a cart*
   Norsk: *svømme, kjøre bil*

3. **Accomplishments** - Gjennomføringer
   Engelsk: *run a mile, walk to school*
   Norsk: *løpe maraton, tegne en tegning*

4. **Achievements** - Prestasjoner
   Engelsk: *recognize, find*
   Norsk: *nå toppen, stoppe*

Et klokkeslettsuttrykk hvor hodet er *fra*, *til* eller *siden*, må beskrive situasjoner hvor aksjonen ikke allerede er avsluttet. Andersson (1977, s. 85) leser videre inn dikotomien *imperfektiv* og *perfektiv* i Vendlers klassifisering, der de to første gruppene, *tilstander* og *aktiviteter*, grupperes som imperfektive. De to siste gruppene beskriver perfektive handlingsforløp. Klokkeslettsuttrykk kan da bare innledes av *fra/til/siden* dersom aksjonsarten i setningen er *imperfektiv*.

(73) **Tilstander**: Jeg visste det fra klokka halv åtte.

(74) **Aktiviteter**: Hun svømte til halv ni.

(75) **Gjennomføringer**: *Hun har kjørt fra Voss til Bergen fra halv ni.[14]

(76) **Prestasjoner**: *Han fant henne fra halv åtte.

Tidsadverbial innledet av *fra/til/siden* kan bare beskrive den temporale utstrekningen til imperfektive aksjonsarter, fordi *fra/til/siden* innleder en terminativ tidsrelasjon som ikke er avsluttet i tid.

Innen

Innen oppfører seg som *før* i mange kontekster, men dette er først og fremst i situasjoner der *innen* opptrer som subjunksjon (som i (77) og (78)). Dersom *innen* tar et klokkeslettsuttrykk som objekt (som i (79)), kan *innen* vanskelig bytte plass med *før* uten at setningen føles ukorrekt.

(77) Jeg kommer innen du går.

(78) Jeg kommer en stund innen du går.

(79) ??Jeg kommer noen minutter innen halv åtte.

(80) *Klokken er innen halv ni.

(79) er ikke tillatt i min grammatikk på grunn av regelen om at denne typen preposisjon ikke kan ta spesifikator dersom objektet er et klokkeslettsuttrykk. *Innen* kan innlede tidsadverbial uavhengig av hvilken aksjonsart verbet beskriver. Vi ser imidlertid at dersom aksjonsarten er *prestasjon*, vil *innen* alltid opptre synonymt med *før* (81).

[14] Denne setningen kan forekomme, men i så fall tolkes det som om bilturen fortsatt ikke er avsluttet, og aksjonsarten blir en *aktivitet* og ikke en *gjennomføring*.

(81) Jeg skal finne henne innen halv ni.

Jeg har ikke lagt inn noen restriksjoner i forhold til hvilke verbtider som godtas når *innen* står som hode i et klokkeslettsuttrykk. Dette vil jeg heller ikke anbefale, men korpussøk avslører at verbet sjelden står i preteritum dersom *innen* innleder et tidsadverbial.[15] *Innen* blir særlig ofte brukt i forbindelse med verb i futurum-former (82), eller presens-former som betegner framtid (83). Dersom verbtiden er fortid, står verbformen som regel i en form av perfektum. Alle de tre setningene nedenfor er hentet fra *Norsk aviskorpus*.

(82) Innen klokken 10 skal trafikken gå som normalt.

(83) 60 000 innbyggerne forlater hjemmene sine innen klokken to i dag.

(84) Sea Kingene var oppe og fløy innen klokken 15.

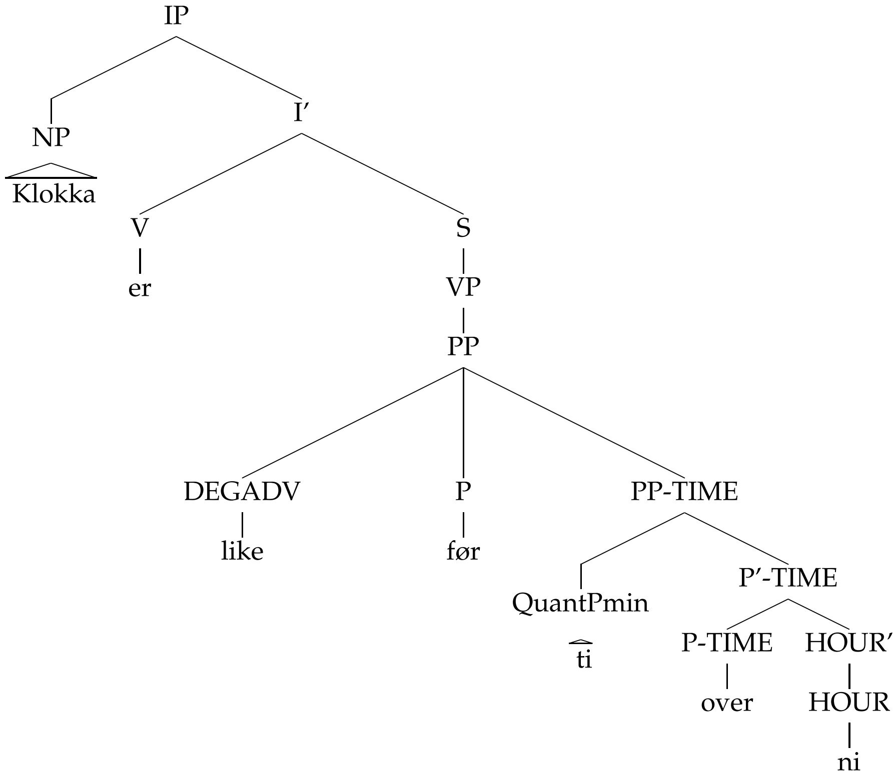
Etter og *før*

Færrest restriksjoner legger preposisjonene *før* og *etter* på verbet i setningen. Dersom disse preposisjonene står som hode der preposisjonsfrasen er et klokkeslettsuttrykk, er det ikke mulig å generalisere over hvilke verbtider eller aksjonsarter som vil gjøre setningen ugrammatisk. I motsetning til de andre preposisjonene som også kan stå som subjunksjoner, kan *etter* og *før* også stå som hode i preposisjonensfraser der den grammatiske funksjonen er PREDLINK (86), i tillegg til i klokkeslettsuttrykk hvor funksjonen er ADJUNCT (85).

(85) Hesten spiste før klokka ti over ni.

[15] Eksempel (84) er et tilfelle hvor dette likevel forekom.

(86) *Klokka er like før ti over ni*



3.4 Oppsummering

I dette kapitlet har jeg sett på hvilke faktorer som griper inn i syntaksen til klokkeslettsuttrykket. De syntaktiske restriksjonene som øves på klokkeslettsuttrykket varierer ut ifra hvilken grammatisk funksjon tidsadverbialet har i setningen, og hvilken preposisjon som står som hode i dette tidsadverbialet. I avsnitt 3.2.1 plasseres klokkeslettsuttrykket innenfor *Norsk referansegrammatikk*s klassifisering av norske adverbialer. Klokkeslettsuttrykkene er enten bundne eller frie tidsadverbial, alt etter hvilken grammatisk funksjon de har i setningen. Klokkeslettsuttrykkets syntaks varierer også etter hvilken grammatisk funksjon tidsadverbialet fyller.

I avsnitt 3.3 har jeg gått grundig gjennom syntaktiske restriksjoner preposisjonene som kan stå som hode i klokkeslettsuttrykk pålegger resten av setningen. Jeg har også sett på hvilke tidsrelasjoner disse preposisjonene introduserer når de står som hode i et tidsadverbial. I denne drøftingen har jeg særlig gått inn på hvordan denne tidsrelasjonen plasserer seg på tidslinjen.

Klokkeslettet som uttrykkes internt i et klokkeslettsuttrykk[16], kan i ut-

[16] For eksempel *halv to*.

gangspunktet ikke fikseres til noe punkt på tidslinjen i forhold til utsagnsøyeblikket. Dette fordi vi vanligvis ikke vet utsagnsøyeblikkets klokkeslett i skriftlige tekster, eller hvilket døgn det er snakk om (dette forandrer seg naturligvis i muntlige sammenhenger). Det er verbets tempus og preposisjonenes inherente egenskaper som hode for tidsadverbialet, som gjør oss i stand til å orientere klokkeslettet mot et punkt på tidslinjen. Jeg har sett på hvordan et klokkeslettsuttrykk brukes som utgangspunkt for å orientere tidsadverbialet til et punkt eller en terminativ *før/etter*-relasjon på tidslinjen, og dessuten belyst hvordan tidsrelasjonen endrer seg ut ifra hvilken preposisjon som står som hode i klokkeslettsuttrykket.

De observasjonene jeg her har gjort, har dannet grunnlaget for XLE-grammatikken jeg har skrevet. Denne grammatikken er tema for neste kapittel, hvor også syntaksen til de forskjellige typer klokkeslettsuttrykk blir drøftet mer i detalj.

Kapittel 4

Implementeringen

4.1 Innledning

Det foreligger så vidt jeg vet ingen tidligere forsøk på å implementere norske klokkeslettsuttrykk i en LFG-formalisme. Jeg har tilpasset min grammatikk etter den eksisterende XLE-koden som LOGON baserer seg på, slik at klokkeslettsuttrykkene jeg behandler enklest mulig skal kunne tilpasses NorGrams overordnede grammatikk.

4.1.1 Presentasjon av kapittelet

I avsnitt 4.2 til avsnitt 4.6 går jeg igjennom XLE-koden for grammatikken jeg har skrevet. Jeg kommer her bare til å legge vekt på c- og f-strukturrepresentasjonen til strukturene, siden det er disse representasjonene som er relevante for hvorvidt uttrykkene blir parset som grammatiske eller ikke. De originale XLE-filene ligger vedlagt i tillegg B-D. I tillegg A gis en oversikt over hvilke trekk som benyttes i XLE-grammatikken.

I avsnitt 4.7 går jeg igjennom hvordan MRS-representasjonen til klokkeslettsuttrykkene ser ut. Jeg diskuterer prinsippet for hvordan klokkeslettet representeres, og tar for meg hvordan representasjonen forandrer seg når syntaksen i setningen endres.

4.1.2 XLE-koden

XLE-koden som er vedlagt i tillegg B-D er en selvstendig minigrammatikk med et svært begrenset ordforråd. Den tar utgangspunkt i konvensjonene til NorGram og bruker også NorGrams templatfil, men en del templat legges til denne. De templatene som brukes i min grammatikk er vedlagt i tillegg D. Frasestrukturreglene jeg har definert, ligger vedlagt i tillegg B, og leksikonen står i tillegg C.

Jeg har funnet det gunstig å bruke en liten, frittstående grammatikk for å implementere de syntaktiske konstruksjonene klokkeslettsuttrykkene representerer. Etter at minigrammatikken parset de ønskede analysene av klokkeslettsuttrykkene, har jeg erstattet mange av mine forenklede frasestrukturregler med de reglene NorGram bruker, og har lagt mine endringer til disse. I stedet for å integrere min minigrammatikk i NorGram, har jeg altså til en viss grad integrert NorGrams regler i min grammatikk. Resultatet er at en fullstendig integrering i NorGram nå vil være enklere. Denne implementeringen vil bli gjort uavhengig av selve masteroppgaven.

4.1.3 Klokkeslettsuttrykket

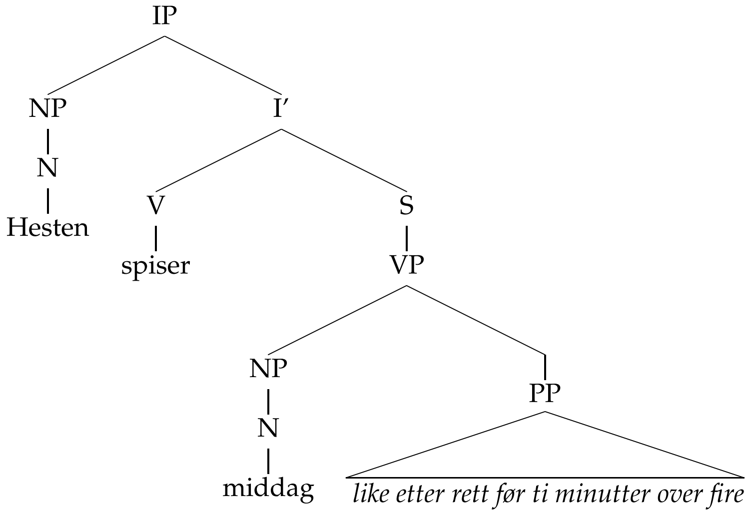
I løpet av oppgaven refereres det gjentatte ganger til begrepet *klokkeslettsuttrykket*. Dette er et begrep jeg selv definerer ut ifra pragmatiske parametre. *Klokkeslettsuttrykket* skal forstås som den forlengelsen jeg har lagd til NorGrams overordnede grammatikk, og som henger som datter under verbfrasen (VP) i c-strukturen.[1] Klokkeslettsuttrykket kan kjennes igjen ved følgende egenskaper i denne avhandlingen:

1. Klokkeslettsuttrykkets grammatiske funksjon er enten PREDLINK eller ADJUNCT.
2. Dersom et *tidsadverbial*[2] som uttrykker et klokkeslett innledes med en preposisjonsfrase (PP), vil hele tidsadverbialet regnes som et klokkeslettsuttrykk så lenge preposisjonsfrasen direkte eller gjennom rekursjon terminerer med kategorier fra TIMES-disjunksjonen[3].
3. Dersom tidsadverbialet ikke innledes med en preposisjonsfrase, vil det alltid bestå av en kategori fra disjunksjonen TIMES. Disse kategoriene er ikke-rekursive.
4. Innad i setningens f-projeksjon vil trekket CLOCK-EXPRESSION + legges til på alle nivå i et tidsadverbial som regnes som et klokkeslettsuttrykk.
5. I c-strukturen vil toppnoden i et klokkeslettsuttrykk henge direkte under VP eller IP[4] og terminere med en kategori fra TIMES-disjunksjonen.

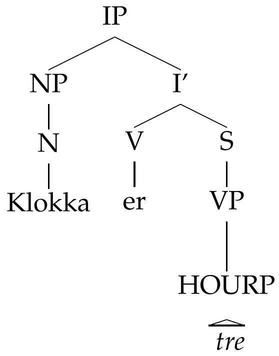
For setning (1) og (2) er tidsadverbialet som defineres som et klokkeslettsuttrykk kursivert i trestrukturen.

[1]Med mindre klokkeslettsuttrykket topikaliseres, se avsnitt 4.6.
[2]Om *tidsadverbial*, se avsnitt 3.2.1.
[3]XLE-koden for TIMES-disjunksjonen er gjengitt i regel (7).
[4]Klokkeslettsuttrykket henger under IP ved topikalisering.

(1) *Hesten spiser middag like etter rett før ti minutter over fire*



(2) *Klokka er tre*



4.1.4 En seksdeling av uttrykkene

Setningsdelene i kursiv i tabell 4.1 er setningens klokkeslettsuttrykk.

	PREDLINK	**ADJUNCT**
HOURP	Klokka er ***ni.***	Jeg kommer hjem ***halv*** *ni.*
PP-TIME	Klokka er *noen minutter* ***på*** *ni.*	Jeg kommer hjem *klokka kvart* ***på*** *ni.*
PP	Klokka er *lenge* ***etter*** *tolv.*	Jeg kommer hjem *like* ***før*** *halv ni.*

Tabell 4.1: En seksdeling av klokkeslettsuttrykkene

Klokkeslettsuttrykkene deles i tabell 4.1 inn etter fem forskjellige kriterier. Denne inndelingen henspiller på hvilken grammatisk funksjon klokkeslettsuttrykket innehar i setningens f-struktur (vannrett inndeling) og hvilken frasekategori som er toppnode i uttrykkets c-struktur (loddrett inndeling). HOURP, PP-TIME og PP er navn på frasekategorier som forekommer i c-strukturen. F-strukturens hode i disse frasene er uthevet.

Jeg tar utgangspunkt i denne tabellen når jeg går igjennom de ulike klokkeslettsuttrykkene jeg har behandlet i denne oppgaven. Over de neste delavsnittene belyser jeg hvilke syntaktiske egenskaper uttrykkene innehar på bakgrunn av hvilken grammatisk funksjon de har i setningen og hva som er toppnoden. Videre drøfter jeg hvordan jeg har representert dette gjennom XLE-koden. Jeg tar her bare hensyn til c- og f-strukturen, og diskuterer MRS-representasjonen i avsnitt 4.7.

4.2 HOURP

	PREDLINK	ADJUNCT
HOURP	**Klokka er *ni*.**	**Jeg kommer hjem *halv ni*.**
PP-TIME	Klokka er *noen minutter på ni.*	Jeg kommer hjem *klokka kvart på ni.*
PP	Klokka er *lenge etter tolv.*	Jeg kommer hjem *like før halv ni.*

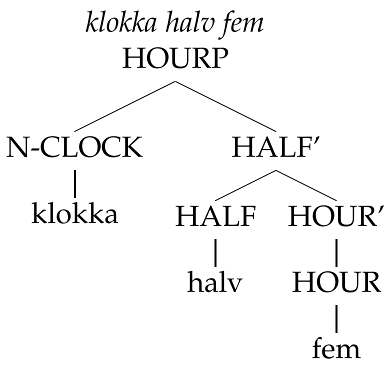
Uttrykk hvor toppnoden i c-struktur er HOURP er ukompliserte uttrykk uten preposisjon. HOURP henger direkte under VP, og frasestrukturregelen for HOURP ser slik ut:

(3) HOURP og underkategorier

```
HOURP -->
          (N-CLOCK: (^ CLOCK-OCCURRENCE)=c +_)
           HOUREXP.

HALF' --> HALF: ^=!;
       { HOUR': (^ OBJ)=!
       | e: (^ OBJ PRED)='pro'
            m::(^ OBJ)=%PREP
            @MRS-EMPTYHOUR}.

HOUR' --> HOUR.

HOUREXP = { HOUR' | HALF' }.
```

Med HOURP som toppnode, kan det mest komplekse uttrykket inneholde tre ord (4).

(4)



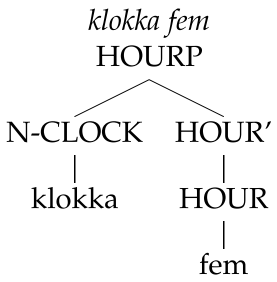
I (4) er *halv* hode i uttrykket, og f-strukturen er:

(5) F-struktur for (4):

$$(5)\ \begin{bmatrix} \text{PRED} & \text{'halv}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'hour-5'} \\ \text{CLOCK-OCCURRENCE} & +\_ \end{bmatrix} \end{bmatrix}$$

I (5) ser vi at *halv* tar den hele klokketimen som objekt. Dersom *halv* ikke forekommer i setningen, vil den hele klokketimen stå som hode i klokkeslettsuttrykket, og klokketimen vil henge direkte som datter av HOURP:

(6)



HOURP henger i c-strukturen like under VP, men i frasestrukturreglene er det to disjunksjoner som fører til HOURP. Disse er TIMEEXP og TIMES. TIMEEXP er en syntaktisk variabel som enten leder til PP eller TIMES. Disjunksjonen legger også til en rekke trekk, og kommer med restriksjoner for hvordan f-strukturen ser ut. Hvilke trekk og restriksjoner som legges til, beror på hvilken grammatisk funksjon klokkeslettsuttrykket har i setningen, og dette gjennomgås når den grammatiske funksjonen til klokkeslettsuttrykk av denne typen behandles (avsnitt 4.2.1 og 4.2.3). TIMES er en variabel som kaller enten HOURP eller PP-TIME. Denne variabelen terminerer alle klokkeslettsuttrykk, og legger til trekket CLOCK-EXPRESSION +.

(7) TIMES-disjunksjonen

```
TIMES = { PP-TIME: (! CLOCK-EXPRESSION)=+
          | HOURP: (! CLOCK-EXPRESSION)=+ }.
```

Klokka/klokken

Regelen HOURP kan ta en opsjonell forekomst av *klokka/klokken* foran hodet i frasen, dette uttrykkes ved `(N-CLOCK: (^ CLOCK-OCCURRENCE)=c +_)`. Når N-CLOCK-oppslaget appliseres (10), innføres den instansierte verdien +_ (om instansierte trekk, se side 21). *Klokka/Klokken* kan opptre som datter også i PP-TIME og i PP, men kan maksimalt forekomme én gang i klokkeslettsuttrykket. Derfor vil ikke (8) bli parset som et grammatisk klokkeslettsuttrykk:

(8) *klokka like før klokka fem over åtte

Innføringen av CLOCK-OCCURRENCE +_ kan gjøres én, og bare én gang i hele klokkeslettsuttrykket, fordi trekket siprer til alle nivå i f-strukturen. Dersom vi derimot skal definere at *klokka/klokken* aldri kan stå i klokkeslettsuttrykket, innføres trekket CLOCK-OCCURRENCE -. Dette trekket innføres blant annet av TIMEEXPs første disjunkt i regel (12). Årsaken til at denne restriksjonen introduseres i (12), er at vi ikke ønsker at en konstruksjon som (9) skal tolkes som en angivelse av klokkeslettet 8.30.[5]

(9) Klokka er klokka halv ni.

(10) Leksikonoppslaget for *klokka* og *klokken*

```
klokka    N-CLOCK * (^ CLOCK-OCCURRENCE)=+_.

klokken   N-CLOCK * (^ CLOCK-OCCURRENCE)=+_.
```

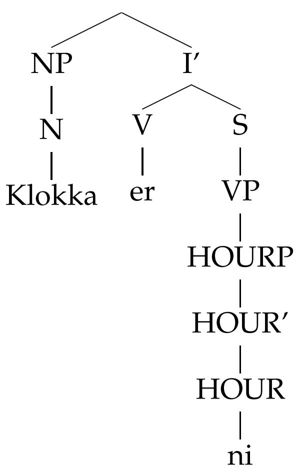
[5]Den eneste gyldige lesningen av denne setningen, er at subjektet *klokka* refererer til en hendelse, se avsnitt 4.2.2.

4.2.1 HOURP og PREDLINK

	PREDLINK	ADJUNCT
HOURP	**Klokka er *ni*.**	Jeg kommer hjem *halv ni.*
PP-TIME	Klokka er *noen minutter på ni.*	Jeg kommer hjem *klokka kvart på ni.*
PP	Klokka er *lenge etter tolv.*	Jeg kommer hjem *like før halv ni.*

Uttrykk hvor klokkeslettets grammatiske funksjon er PREDLINK og det øverste c-struktur-nivået er HOURP, kan i praksis bare generere noen få klokkeslettsuttrykk. I (11) ser vi hvordan c- og f-strukturen for *Klokka er ni* ser ut:

(11) *Klokka er ni*



(11)

$$\left[\begin{array}{ll} \text{PRED} & \text{'være}\left\langle\left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{PREDLINK}\right)\right\rangle\text{'} \\ \text{SUBJ} & \left[\text{PRED} \quad \text{'klokke'}\right] \\ \text{PREDLINK} & \left[\begin{array}{ll} \text{PRED} & \text{'hour-9'} \\ \text{CLOCK-PREDLINK} & + \\ \text{CLOCK-EXPRESSION} & + \\ \text{CLOCK-OCCURRENCE} & - \end{array}\right] \end{array}\right]$$

Klokkeslettsuttrykkets toppnode, HOURP, ble diskutert i avsnitt 4.2. Der så vi på HOURPs egenskaper isolert fra den øvrige setningen. Når en hel setning parses, vil imidlertid flere trekk legges til i f-strukturen ut ifra hvilken grammatisk funksjon klokkeslettet har, og hva subjektet i setningen refererer til. Disse trekkene legges til av disjunksjonene TIMES og TIMEEXP, som også sørger for at nødvendige restriksjoner som ivaretar grammatikaliteten blir lagt til.

(12) TIMES-variabelen og utdrag fra TIMEEXP-disjunksjonen

```
TIMEEXP =  (...)

            TIMES: {{(...)
                   (^ PREDLINK)=!
                   {   (^ SUBJ PRED FN)=c klokke
                       (^ SUBJ DEF)=c +
                   |   (^ SUBJ PRON-FORM)=c den
                       (^ SUBJ GEND NEUT)=c -}
                     (! CLOCK-PREDLINK)=+
                     (! CLOCK-OCCURRENCE)= -

                    | (^ PREDLINK)=!
                      (^ SUBJ CLOCK-EVENT)=+
                      { ~(^ SUBJ PRED FN)= klokke
                      |(^ SUBJ PRED FN)= klokke
                       (! CLOCK-OCCURRENCE)=c +_}}}.


TIMES = { PP-TIME: (! CLOCK-EXPRESSION)=+
         | HOURP: (! CLOCK-EXPRESSION)=+ }.
```

I TIMEEXP-disjunksjonen er det to ulike disjunkt som leder til HOURP og angir den grammatiske funksjonen PREDLINK. I det første disjunktet ser vi at det stilles et krav til subjektet om at dette enten må være *klokka/klokken* eller *den*. Hele uttrykket gjenspeiler den konvensjonelle måten for å uttrykke hvor mye klokka er. I disse tilfellene legges trekkene CLOCK-PREDLINK + og CLOCK-OCCURRENCE - til i f-strukturen som vist i (11). CLOCK-OCCURRENCE - er en restriksjon som hindrer at *klokka/klokken* opptrer innad i klokkeslettsuttrykket, som vist i eksempel (9). Trekket CLOCK-EXPRESSION + tilordnes under TIMES-variabelen.

Det andre relevante disjunktet legger ikke noen trekk til i f-strukturen til PREDLINK, men legger derimot CLOCK-EVENT + til i subjektets f-struktur. Analyser under dette disjunktet representerer tilfeller hvor subjektet i setningen beskriver en hendelse, og denne hendelsen tidfestes av klokkeslettsuttrykket. Et typisk eksempel på dette ser vi i (13).

(13) *Kampen er halv ni*

$$(13)\quad \begin{bmatrix} \text{PRED} & \text{'være}\left\langle\left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{PREDLINK}\right)\right\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'kamp'} \\ \text{CLOCK-EVENT} & + \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{'halv}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED 'hour-9'} \end{bmatrix} \\ \text{CLOCK-EXPRESSION} & + \end{bmatrix} \end{bmatrix}$$

Internt i dette andre disjunktet i (12) settes det restriksjoner knyttet til setningens subjekt, og disse restriksjonene drøftes i neste avsnitt.

4.2.2 Subjekt som hendelse

Dersom et klokkeslettsuttrykk tilordnes funksjonen PREDLINK forventes to potensielle subjektstyper. I det første tilfellet er subjektet et symbol for tid, eventuelt klokketid (Vannebo, 1979, s. 79), og hele setningen uttrykker den konvensjonelle måten å uttrykke et klokkeslett på. Denne analysen korresponderer med det første disjunktet til XLE-regelen for TIMEEXP i (12).

Dersom setningen ikke uttrykker et klokkeslett, men PREDLINK likevel er et klokkeslettsuttrykk, forventes det at subjektet refererer til noe som kan tidfestes. Dette er et krav setning (13) tydelig tilfredstiller. I utgangspunktet skulle man kanskje tro at en automatisk parser kunne ivareta dette kravet dersom alle nomen som kan la seg tidfeste, på forhånd tagges i leksikon. Dette lar seg dessverre ikke gjøre, fordi alle nomen kan gis en kontekst hvor de kan referere til noe som lar seg tidfeste. I (14) får et særnavn denne egenskapen:

(14) - Det stod i programmet at foredraget til Hans Blix var klokka ni, men i avisen står det at Muhammed ElBaradei er klokka ni.

For en parser er denne ubestemmeligheten et stort problem. I og med at alle subjekt må tolkes som potensielle referenter til en hendelse, ville dette i min grammatikk medført at alle uttrykk som kunne analyseres fra det første disjunktet til TIMEEXP i (12), også ville fått en analyse i det andre disjunktet. Med andre ord ville uttrykket *Klokka er halv fem* fått to analyser, i og med at *klokka* teoretisk sett kan referere til en hendelse. En kontekst hvor denne lesningen ikke er helt utenkelig er:

(15) Auksjonen på bildet er klokka tre, sofaen er halv fire, og klokka er halv fem.

Subjektsrestriksjonen jeg har satt inn i regel (12) tillater ikke denne lesningen. Dersom subjektet *klokka/klokken* skal vise til en hendelse, må trekket CLOCK-OCCURRENCE +_ være til stede i f-strukturen til PREDLINK. Utsnittet av TIMEEXP-disjunktet demonstrerer dette:

(16) Subjektsrestriksjonen til TIMEEXP-disjunktet dersom subjektet viser til en hendelse

```
(^ SUBJ CLOCK-EVENT)=+
{ ~(^ SUBJ PRED FN)= klokke
 |(^ SUBJ PRED FN)= klokke
 (! CLOCK-OCCURRENCE)=c +_}
```

Det denne restriksjonen gjør, er å definere at subjektet må vise til en hendelse (ved `(^ SUBJ CLOCK-EVENT)=+`). Dernest kan ikke subjektet være *klokke* (ivaretatt av `~(^ SUBJ PRED FN)= klokke`), eller så er subjektet *klokke* dersom `CLOCK-OCCURRENCE +_` er instansiert. I praksis betyr dette at subjektet *klokka* bare leses som en hendelse i en setning som

(17) Klokka er klokka halv tre

siden trekket CLOCK-OCCURRENCE +_ her instansieres i f-strukturen til klokkeslettsuttrykket. Dette medfører ikke tvetydighet siden (17) ikke vil få en løsning under det første TIMEEXP-disjunktet i (12), i og med at dette disjunktet krever at trekket CLOCK-OCCURRENCE +_ ikke er til stede. Intuisjonen er at i en ytring hvor *klokka/klokken* er ment å referere til en hendelse, vil det være vanligst å markere dette ved å skyte inn *klokka/klokken* på nytt i PREDLINK, slik at det ikke vil være mulig å tolke hele setningen som en referanse til klokkeslettet. Løsningen jeg har valgt er utvilsomt diskutabel rent lingvistisk, men i praksis vil dette være tjenlig for de aller fleste applikasjoner. Dette gjør at setningene

(18) Klokka er halv fem

og

(19) Klokka er klokka fem

bare får en løsning hver. Imidlertid vil setningen

(20) Den er halv åtte

få to løsninger, siden det her er umulig for parseren å avgjøre hva subjektet viser til.

4.2.3 HOURP og ADJUNCT

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	**Jeg kommer hjem *halv ni.***
PP-TIME	Klokka er *noen minutter på ni.*	Jeg kommer hjem *klokka kvart på ni.*
PP	Klokka er *lenge etter tolv.*	Jeg kommer hjem *like før halv ni.*

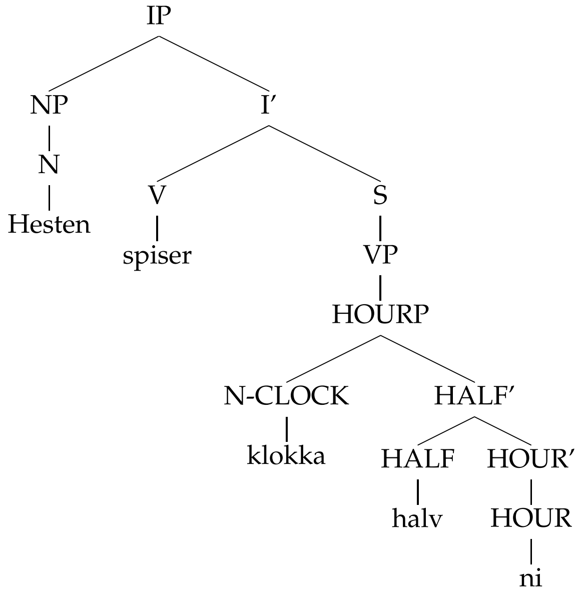
Klokkeslettsuttrykk med den grammatiske funksjonen ADJUNCT definerer et klokkeslett som tidfester en hendelse som introduseres i setningen. XLE-koden til HOURP er gjengitt i (3). HOURP vil henge under VP eller IP, og blir kalt av disjunksjonene TIMEEXP og TIMES. Den grammatiske funksjonen blir tilordnet av TIMEEXP-disjunksjonen:

(21) Utsnitt av TIMEEXP-disjunksjonen

```
TIMEEXP = {
            (...)
            | TIMES: { !$(^ ADJUNCT)
                        @MRS-CLOCKADJUNCT
            (...)}}.
```

Egenskapene til HOURP ble gjennomgått i avsnitt 4.2. Det stilles ingen ytre restriksjoner verken til eller fra andre setningsdeler til hvordan klokkeslettsuttrykket ser ut. C- og f-strukturen for *Hesten spiser klokka halv ni* demonstreres i (22).

(22) *Hesten spiser klokka halv ni*



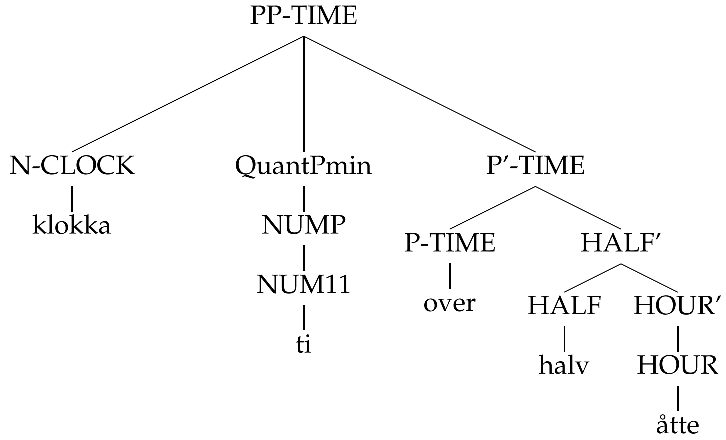
(22)
$$\left[\begin{array}{ll} \text{PRED} & \text{'spise}\left\langle\left(\uparrow \text{SUBJ}\right)\right\rangle\text{'} \\ \text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{'hest'}\end{array}\right] \\ \text{ADJUNCT} & \left\{\left[\begin{array}{ll} \text{PRED} & \text{'halv}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\ \text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{'hour-9'}\end{array}\right] \\ \text{CLOCK-EXPRESSION} & + \\ \text{CLOCK-OCCURRENCE} & +\_ \end{array}\right]\right\} \end{array}\right]$$

4.3 PP-TIME

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	Jeg kommer hjem *halv ni.*
PP-TIME	**Klokka er *noen minutter på ni.***	**Jeg kommer hjem *klokka kvart på ni.***
PP	Klokka er *lenge etter tolv.*	Jeg kommer hjem *like før halv ni.*

PP-TIME beskriver frasen mange intuitivt tenker på når de tenker på et klokkeslettsuttrykk med en preposisjon. Dette er en preposisjonsfrase der enten *over* eller *på* står som hode. Det er sterke restriksjoner for hvordan denne preposisjonsfrasen kan se ut i forhold til vanlige preposisjonsfraser (PP). Vi ser hvordan et klokkeslettsuttrykk isolert fra konteksten blir representert i c- og f-struktur i (23).

(23) *klokka ti over halv åtte*



(23)
$$\begin{bmatrix} \text{PRED} & \text{'over}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'halv}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix}\text{PRED 'halv'}\end{bmatrix} \end{bmatrix} \\ \text{SPEC} & \begin{bmatrix} \text{SPEC} \begin{bmatrix} \text{NUMBER} \begin{bmatrix}\text{PRED 'ti'}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{PTYPE-CLOCK} & \begin{bmatrix}\text{TEMPFINALE +}\end{bmatrix} \\ \text{CLOCK-OCCURRENCE} & +\_ \end{bmatrix}$$

Frasestrukturreglene for PP-TIME ser slik ut:

(24) PP-TIME og underkategorier

```
PP-TIME -->
           (N: (^ CLOCK-OCCURRENCE)=c +_)
            TIME-MOD: (^ SPEC)=!
            @MRS-MINUTEMEASURE
           P'-TIME: ^=!
                   (^ PTYPE-CLOCK TEMPFINALE)=c +.

TIME-MOD = { QuantPmin: { (! PRED FN)=c minutt
                        | (! PRED FN)=c sekund
                        | (! PRED FN)=c pro }
          | QUART
          | N: { (! PRED FN)=c minutt
               | (! PRED FN)=c sekund }
```

```
               m::!=%OBJECT
               @(MRS-COVQUANT _få_q 3 pl) }.

P'-TIME -->
           P-TIME: ^=!
                   (! PTYPE-CLOCK TEMPFINALE)=c +;
           { (HOUREXP: (^ OBJ)=!)
            |e: (^ OBJ PRED)='pro'
                m::(^ OBJ)=%PREP
                @MRS-EMPTYHOUR}.

 QuantPmin --> { { Q: ^=!
                        (^ CHECK _PREDET)=+
            | NUMdig: ^=!
                      (^ CHECK _PREDET)=+
             | NUMP: (^ SPEC NUMBER)=!
                     @LITNUM
                     (^ CHECK _PREDET)=+ }
            N: ~(^ DEF)=+
          | { NUMdig: ^=!
                       (^ PRED)='pro'
                       "</mrs>" @(MRS-OBJECTGVN minutt)"<mrs/>"
              | NUMP: (^ PRED)='pro'
                      (^ SPEC NUMBER)=!
                      @LITNUM
                      (s::M* arg1 ind given)=+
                      "</mrs>"@(MRS-OBJECTGVN minutt)"<mrs/>"}}.
```

Foran preposisjonen kan kun fraser fra TIME-MOD-disjunksjonen stå. Disse blir i f-strukturen spesifikator til hodet i klokkeslettsuttrykket. Dersom dette er en kvantorfrase, må nomenet være *minutt* eller *sekund*. Nomenet kan også være undertrykt og få en analyse gjennom QuantPmin, men dette fordrer at kvantoren i QuantPmin er et tallord. Underforstått blir da nomenet *minutter* fortrengt, som i uttrykket *fem på åtte*. Andre spesifikatorer som kan stå foran preposisjonen i PP-TIME er *kvart* eller *minutter/sekunder*, som i eksempel (25) og (26).

(25) kvart på ni

(26) sekunder over halv åtte

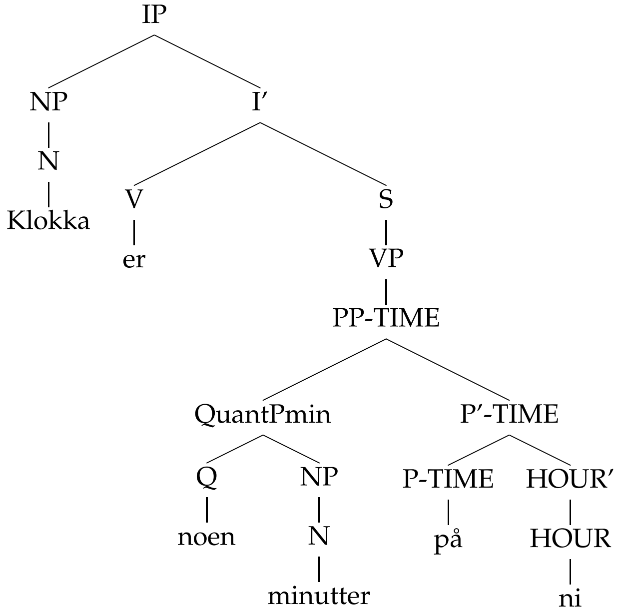
Restriksjonen PTYPE-CLOCK TEMPFINALE + i PP-TIME-regelen ivaretar at bare *over* og *på* kan stå som hode i denne frasen. Vi kan også legge merke til at uttrykk som *like over åtte* i min grammatikk ikke analyseres som en PP-TIME. Her anvendes et annet leksikonoppslag av *over*, og klokkeslettsuttrykket analyseres under PP. *Over*s egenskaper som hode i PP-TIME og PP blir drøftet i kapittel 3, side 41.

4.3.1 PP-TIME og PREDLINK

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	Jeg kommer hjem *halv ni.*
PP-TIME	**Klokka er *noen minutter på ni.***	Jeg kommer hjem *klokka kvart på ni.*
PP	Klokka er *lenge etter tolv.*	Jeg kommer hjem *like før halv ni.*

C- og f-strukturen til *Klokka er noen minutter på ni* ser slik ut:

(27) *Klokka er noen minutter på ni*



(27)
$$
\left[\begin{array}{ll}
\text{PRED} & \text{'være}\left\langle\left(\uparrow \text{SUBJ}\right)\left(\uparrow \text{PREDLINK}\right)\right\rangle\text{'} \\
\text{SUBJ} & \left[\text{PRED}\quad\text{'klokke'}\right] \\
\text{PREDLINK} & \left[\begin{array}{ll}
\text{PRED} & \text{'på}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\
\text{OBJ} & \left[\text{PRED 'hour-9'}\right] \\
\text{PTYPE-CLOCK} & \left[\text{TEMPFINALE} +\right] \\
\text{SPEC} & \left[\begin{array}{l}
\text{PRED 'minutt'} \\
\text{NTYPE}\left[\text{CLOCK-MODIFIER} +\right] \\
\left[\left[\text{SPEC}\left[\text{QUANT}\left[\text{PRED 'noen'}\right]\right]\right]\right]
\end{array}\right] \\
\text{CLOCK-EXPRESSION} & + \\
\text{CLOCK-OCCURRENCE} & - \\
\text{CLOCK-PREDLINK} & +
\end{array}\right]
\end{array}\right]
$$

I forhold til de grammatiske funksjonene, gjelder de samme restriksjonene for uttrykk med PP-TIME som toppnode, som for dem som hadde HOURP som toppnode (se avsnitt 4.2.1). Trekk og restriksjoner i f-strukturen bygges opp av frasestrukturreglene i (12), og de samme trekkene legges til uavhengig av hva toppnoden i c-strukturen er. Avsnitt 4.2.2s drøfting av når subjektet skal tolkes som en hendelse, har samme gyldighet for uttrykk med PP-TIME som toppnode. Dette betyr at uttrykkene

(28) Kampen er ti over åtte

og

(29) Klokka er ti over åtte

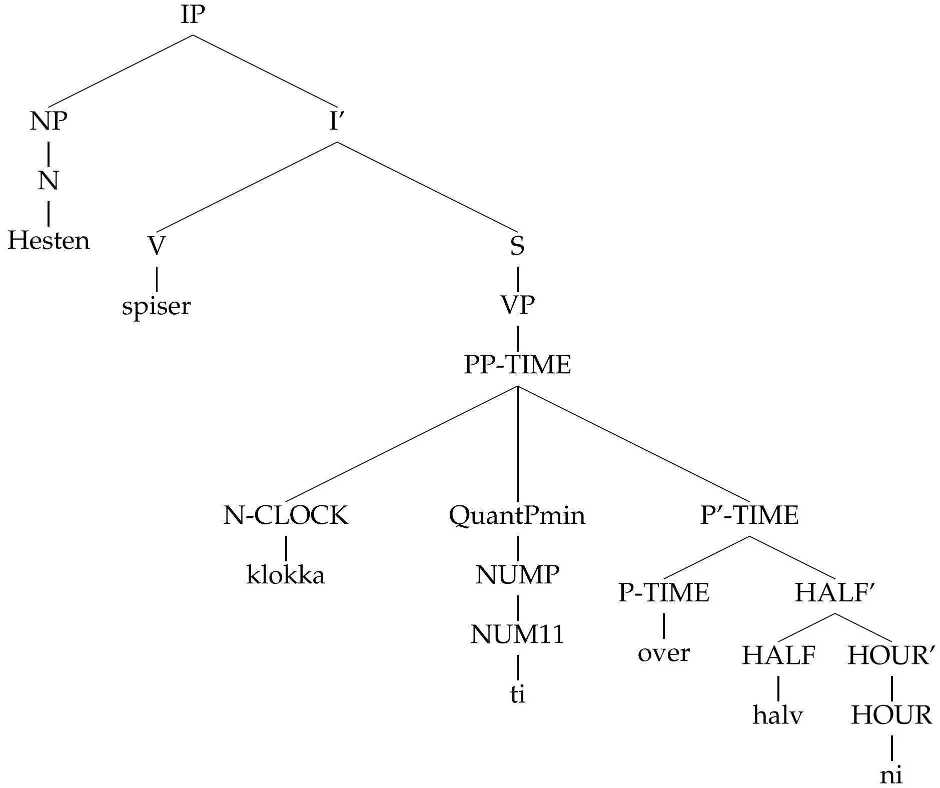
får forskjellige trekkstrukturer, siden subjektet til eksempel (28) regnes som en referanse til en hendelse, og trekket CLOCK-EVENT + legges til i subjektets f-struktur.

4.3.2 PP-TIME og ADJUNCT

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	Jeg kommer hjem *halv ni.*
PP-TIME	Klokka er *noen minutter på ni.*	**Jeg kommer hjem *klokka kvart på ni.***
PP	Klokka er *lenge etter tolv.*	Jeg kommer hjem *like før halv ni.*

Det legges ingen videre restriksjoner i f-strukturen på klokkeslettsuttrykkene med den grammatiske funksjonen ADJUNCT. Den relevante XLE-regelen for tilordningen av den grammatiske funksjonen er vist i (21). Utover dette følges PP-TIME-regelen som ble gjennomgått i avsnitt 4.3. C- og f-strukturen for uttrykket *Hesten spiser klokka ti over halv ni* får denne representasjonen:

(30) *Hesten spiser klokka ti over halv ni*



$$
(30)\quad \left[\begin{array}{ll}
\text{PRED} & \text{'spise}\left\langle(\uparrow \text{SUBJ})\right\rangle\text{'} \\
\text{SUBJ} & [\text{PRED 'hest'}] \\
\text{ADJUNCT} & \left\{ \left[\begin{array}{ll}
\text{PRED} & \text{'over}\left\langle(\uparrow\text{OBJ})\right\rangle\text{'}\\
\text{OBJ} & \left[\begin{array}{l} \text{PRED 'halv}\left\langle(\uparrow\text{OBJ})\right\rangle\text{'} \\ \text{OBJ } [\text{PRED 'hour-9'}]\end{array}\right]\\
\text{SPEC} & \left[\begin{array}{l}\text{PRED 'pro'}\\ \left[\text{SPEC }\left[\text{NUMBER }[\text{PRED 'ti'}]\right]\right]\end{array}\right]\\
\text{PTYPE-CLOCK} & [\text{TEMPFINALE} +]\\
\text{CLOCK-EXPRESSION} & +\\
\text{CLOCK-OCCURRENCE} & +\_
\end{array}\right] \right\}
\end{array}\right]
$$

Vi kan legge merke til at *klokka* henger direkte under PP-TIME i c-strukturen i (30). Denne forekomsten tillates i frasestrukturregelen for PP-TIME, se regel (24). For uttrykkene i den grammatiske funksjonen ADJUNCT foreligger ingen ytre restriksjoner i forhold til forekomsten av dette ordet.

4.4 PP

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	Jeg kommer hjem *halv ni.*
PP-TIME	Klokka er *noen minutter på ni.*	Jeg kommer hjem *klokka kvart på ni.*
PP	**Klokka er *lenge etter tolv.***	**Jeg kommer hjem *like før halv ni.***

I klokkeslettsuttrykk har *på* og *over* andre egenskaper enn disse preposisjonene har i vanlige PP-fraser. Dette gjør at jeg har skilt ut en egen spesialfrase for klokkeslettsuttrykk der disse preposisjonene er involvert, og denne har jeg kalt PP-TIME. De leksikalske oppslagene for *på* og *over* har egne sett av trekk og MRS-representasjoner når de står innenfor denne frasen, og de representerer ordklassen P-TIME i stedet for P. I disse tilfellene kan vi si at PP-TIME-reglene utgjør et idiom-skjelett for hvordan syntaksen kan forandre seg, samtidig som PP-TIME beholder sin idiomatiske karakter.

For de andre preposisjonene jeg har behandlet finnes det ikke lignende bevis for idiomatisering. Bare *over* og *på* kan ta nye egenskaper dersom de er hode i et klokkeslettsuttrykk. De andre preposisjonene beholder sin grunnleggende semantiske betydning også dersom dets objekt er et klokkeslett. Jeg har derfor ikke sett noen grunn til å skille de andre preposisjonene ut fra den normale PP-frasen til NorGram når jeg har behandlet disse preposisjonene.

Preposisjonen *over* kan stå som hode i klokkeslettsuttrykk både i en PP- og en PP-TIME-frase. Dette gjør at en del setninger vil bli parset med to analyser, hvor det ikke er avgjort hvorvidt *over* skal være hode i PP eller PP-TIME. Dette fenomenet blir drøftet i avsnitt 4.4.2.

I avsnitt 3.3 drøftet jeg hvilke syntaktiske restriksjoner preposisjonene stiller dersom de står som hode i et klokkeslettsuttrykk. Alle restriksjonene fra tabell 3.1 er inkorporert i minigrammatikken min. Som jeg diskuterte i avsnittet hvor jeg behandlet norske adverbialstyper (avsnitt 3.2.2), er klokkeslettsuttrykkene enten frie eller bundne adverbial. I min grammatikk kan disse kategoriene bare henge under VP eller IP[6], en restriksjon som er typisk for adverbialer.

4.4.1 Generelle regler for PP

PP-regelen jeg bruker i min grammatikk er en versjon av NorGrams PP-regel. For denne regelen har jeg lagt til de nødvendige reglene som gjør

[6] Ved topikalisering.

at preposisjonsfrasen er i stand til å parse klokkeslettsuttrykk med PP som toppnode. En applisering av disse reglene fordrer imidlertid at trekket CLOCK-EXPRESSION + instansieres. Slik ser mine tillegg i PP-regelen ut:

(31) En strippet versjon av PP

```
PP --> {
         (N-CLOCK: (^ CLOCK-EXPRESSION)=c +
                   (^ CLOCK-OCCURRENCE)=c +_
                   (^ ADJUNCT TEMP-SPEC)=c +)
         (PP-TIME-MOD: (^ CLOCK-EXPRESSION)=c +_
                       ~(^ PTYPE-CLOCK SUBTEMP)= +
                        (^ PTYPE-CLOCK TEMP)=c +
                        (! TEMP-SPEC)=+)

                (...)
           }
           { P: ^=!
               (^ CLOCK-EXPRESSION)=(^ OBJ CLOCK-EXPRESSION)
               (^ PTYPE)=c sem
               @OBJCASE
               s::M*=s::*
               s::M*=s::^
              (s::M* arg2)=s::(^ OBJ)
               @(COMPLETE (^ OBJ CHECK _PREPEXISTS))
               @(COMPLETE (^ OBJ CHECK _P-SELFORM))
               @(COMPLETE (^ PSEM))
               @(COMPLETE (^ PTYPE))
             | (...) }
             {  NP: (^ OBJ)=!
                @COMPLETECHECKFEATS
                ~(! CHECK _P-SELFORM)
                { ~(! PRON-TYPE)=refl
                | (! PRON-TYPE)=c refl
                  (^ CHECK _REFL)=c +}
                { ~(! PRON-TYPE)=refrefl
                | (! PRON-TYPE)=c refrefl
                  @PPREFREFLEXIVE }
                  @REFERENTIAL?
                 @PPPERCOLATEANT

                (...)
```

```
          | PP-CLOCK: (^ OBJ)=!
                      (^ CLOCK-EXPRESSION)=c +
                      (^ PTYPE-CLOCK TEMP)=c +
                      (^ CLOCK-OCCURRENCE)=(^ OBJ CLOCK-OCCURRENCE)}
            (...).
```

Spesifikatorene til PP

I et klokkeslettsuttrykk kan PP ha to døtre på spesifikatorplass før hodet i frasen, disse er N-CLOCK og en kategori fra PP-TIME-MOD-variabelen. Forekomsten *klokka/klokken* kan stå i denne posisjonen, men dette fordrer at variabelen PP-TIME-MOD anvendes. Dette kan være en setning som:

(32) Banen åpnet igjen klokka like før halv ni. [7]

Innføringen av *klokka/klokken* instansierer trekket CLOCK-OCCURRENCE +_. Dette blir derfor en restriksjon som sørger for at uttrykket *Klokka er klokka like etter ti* bare kan tolkes som en tidfestelse av en hendelse hvor subjektet *Klokka* står som referanse til denne hendelsen. Restriksjonen hindrer også at *klokka/klokken* forekommer to ganger internt i klokkeslettsuttrykket (33).

(33) *Vi spiser middag klokka like før klokka halv ni.

PP-TIME-MOD definerer de temporale modifikasjonene som er grammatiske dersom PP er toppnode i et klokkeslettsuttrykk. Dersom PP-TIME-MOD-variabelen appliseres, kreves det at hodet i PP har trekket PTYPE-CLOCK TEMP + i f-strukturen, og at trekket PTYPE-CLOCK SUBTEMP + ikke forekommer. Den siste restriksjonen hindrer at preposisjonene *innen, siden, fra* og *til* får en spesifikator, da dette er ugrammatisk (dette blir også diskutert i kapittel 3 fra side 44 og utover). Disse restriksjonene ivaretas i PP-regelen (31). De ordklassene som kan stå som spesifikator, defineres i PP-TIME-MOD:

(34) PP-TIME-MOD-variabelen

```
PP-TIME-MOD =
```

[7] Hentet fra TV2 Nettavisen, http://pub.tv2.no/nettavisen/innenriks/article247686.ece, 09.12.2005

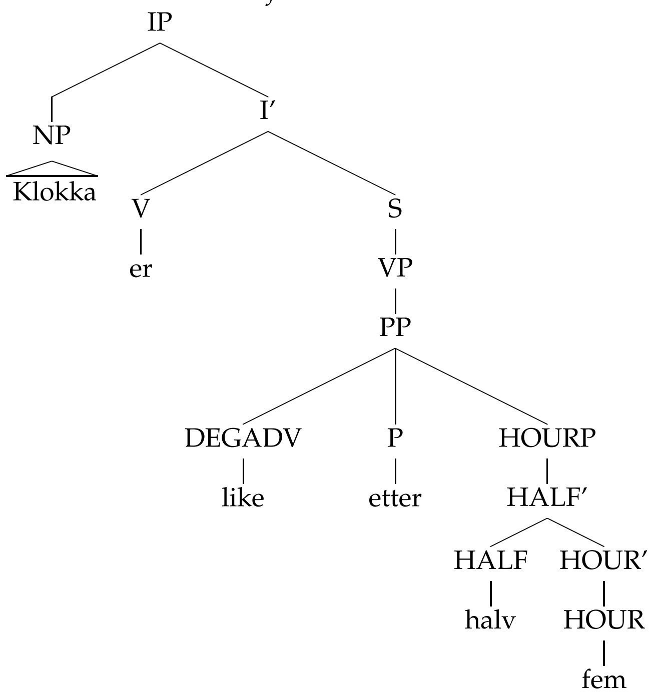
```
{ DEGADV: !$(^ ADJUNCT)
          "</mrs>" @MRS-DEGADVSYNT "<mrs/>"
          (! ATYPE CLOCK)=c +
| QuantP: !$(^ ADJUNCT)
          @MRS-MINUTEMEASURE
          (! NTYPE CLOCK-MODIFIER)=c +
| NP: !$(^ ADJUNCT)
      @MRS-MINUTEMEASURE
      (! NUM)=c pl
      (! CHECK _DEF-MORPH)=c -
      (! NTYPE CLOCK-MODIFIER)=c + }.
```

PP-TIME-MOD er preposisjonsfrasens motstykke til TIME-MOD, som er spesifikator-variabelen for PP-TIME-frasen (se avsnitt 4.3). De tre ordklassene som kan stå foran hodet i PP er gradsadverbial, kvantorfraser og nomenfraser.

Gradsadverbialet

Visse typer gradsadverbial kan potensielt stå som spesifikator til PP. Disse adverbialene må ha trekket ATYPE CLOCK = + i leksikon for å markere at de kan opptre som spesifikator i klokkeslettsuttrykkene, da det ikke er alle typer gradsadverbial som har muligheten til semantisk å modifisere et klokkeslettsuttrykk. Typiske gradsadverbial i denne gruppa er *like, rett* og *straks*. (35) viser en typisk forekomst hvor gradsadverbialet står som spesifikator foran hodet i PP.

(35) *Klokka er like etter halv fem*

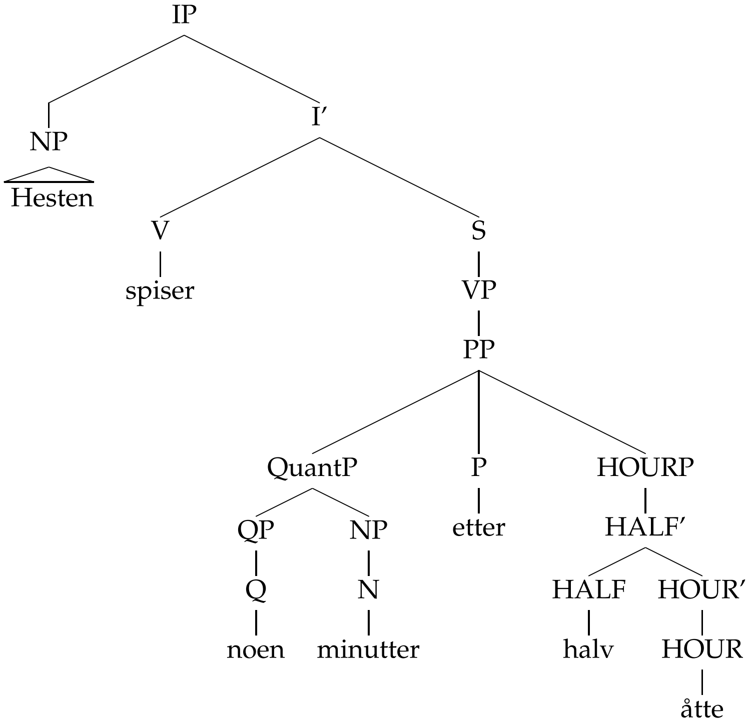


Kvantorfrasen og nomenfrasen

En kvantorfrase eller en nomenfrase som står foran hodet i PP må ha et hode som har trekket NTYPE CLOCK-MODIFIER = + i leksikon. Dette trekket er tagget på nomen som alene eller sammen med en kvantor semantisk kan modifisere et klokkeslettsuttrykk. Typiske eksempel på slike forekomster ser vi i (36) (NP) og (37) (QuantP).

(36) Hesten spiser sekunder etter halv åtte.

(37) *Hesten spiser noen minutter etter halv åtte*



For nomenfrasen må nomenet stå i flertall ubestemt.

Terminering eller rekursjon

Variabelen PP-CLOCK fører til preposisjonsfrasenes objekt dersom PP er et klokkeslettsuttrykk. I PP-regelen (se (31)) stilles følgende restriksjoner til denne disjunksjonen:

(38) Utsnitt av PP-regelen der PP-CLOCK kalles

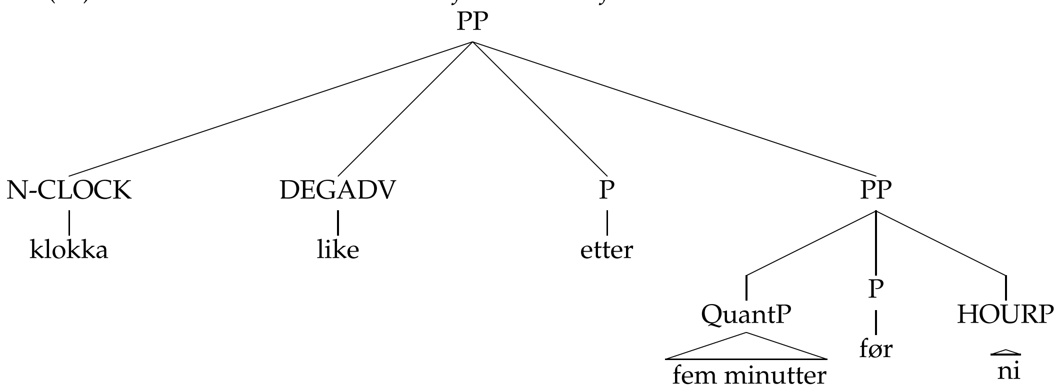
```
(PP-CLOCK: (^ CLOCK-EXPRESSION)=c +
           (^ PTYPE TEMP)=c +
           (^ CLOCK-OCCURRENCE)=(^ OBJ CLOCK-OCCURRENCE))
```

Disjunksjonen PP-CLOCK blir alltid anvendt dersom preposisjonsfrasen er del av et klokkeslettsuttrykk. De to første restriksjonene krever at preposisjonsfrasen er et klokkeslettsuttrykk[8], og at hodet i PP er tagget med trekket PTYPE TEMP +. Den siste restriksjonen sørger for at det eventuelle trekket CLOCK-OCCURRENCE +_ smitter mellom f-struktur-nivåene i PP. Dersom det forekommer i det ytre nivået, vil det også bli instansiert i objektet til hodet og omvendt. Dette er viktig siden trekket bare kan instansieres en gang i løpet av klokkeslettsuttrykket, det skal derfor stå på alle nivå i klokkeslettsuttrykkets f-struktur dersom det blir instansiert. PP-CLOCK er bare en enkel variabel som definerer om klokkeslettsuttrykket er rekursivt (som i (40)), eller om det terminerer umiddelbart (41).

(39) PP-CLOCK

```
PP-CLOCK = {PP | TIMES }.
```

(40) *klokka like etter fem minutter før ni*



[8]Restriksjonenskravet blir oppfylt når PP-CLOCK på dette nivå eller ved rekursjon velger TIMES-variabelen som objekt til PP.

(40)
$$
\left[\begin{array}{ll}
\text{PRED} & \text{'etter}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\
\text{OBJ} & \left[\begin{array}{l}
\text{PRED 'før}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\
\text{OBJ} \left[\begin{array}{ll} \text{PRED} & \text{'hour-9'} \\ \text{CLOCK-EXPRESSION} & + \\ \text{CLOCK-OCCURRENCE} & +\_ \end{array}\right] \\
\text{ADJUNCT} \left\{\left[\begin{array}{ll} \text{PRED} & \text{'minutt'} \\ \text{NTYPE} & \left[\text{CLOCK-MODIFIER} +\right] \\ \text{SPEC} & \left[\text{NUMBER} \left[\text{PRED 'fem'}\right]\right] \\ \text{TEMP-SPEC} & + \end{array}\right]\right\} \\
\text{PTYPE-CLOCK} \left[\text{TEMP} +\right] \\
\text{CLOCK-EXPRESSION} + \\
\text{CLOCK-OCCURRENCE} +\_
\end{array}\right] \\
\text{ADJUNCT} & \left\{\left[\begin{array}{ll} \text{PRED} & \text{'like'} \\ \text{ATYPE} & \left[\text{CLOCK} +\right] \\ \text{TEMP-SPEC} & + \end{array}\right]\right\} \\
\text{PTYPE-CLOCK} & \left[\text{TEMP} +\right] \\
\text{CLOCK-EXPRESSION} & + \\
\text{CLOCK-OCCURENCE} & +\_
\end{array}\right]
$$

(41) Klokka er rett etter åtte.

I f-strukturen til (40) ser vi at trekkene med attributtene CLOCK-EXPRESSION og CLOCK-OCCURRENCE står på flere nivå i f-strukturen til klokkeslettsuttrykket. Disse trekkene vil alltid stå på samme nivå som hodet i PP dersom PP er del av et klokkeslettsuttrykk. Ved rekursjon siprer trekkene enten innover fra hode-objekt, eller utover fra objekt-hode. Trekket CLOCK-EXPRESSION markerer: "Dette nivå i f-strukturen er et klokkeslettsuttrykk." Det er derfor avgjørende at dette trekket siprer helt opp til ytterste ledd i f-strukturen, og til det nivået hvor hodet i hele klokkeslettsuttrykket står. Variabelen som innfører dette trekket, terminerer også rekursjonen for klokkeslettsuttrykk i PP (42).

(42) TIMES-disjunksjonen

```
TIMES = { PP-TIME: (! CLOCK-EXPRESSION)=+
          | HOURP: (! CLOCK-EXPRESSION)=+ }.
```

Alle klokkeslettsuttrykk med PP som toppnode vil terminere med at denne variabelen velges av PP-CLOCK (se regel (39)). Når et av disse disjunktene velges vil objektet til PP bli enten PP-TIME eller HOURP. Egenskapene til PP-TIME er gjennomgått i avsnitt 4.3, og HOURPs egenskaper er gjennomgått i avsnitt 4.2.1 og avsnitt 4.2.3.

4.4.2 Doble analyser for *over*

Som tidligere nevnt kan preposisjonen *over* stå som hode både i frasen PP-TIME og i klokkeslettsuttrykk i PP. Bakgrunnen for dette drøftes i avsnittet *Forskjellene mellom over og på*, side 41. I korthet går dette ut på at *over*s grunnleggende betydning i en vanlig preposisjonsfrase kan være *senere enn/mer enn*, i tillegg til at *over* kan brukes i mer idiomatiske uttrykk internt i PP-TIME (for eksempel i uttrykket *ti over åtte*). Dette gjør at leksikonoppslaget til *over* er relativt komplekst:

(43) Leksikonoppslaget til *over*

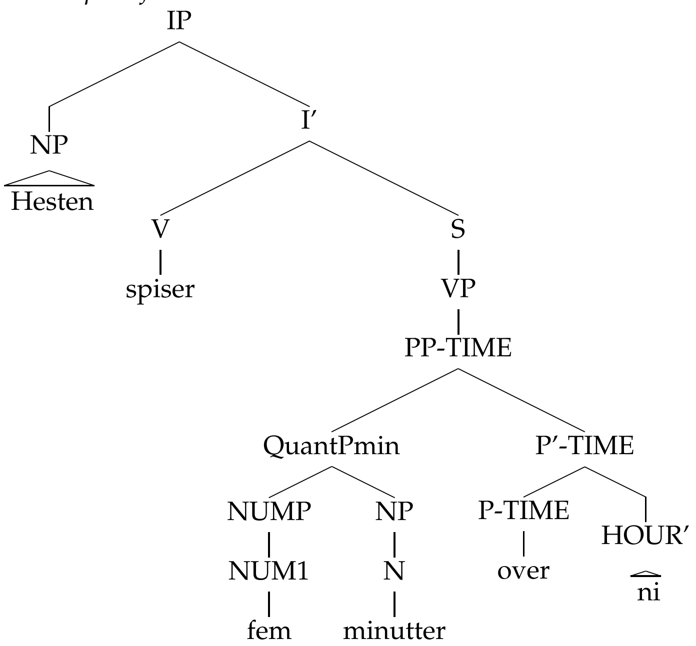
```
over      P-TIME * (^ PRED)='over<(^ OBJ)>'
                   (^ PTYPE-CLOCK TEMPFINALE)=+
                   @MRS-PAST;
          P * @(SEMPREP over over)
              { (^ PTYPE-CLOCK TEMP)=+
                (^ CLOCK-EXPRESSION)=c +
                { (^ CLOCK-PREDLINK)=c +
                |
                  (^ ADJUNCT TEMP-SPEC)=c +}
               |
                ~(^ CLOCK-EXPRESSION)=+}.
```

Problemet med denne løsningen er at en del frekvente klokkeslettsuttrykk der *over* er hode vil få to løsninger. Dette gjelder tilfeller hvor spesifikatorene til *over* kan analyseres både av TIME-MOD- (se regel (24)) og PP-TIME-MOD-disjunksjonen (se regel (34)). Denne situasjonen inntreffer for eksempel ved en analyse av setning (44).

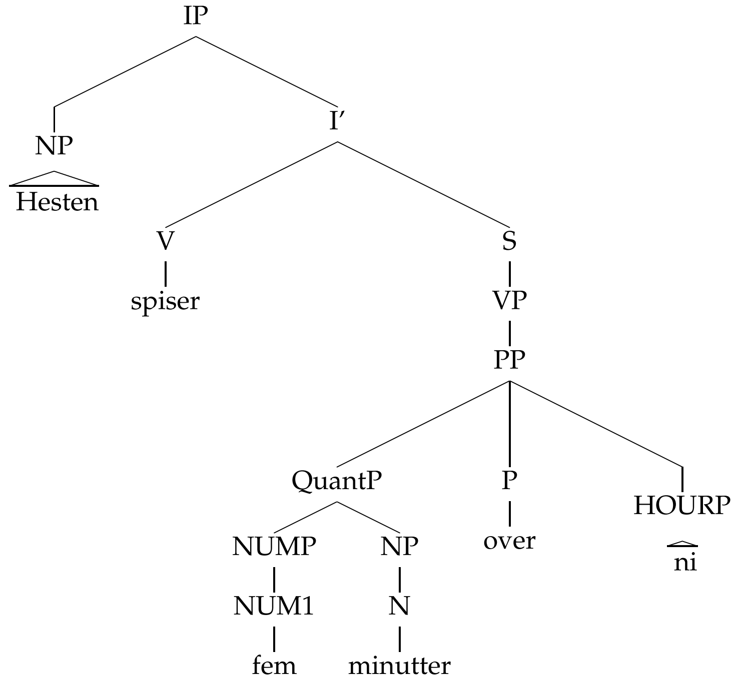
(44) Hesten spiser fem minutter over ni.

I c-struktur (45) er toppnoden i klokkeslettsuttrykket PP-TIME, mens i (46) er toppnoden PP:

(45) *Hesten spiser fem minutter over ni*



(46) *Hesten spiser fem minutter over ni*



I de to c-strukturene (45) og (46) kan vi legge merke til at modernoden for *ni* er ulik. I en PP kan *klokka/klokken* stå foran objektet til hodet i frasen, mens dette ikke er mulig i PP-TIME. Dette betyr at setning (47)

(47) Hesten spiser fem minutter over klokka ni.

vil gi en entydig analyse, hvor klokkeslettsuttrykkets toppnode må være PP. Internt i PP-TIME kan ikke *klokka/klokken* stå foran hodet til objektet, og et klokkeslettsuttrykk som *ti minutter på klokka ni* tillates ikke. I *Norsk aviskorpus* (NA) og *Oslo-korpuset* (OK) er det til sammen 15 forekomster av trigrammet ”minutter over klokka/klokken”, mens søk på ”minutter på klokka/klokken” bare ga ett treff[9]. Dette er et sterkt argument for å tillate at *over* kan stå i PP i et klokkeslettsuttrykk, og på denne måten skille *over* fra *på*. Når *over* står internt i PP-TIME skal derimot ikke *klokka/klokka* tillates foran objektet, og klokkeslettsuttrykket (48) tolkes ikke som grammatisk:

(48) *kvart over klokka ni

Det er resultatet av dette forholdet som gjør at vi får den uheldige situasjonen hvor det oppstår to analyser for setning (44). Ut ifra den grammatikken jeg har definert, kan jeg ikke si at den ene eller den andre løsningen er å foretrekke. I evalueringen, side 112, diskuterer jeg en sammenslåing av PP-TIME og PP, noe som ville bøtet på det problemet jeg her har skissert.

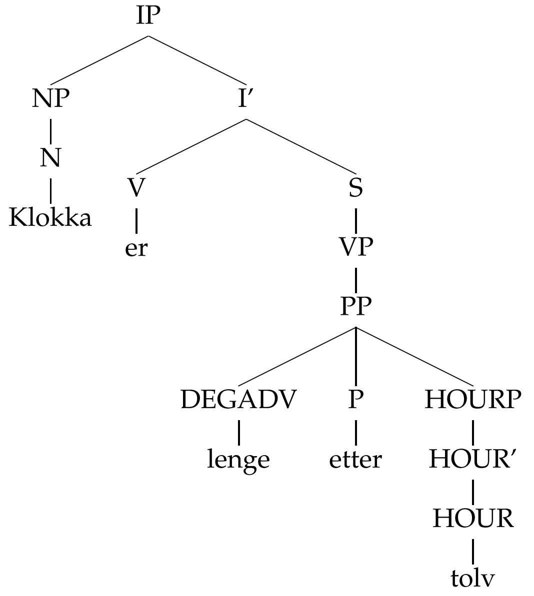
[9] Korpusene ble oppsøkt 12.01.2006.

4.4.3 PP og PREDLINK

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	Jeg kommer hjem *halv ni.*
PP-TIME	Klokka er *noen minutter på ni.*	Jeg kommer hjem *klokka kvart på ni.*
PP	**Klokka er *lenge etter tolv.***	Jeg kommer hjem *like før halv ni.*

C- og f-strukturen til *Klokka er lenge etter tolv* ser slik ut:

(49) *Klokka er lenge etter tolv*



(49)

$$\left[\begin{array}{ll}
\text{PRED} & \text{'være}\Big\langle\Big(\uparrow \text{SUBJ}\Big)\Big(\uparrow \text{PREDLINK}\Big)\Big\rangle\text{'} \\
\text{SUBJ} & \left[\text{PRED} \quad \text{'klokke'}\right] \\
\text{PREDLINK} & \left[\begin{array}{ll}
\text{PRED 'etter}\Big\langle\Big(\uparrow \text{OBJ}\Big)\Big\rangle\text{'} & \\
\text{OBJ} & \left[\begin{array}{l}\text{PRED 'hour-8'} \\ \text{CLOCK-EXPRESSION +} \\ \text{CLOCK-OCCURRENCE +_}\end{array}\right] \\
\text{ADJUNCT} & \left\{\left[\begin{array}{l}\text{PRED 'lenge'} \\ \text{ATYPE} \left[\text{CLOCK +}\right] \\ \text{TEMP-SPEC +}\end{array}\right]\right\} \\
\text{PTYPE-CLOCK} & \left[\text{TEMP +}\right] \\
\text{CLOCK-EXPRESSION} & + \\
\text{CLOCK-OCCURRENCE} & - \\
\text{CLOCK-PREDLINK} & +
\end{array}\right]
\end{array}\right]$$

Den grammatiske funksjonen PREDLINK setter restriksjoner for hvordan klokkeslettsuttrykket kan se ut. Disse restriksjonene er svært like de kravene som stilles til klokkeslettsuttrykk fra TIMES-disjunksjonen, og dette ble drøftet i avsnitt 4.2.1. Subjektsrestriksjonene er identiske i disse tilfellene, men i PP stilles det i tillegg krav til hvilke preposisjoner som godtas i forhold til om subjektet beskriver en hendelse eller om hele setningen er en referanse til et tidspunkt (*Subjekt som hendelse* ble drøftet i avsnitt 4.2.2). Disjunksjonen TIMEEXP setter restriksjonene for hvordan en PP som PREDLINK kan se ut:

(50) TIMEEXP-disjunksjonen for PP i PREDLINK

```
TIMEEXP = { PP: { (^ SUBJ PRED FN)=c klokke
                  (^ SUBJ DEF)=c +
                  | (^ SUBJ PRON-FORM)=c den
                    (^ SUBJ GEND NEUT)=c -}

                    (^ PREDLINK)=!
                    (! CLOCK-PREDLINK)=+
                    (! CLOCK-OCCURRENCE)=-
                    (! PTYPE-CLOCK TEMP)=c +
                   ~(! PTYPE-CLOCK SUBTEMP)=+
```

```
| PP: (^ PREDLINK)=!
  (^ SUBJ CLOCK-EVENT)=+
  (! PTYPE-CLOCK TEMP)=c +
  { ~(^ SUBJ PRED FN)= klokke
   |(^ SUBJ PRED FN)= klokke
    (! CLOCK-OCCURRENCE)=c +_}
    (...)}.
```

Vi ønsker at setningen:

(51) Kampen er fra klokka halv ni

skal tillates i grammatikken. Derimot ønsker vi ikke at uttrykket

(52) *Klokka er fra halv ni

skal analyseres som et uttrykk som definerer et klokkeslett. Preposisjonen *fra* har, i likhet med *til, siden* og *innen* iboende egenskaper som gjør at den er uegnet som hode i en PP i kontekster som (52). *Innen, fra, siden* og *til* har følgende trekk i f-strukturen:

$$\left[\text{PTYPE-CLOCK}\begin{bmatrix}\text{SUBTEMP} & + \\ \text{TEMP} & +\end{bmatrix}\right]$$

Trekket SUBTEMP + hindrer at det første disjunktet i TIMEEXP-disjunksjonen fra regel (50) velges. Det betyr at dersom den grammatiske funksjonen til klokkeslettsuttrykket er PREDLINK, kan *fra, innen, siden* og *til* bare stå som hode i uttrykk hvor subjektet refererer til en hendelse. Ved å sette disse restriksjonene inn i grammatikken, følger XLE-grammatikken de slutningene jeg trakk i avsnitt 3.3.3, og den videre teoretiseringen rundt preposisjonene *innen, siden, fra* og *til* fra side 44 med hensyn til de syntaktiske rammene disse preposisjonene kan opptre i.

4.4.4 PP og ADJUNCT

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	Jeg kommer hjem *halv ni.*
PP-TIME	Klokka er *noen minutter på ni.*	Jeg kommer hjem *klokka kvart på ni.*
PP	Klokka er *lenge etter tolv.*	**Jeg kommer hjem *like før halv ni.***

Klokkeslettsuttrykk med PP som toppnode i c-strukturen, og med den grammatiske funksjonen ADJUNCT, analyseres med den vanlige PP-regelen som henger under VP. VP-regelen vises i (53).

(53) VP-regelen

```
VP --> (V: ^=!)
       (SADV: ! $ (^ ADJUNCT))
       ({ NP: (^ OBJ)=!
        | PRON: (^ OBJ)=!})
        PP*: @ADJUNCTPP,
       (TIMEEXP: (! CLOCK-EXPRESSION)=c +).
```

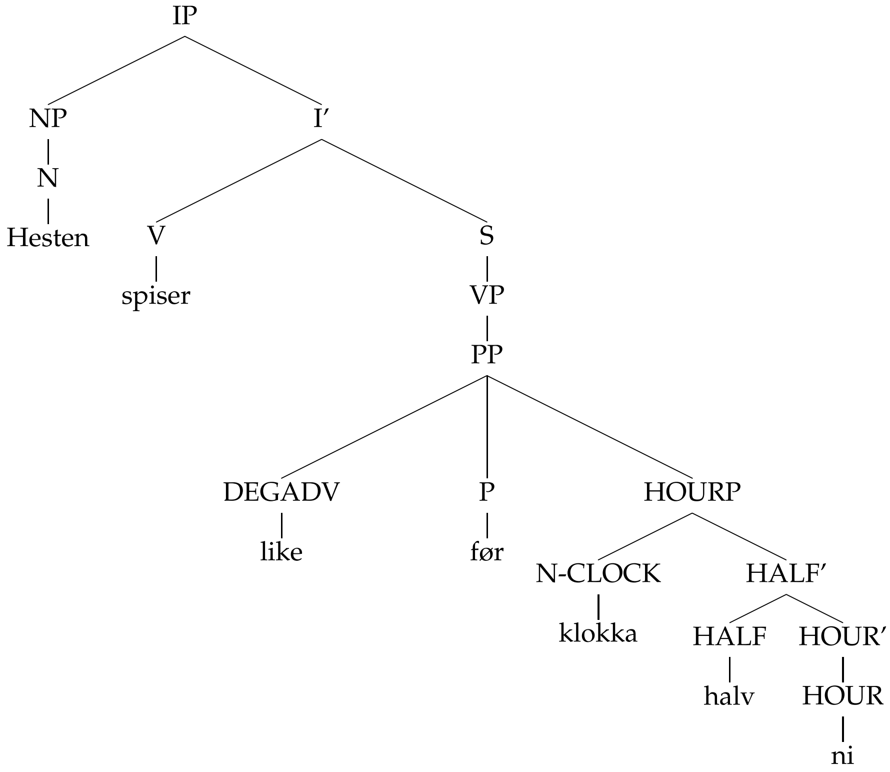
Dette er de eneste klokkeslettsuttrykkene som blir analysert av en standard NorGram-regel. Alle andre klokkeslettsuttrykk blir analysert av TIMEEXP-variabelen. Men vi ønsker at en PP som klokkeslettsuttrykk kun skal henge under VP og IP. Derfor har jeg hardkodet at PP-fraser under NP (54) ikke kan være klokkeslettsuttrykk. Dette hindrer overgenereringer:

(54) NP-regelen

```
NP --> (D)
        A*:!$(^ ADJUNCT);
        N: ^=!;
        PP*: @ADJUNCTPP
            ~(! CLOCK-EXPRESSION)=+.
```

C- og f-strukturen til *Hesten spiser like før klokka halv ni* er gjengitt i (55).

(55) *Hesten spiser like før klokka halv ni*



(55)
$$\left[\begin{array}{ll}
\text{PRED} & \text{'spise}\left\langle\left(\uparrow \text{SUBJ}\right)\right\rangle\text{'} \\
\text{SUBJ} & \left[\text{PRED} \quad \text{'hest'}\right] \\
\text{ADJUNCT} & \left\{\left[\begin{array}{ll}
\text{PRED} & \text{'før}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\
\text{OBJ} & \left[\begin{array}{l}
\text{PRED 'halv}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\
\text{OBJ} \left[\text{PRED 'hour-9'}\right] \\
\text{CLOCK-EXPRESSION +} \\
\text{CLOCK-OCCURRENCE +_}
\end{array}\right] \\
\text{ADJUNCT} & \left\{\left[\begin{array}{l}
\text{PRED 'like'} \\
\text{ATYPE} \left[\text{CLOCK +}\right] \\
\text{TEMP-SPEC +}
\end{array}\right]\right\} \\
\text{PTYPE-CLOCK} & \left[\text{TEMP +}\right] \\
\text{CLOCK-EXPRESSION} & + \\
\text{CLOCK-OCCURRENCE} & +_
\end{array}\right]\right\}
\end{array}\right]$$

Ingen spesielle restriksjoner legges her til klokkeslettsuttrykket, og uttrykket blir analysert ved hjelp av de disjunksjonene/reglene jeg har lagd som tillegg til NorGrams PP-regel. Klokkeslettsuttrykk som er frie tidsadverbial og har den grammatiske funksjonen ADJUNCT har generelt færre restriksjoner knyttet til syntaksen enn det de bundne klokkeslettsadverbialene i PREDLINK har.

4.4.5 Temporale uttrykk i PP

I forhold til NorGrams grammatikk, må bare noen få linjer legges til for at preposisjonsfrasen skal kunne behandle klokkeslettsuttrykk. Mitt tillegg til preposisjonsfrasen vil bare gjøre PP i stand til å analysere klokkeslettsuttrykk, og tillegget vil ikke kunne analysere kontekster hvor en del av de samme spesifikatorene (fra PP-TIME-MOD-disjunksjonen, se regel (34)) er brukt i uttrykk av en annen art. Dette kunne vært uttrykk som:

(56) Snøen kommer like før jul.

Dette tidsuttrykket har allerede en analyse i NorGram, men jeg har ikke for alvor utforsket mulighetene for å slå sammen en kategori for generelle temporale spesifikatorer med de spesifikatorene jeg har implementert for klokkeslettsuttrykkene. Jeg vil anta at dette er mulig, men det må helt sikkert gjøres tilpasninger, for spesifikatorer til klokkeslettsuttrykk vil ikke nødvendigvis kunne anvendes parallelt med en hvilken som helst spesifikator i et temporalt uttrykk. Min variabel PP-TIME-MOD kan sannsynligvis danne et godt grunnlag for en generell regel for andre tidsuttrykk, men utarbeidingen av slike generelle regler faller utenfor denne oppgavens rammer.

En vurdering av hvordan mitt tillegg om klokkeslettsuttrykkene kan utvides til å gjelde mer generelle tidsuttrykk blir drøftet i evalueringen, se avsnitt 5.2.3.

4.5 Uuttrykt klokketime

I klokkeslettsuttrykk kan klokketimen ofte undertrykkes. Dette gjøres når meningen er å vise til tidspunkt som gjentas i løpet av et døgn (eksempel (57)), eller det gjøres når det er underforstått hvilken time det vises til (58).

(57) Bussen går alltid klokka ti over halv.

(58) Jeg burde vært her klokka ti, men jeg rakk ikke å komme før kvart over.

I frasestrukturregelen for P′-TIME (24) og HALF′ (3) tillates at objektet til enten preposisjonen i PP-TIME, eller *halv*s objekt undertrykkes. I f-strukturen representeres den undertrykte klokketimen med ′pro′ som sitt PRED (60).

(59) Klokka er ti over.

(59)
$$\left[\begin{array}{ll} \text{PRED} & \text{'være}\left\langle(\uparrow \text{SUBJ})(\uparrow \text{PREDLINK})\right\rangle\text{'} \\ \text{SUBJ} & \left[\text{PRED} \quad \text{'klokke'}\right] \\ \text{PREDLINK} & \left[\begin{array}{ll} \text{PRED} & \text{'over}\left\langle(\uparrow \text{OBJ})\right\rangle\text{'} \\ \text{OBJ} & \left[\text{PRED 'pro'}\right] \\ \text{PTYPE-CLOCK} & \left[\text{TEMPFINALE +}\right] \\ \text{SPEC} & \left[\begin{array}{l} \text{PRED 'pro'} \\ \text{SPEC} \left[\text{NUMBER} \left[\text{PRED 'ti'}\right]\right] \end{array}\right] \\ \text{CLOCK-PREDLINK} & + \\ \text{CLOCK-EXPRESSION} & + \\ \text{CLOCK-OCCURRENCE} & - \end{array}\right] \end{array}\right]$$

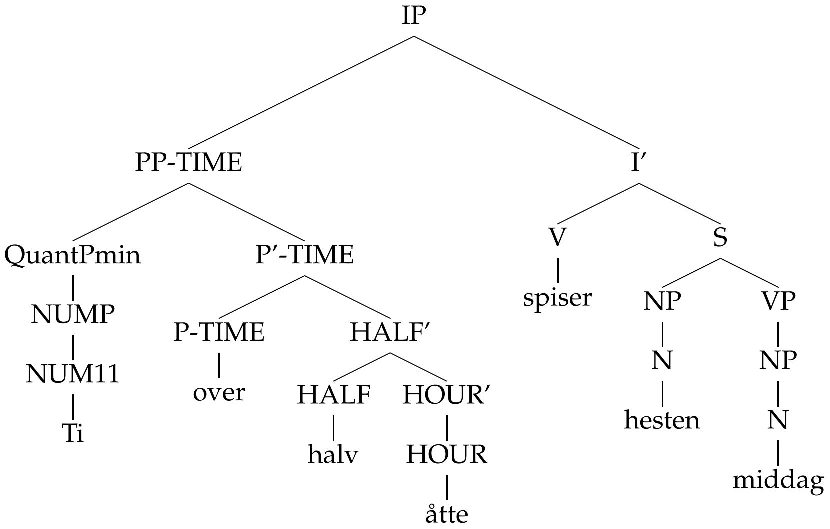
4.6 Topikalisering

I avsnittet *Syntaktiske egenskaper for klokkeslettsuttrykkene* (avsnitt 3.2.3) drøftet jeg hvordan hele eller deler av klokkeslettsuttrykket kan flyttes fram til forfeltet i setningen. Frie klokkeslettsadverbialer med den grammatiske funksjonen ADJUNCT kan ikke deles opp, og ved topikalisering må hele klokkeslettsuttrykket flyttes til forfeltet. En oppdeling av klokkeslettsuttrykket fører til at setningen blir ugrammatisk:

(60) *Halv ni spiser hesten middag ti over.

Setning (61) følger derimot kravet om at hele klokkeslettsuttrykket må følge med under topikalisering og er grammatisk.

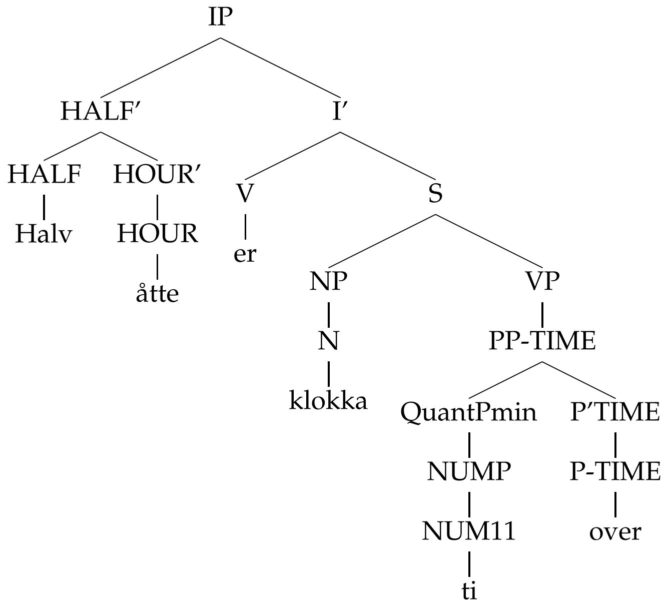
(61) *Ti over halv åtte spiser hesten middag*



I f-strukturen i min grammatikk markeres det ikke hvilket setningsledd som topikaliseres, og f-strukturen for (61) vil være identisk med f-strukturen for *Hesten spiser middag ti over halv åtte*.

De bundne klokkeslettsadverbialene med den grammatiske funksjonen PREDLINK lar seg imidlertid dele opp ved topikalisering. Min grammatikk tillater de syntaktiske strukturene som beskrives i avsnitt 3.2.3. Det vil si at dersom klokkeslettsuttrykket er delt opp, vil den topikaliserte delen av uttrykket være objektet til hodet i PP/PP-TIME.

(62) *Halv åtte er klokka ti minutter over*



I frasestrukturregelen for IP defineres alle de alternative ekspansjonene som muliggjør ulike former for topikalisering.

(63) Frasestrukturregelen for IP

```
IP --> { { NP: (^ SUBJ)=!
         | PRON: (^ SUBJ)=!}
         |PP*: @ADJUNCTPP
         |TIMEEXP: (! CLOCK-EXPRESSION)=c +
         | HOUR': (^ PREDLINK OBJ)=!
         |HALF':(^ PREDLINK OBJ)=!}
         I': @MRS-PRPSTN.
```

Min klokkeslettsgrammatikk er svært enkel med hensyn til hvilke strukturer som kan analyseres. Dersom klokkeslettsgrammatikken skal takle mer komplisert langdistanseavhengighet som involverer funksjonell usikkerhet[10], må det skrives spesialtilpassede regler i forhold til hvordan den omkringliggende grammatikken ser ut. Dette blir drøftet i oppsummeringen av kapitlet.

4.7 MRS-representasjonen

MRS-projeksjonen som gis for klokkeslettsuttrykkene er inspirert av den konvensjonelle måten å representere MRSer innen NorGram. Imidlertid er ikke NorGram-grammatikken bygd for å parse klokkeslettsuttrykk av noen slag, og den helhetlige MRS-representasjonen for klokkeslettsuttrykkene måtte bygges opp fra grunnen av.

Flere av EPene som benyttes i min grammatikk er inspirert av EPer i MRS-projeksjonen i LinGO ERG-grammatikken (CSLI LinGO Lab). Her innføres 'measure'-EPer for å definere et tidsrom, og 'numbered_hour'-relasjoner for å uttrykke den hele klokketimen. Disse EPene brukes også i min grammatikk, men forholdet mellom EPene i min grammatikk og ERG-grammatikken er ikke fullstendig parallelt. Min grammatikk er også langt mer robust i forhold til å takle variasjoner over måter å uttrykke klokkeslett på.

Mange av templatene som brukes for å derivere MRS-representasjoner er eksisterende NorGram-templater. I tillegg har andre NorGram-templater blitt modifisert, slik at de kan brukes også innenfor et klokkeslettsuttrykk.

[10] *Functional uncertainty*, se Falk (2001, s. 152).

En rekke templater har imidlertid måttet bygges opp fra grunnen av, siden klokkeslettsuttrykkenes grammatikk skiller seg såpass mye fra den øvrige grammatikken. Min veileder, Helge Dyvik, har skrevet disse templatene. Jeg kommer bare til å presentere hvordan selve MRS-projeksjonen ser ut i web-grensesnittet (fra siden: http://hilarion.aksis.uib.no:8030/logon/xle.xml), og vil ikke gå innpå XLE-koden. I avsnitt 4.7.5 kommenterer jeg hvordan videre utvideler kan gjøres for å optimalisere MRS-representasjonen. XLE-koden for templatene som benyttes, gjengis i tillegg D.

4.7.1 HOURP

	PREDLINK	ADJUNCT
HOURP	**Klokka er *ni*.**	**Jeg kommer hjem *halv ni*.**
PP-TIME	Klokka er *noen minutter på ni.*	Jeg kommer hjem *klokka kvart på ni.*
PP	Klokka er *lenge etter tolv.*	Jeg kommer hjem *like før halv ni.*

I alle klokkeslettsuttrykk introduseres EPen 'temp-loc_rel'. Denne innføres i leksikon ved at templatet @NUMBEREDHOUR kalles av tallordet som refererer til klokketimen, og dette templatet kaller igjen @MRS-HOUR.[11] Dersom det foreligger en modifikator til klokketimen (for eksempel *halv* eller *ti minutter*) vil 'temp-loc_rel's ARG1 være bundet til ARG2-variabelen som denne modifikatoren innfører. Om det ikke foreligger noen modifikator, vil ARG1-variabelen til 'temp-loc_rel' være bundet til ARG2 til 'clock-id_rel'-EPen (dersom klokkeslettsuttrykket er PREDLINK) eller til ARG2 av EPen 'unspec_loc_rel' (når uttrykkets funksjon er ADJUNCT).

Timepoint

I tillegg til at 'temp-loc_rel' står i alle klokkeslettsuttrykk, blir også EPen 'timepoint_rel' innført én eller flere ganger. Hver tidsrelasjon innen HOURP og PP-TIME får sin egen 'timepoint_rel'-EP. Dette gjelder altså den hele klokketimen, og de modifikasjonene som skal legges til eller trekkes fra denne klokketimen når man skal regne seg fram til klokkeslettet som uttrykkes. Dette betyr at i klokkeslettsuttrykket *halv åtte* uttrykkes to timepoint-relasjoner, og i *ti på halv åtte* defineres tre slike relasjoner. 'temp-loc_rel's ARG1 binder 'timepoint_rel's ARG0-verdi for den timepoint-relasjonen som innføres av den hele klokketimen (fra templatet @NUMBEREDHOUR). Når timepoint-relasjonen innføres av en modifikasjon, bindes dens ARG0 av den modifikatoren som innfører den.

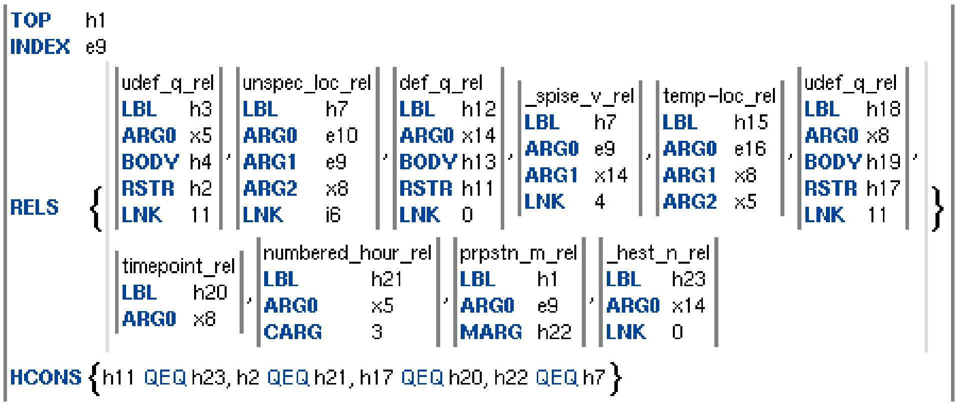
ADJUNCT

I setning (64) forekommer ingen modifiseringer av klokketimen.

[11]Dersom klokketimen er uuttrykt, kalles templatet @MRS-EMPTYHOUR som innfører 'temp-loc_rel', se avsnitt 4.7.4.

(64) Hesten spiser klokka tre.

I dette tilfellet viser 'temp-loc's ARG1 til 'unspec_loc_rel's ARG2-variabel. Vi finner også igjen variabelen i 'timepoint_rel's ARG0, som i dette tilfellet innføres av @MRS-HOUR. Hele MRS-projeksjonen for (64) blir representert i figur 4.1.



Figur 4.1: MRS-representasjonen av *Hesten spiser klokka tre*

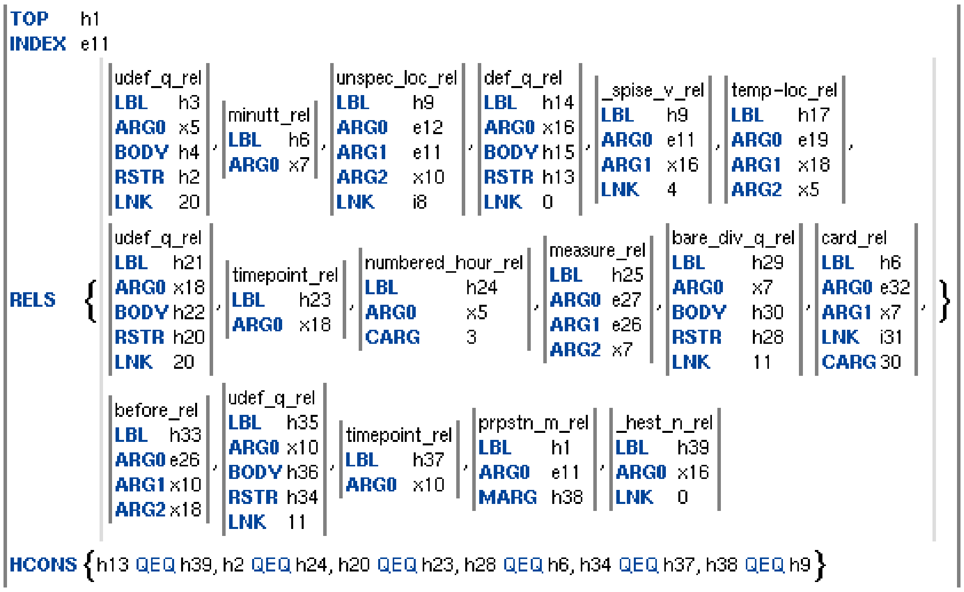
Dersom HOURP er toppnode, er *halv* det eneste elementet som kan modifisere tidspunktet som angis av den hele klokketimen. I uttrykket *halv ni* kan uttrykket forstås som 'trekk 30 minutter fra klokka ni'. Det er akkurat denne intuisjonen som reflekteres i MRSen for disse uttrykkene. For uttrykket

(65) Hesten spiser klokka halv tre.

vil 'temp-loc_rel's ARG1 vise til ARG2 i EPen 'before_rel'. MRSen for (65) gjengis i figur 4.2.

I MRS-representasjonen i figur 4.2 er timepoint-relasjonen med etiketten h23 innført av @MRS-HOUR, og dens ARG0 er koindeksert med 'temp-loc_rel's ARG1 og 'before_rel's ARG2. En ny viktig EP introduseres også i figur 4.2, nemlig 'measure_rel'. Vi ser at 'before_rel's ARG0 er eventen e26. Denne knytter 'before_rel' sammen med 'measure_rel's ARG1. 'measure_rel' måler opp et antall minutter eller sekunder som skal legges til eller trekkes fra den hele klokketimen for å komme fram til det riktige tidspunktet. 'measure_rel's ARG2 viser her til tidsrelasjonen som skal trekkes fra *tre*. *Halv* representeres ved minuttrelasjonen 'minutt_rel', og 'card_rel', som definerer antall minutter, samt 'before_rel' og 'measure_rel' (se figur 4.3).

I 'before_rel's ARG1 koindekseres verdien x10 med 'unspec_loc_rel's ARG2 og ARG0 til timepoint-relasjonen med h37-etiketten. Denne before-relasjonen innføres av @MRS-HALV, og ARG1 refererer her til tidspunkt-



Figur 4.2: MRS-representasjon av *Hesten spiser klokka halv tre*

$$\begin{bmatrix}\texttt{minutt\_rel} & \\ \text{LBL} & \text{h6} \\ \text{ARG0} & \text{x7}\end{bmatrix}\begin{bmatrix}\texttt{card\_rel} & \\ \text{LBL} & \text{h6} \\ \text{ARG0} & \text{e32} \\ \text{ARG1} & \text{x7} \\ \text{LNK} & \text{i31} \\ \text{CARG} & 30\end{bmatrix}\begin{bmatrix}\texttt{before\_rel} & \\ \text{LBL} & \text{h33} \\ \text{ARG0} & \text{e26} \\ \text{ARG1} & \text{x10} \\ \text{ARG2} & \text{x18}\end{bmatrix}\begin{bmatrix}\texttt{measure\_rel} & \\ \text{LBL} & \text{h25} \\ \text{ARG0} & \text{e27} \\ \text{ARG1} & \text{e26} \\ \text{ARG2} & \text{x7}\end{bmatrix}$$

Figur 4.3: EPene som representerer *halv* i figur 4.2

relasjonen på dette nivået i f-strukturen, mens ARG2 refererer til tidspunktet som defineres som objektet til *halv*. At ARG1 er koindeksert med 'unspec_loc_rel' antyder at denne before-relasjonen er innført av hodet i klokkeslettsuttrykket.

'measure_rel' innføres enten av templatet @MRS-HALV i leksikon-oppslaget til *halv*, eller gjennom templatet @MRS-MINUTEMEASURE, som blir kalt i frasestrukturreglene. Intuisjonen er at det er det samme klokkeslettet som uttrykkes av *halv åtte* og *tretti (minutter) på åtte*. Disse klokkeslettsuttrykkene får derfor semantisk lik MRS-representasjon[12], selv om deres c- og f-struktur er forskjellig. Semantisk kan hele klokkeslettet skrives om til 7.30.

PREDLINK

Dersom klokkeslettsuttrykkets grammatiske funksjon er PREDLINK, innføres EPen 'clock-id_rel'. Denne EPen oppfører seg slik 'unspec_loc_rel' gjør i en setning hvor klokkeslettsuttrykket står som ADJUNCT. Siden klokkeslettsuttrykket i PREDLINK er en del av argumentstrukturen til kopulaverbet *være*, kalles templatet @CLOCKCOPULA av *være* i leksikon. I MRS-representasjonen for

(66) Klokka er tre.

ser vi at 'clock-id_rel's ARG0 er koindeksert med verdien for INDEX i MRSen (figur 4.4). I f-strukturen korresponderer dette med verbet *være*, hvis PRED står på ytterste nivå i strukturen (se for eksempel f-strukturen til *Klokka er ni*, eksempel (11)).

I figur 4.4 ser vi at 'clock-id_rel's ARG2 tar x11 som verdi. Denne verdien gjenfinnes som ARG1 i EPen til 'temp-loc_rel'. Dersom den grammatiske funksjonen til klokkeslettsuttrykket er ADJUNCT, ville verdien til 'temp-loc_rel's ARG1 vært koindeksert med 'unspec_loc_rel's ARG2 i stedet (se figur 4.1).

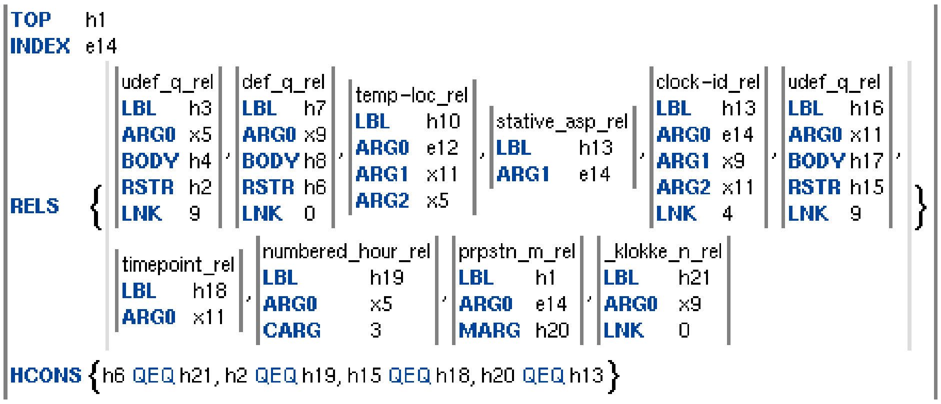
Dersom det foreligger temporale modifikasjoner i forhold til klokketimen, vil derimot 'temp-loc_rel's ARG1 referere til denne past/before-relasjonen, også dersom klokkeslettsuttrykkets funksjon er PREDLINK. For setning (67):

(67) Klokka er halv tre.

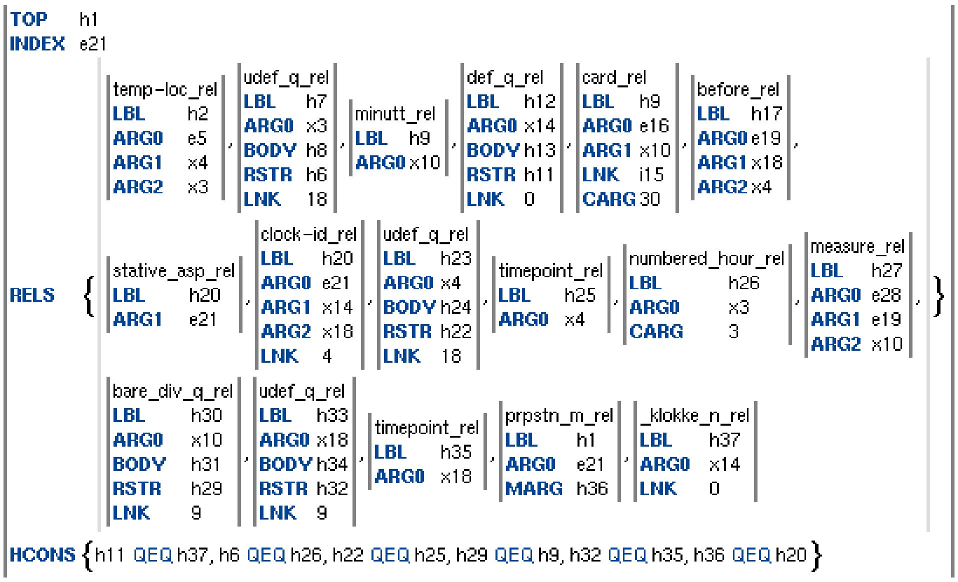
illustreres MRS-representasjonen i figur 4.5.

---

[12]Det finnes visse forskjeller mellom MRS-representasjonene, blant annet med hensyn til hvordan 'card_rel' og 'minutt_rel' ser ut. I uttrykket *tretti minutter på tre* innføres for eksempel EPen '_minutt_n_rel', mens i *halv tre* brukes 'minutt_rel'. Disse forskjellene endrer imidlertid ikke det semantiske som uttrykkes.



Figur 4.4: MRS-representasjonen av *Klokka er tre*



Figur 4.5: MRS-representasjon for *Klokka er halv tre*

Vi ser i MRSen i figur 4.5 at 'before_rel's ARG1 har samme verdi som 'clock-id_rel's ARG2. Slik indikeres her at den grammatiske funksjonen til klokkeslettsuttrykket *halv tre* er PREDLINK, i forhold til klokkeslettsuttrykket i figur 4.2. Den EPen som f-struktur-hodet i klokkeslettsuttrykket innfører, vil alltid ha en ARG1 som deler verdi med 'clock-id_rel's ARG2 når klokkeslettsuttrykkets grammatiske funksjon er PREDLINK.

4.7.2 PP-TIME

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	Jeg kommer hjem *halv ni.*
PP-TIME	**Klokka er *noen minutter på ni.***	**Jeg kommer hjem *klokka kvart på ni.***
PP	Klokka er *lenge etter tolv.*	Jeg kommer hjem *like før halv ni.*

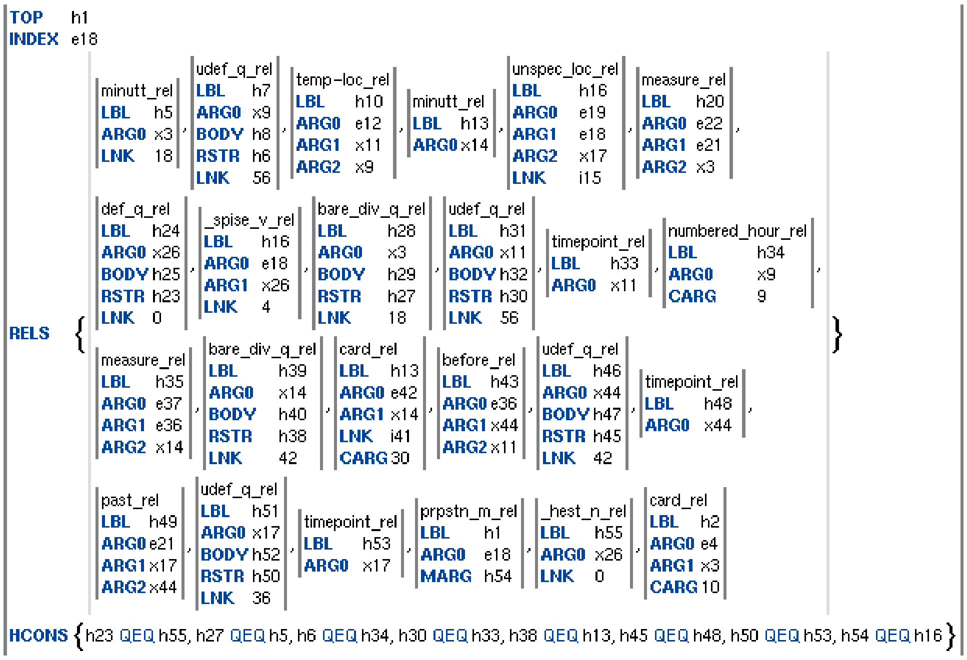
Som jeg var inne på tidligere, er det et mål at MRS-representasjonen skal uttrykke klokkeslett som refererer til det samme tidspunktet på en mest mulig ensartet måte, selv om c-strukturen til ytringene ikke er identiske. Når toppnoden i c-strukturen var HOURP, kunne bare *halv* skytes inn som temporal modifikator for klokketimen. For PP-TIME er det flere muligheter for å modifisere klokkeslettet. I setningen

(68) Hesten spiser klokka ti over halv ni.

defineres et tidsrom i spesifikatoren til hodet i preposisjonsfrasen PP-TIME. *Ti* uttrykker i denne sammenhengen *ti minutter*. På/over-preposisjonen i hodet til PP-TIME vil avgjøre om det definerte tidsrommet i spesifikatoren skal trekkes fra eller legges til klokkeslettet som defineres av objektet til preposisjonen (se f-strukturen til *Hesten spiser klokka ti over halv ni* (30)). De to tidsrommene som her defineres, blir i MRS-representasjonen (se figur 4.6) uttrykt med hver sine 'measure_rel'-EPer, og disse modifikatorene vil også ha sine timepoint-EPer.

De to measure-EPene uttrykker at det er to forskjellige tidsrom som skal trekkes fra/legges til klokketimen *ni*. *Halv ni*-uttrykket representeres med de samme EPene som vi så i setning (65) og figur 4.2. Vi ser imidlertid at EPen 'past_rel' er kommet til, og dens ARG2 er identisk med 'before_rel's ARG1.

'past_rel'-EPen innføres av @MRS-PAST-templatet som kalles av leksikon-oppslaget til *over* dersom *over* står i PP-TIME. 'past_rel's ARG2 kobler EPen til 'before_rel'-EPens ARG1 ved at de viser til samme verdi: x44. I f-strukturen for *ti over halv ti* står *over* ytterst, og tar *halv ti* som sitt objekt. Reglene for @MRS-PAST og @MRS-BEFORE er definert slik at deres EPs ARG2 blir identisk med ARG1-verdien til EPen som objektets hode introduserer. @MRS-MINUTEMEASURE innfører 'measure_rel'-EPene for spesifikatoren i *ti over halv ti*, mens 'measure_rel'-EPen for *halv* i *halv ti* blir innført av @MRS-HALV-templatet i leksikon-oppslaget til *halv*.



Figur 4.6: MRS-representasjon for *Hesten spiser klokka ti over halv ni*

Vage tidsrom

'measure_rel' trenger ikke å referere til et definert antall minutter eller sekunder. Innen PP-TIME-frasen kan uttrykk som *noen minutter over ni* forekomme. En slik konstruksjon gjør at det blir umulig å regne seg fram til en digital representasjon av klokkeslettsuttrykket, men selve representasjonen i MRS skiller seg ikke nevneverdig fra måten *ti over ni* representeres på. I stedet for at EPen 'card_rel' definerer antall minutter, kvantifiseres minuttene av EPen som innføres av *noen*. I dette tilfellet kan et utsnitt av MRSen for *noen minutter over ni* være identisk med figur 4.7.

$$
\begin{bmatrix}\texttt{measure\_rel} & \\ \text{LBL} & \text{h23} \\ \text{ARG0} & \text{e24} \\ \text{ARG1} & \text{e20} \\ \text{ARG2} & \text{x3}\end{bmatrix}
\begin{bmatrix}\texttt{noen\_q\_rel} & \\ \text{LBL} & \text{h9} \\ \text{ARG0} & \text{x3} \\ \text{BODY} & \text{h10} \\ \text{RSTR} & \text{h8} \\ \text{LNK} & \text{10}\end{bmatrix}
\begin{bmatrix}\texttt{\_minutt\_n\_rel} & \\ \text{LBL} & \text{h2} \\ \text{ARG0} & \text{x3} \\ \text{LNK} & \text{13}\end{bmatrix}
$$

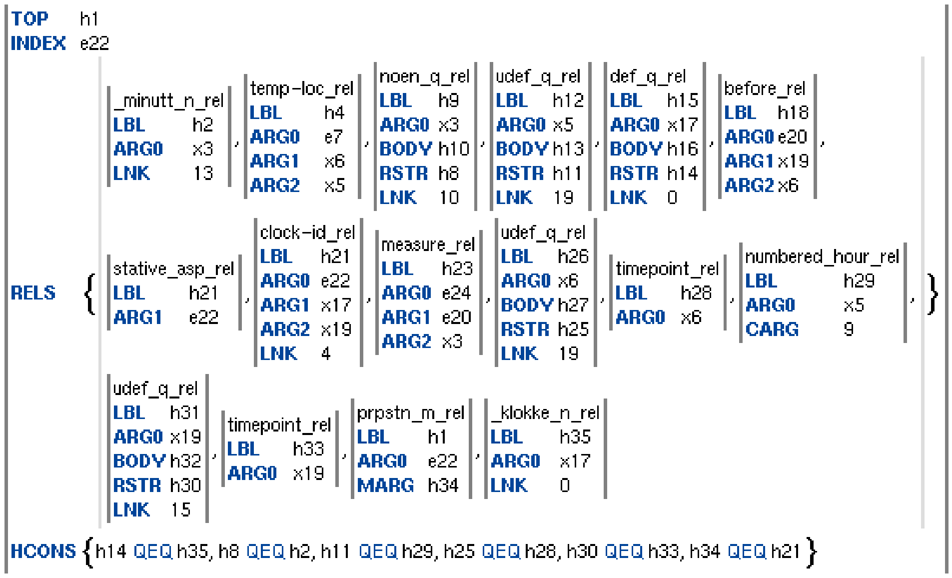
$$
\begin{bmatrix}\text{HCONS} & \text{h8-QEQ-h2}\end{bmatrix}
$$

Figur 4.7: EP-ene som representerer *noen minutter*

Hele MRSen til uttrykket

(69) Klokka er noen minutter på ni.

illustreres i figur 4.8.



Figur 4.8: MRS-representasjon for *Klokka er noen minutter på ni*

4.7.3 PP

	PREDLINK	ADJUNCT
HOURP	Klokka er *ni.*	Jeg kommer hjem *halv ni.*
PP-TIME	Klokka er *noen minutter på ni.*	Jeg kommer hjem *klokka kvart på ni.*
PP	**Klokka er *lenge etter tolv.***	**Jeg kommer hjem *like før halv ni.***

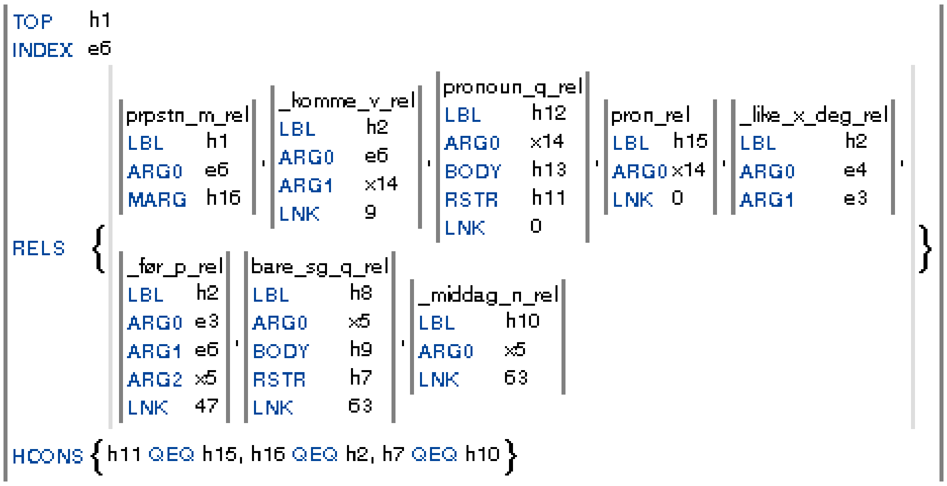
I klokkeslettsuttrykk hvor toppnoden er PP anvendes de konvensjonelle leksikonoppslagene for preposisjonene som står som hode i frasen. For PP brukes det vanlige templatet for preposisjoner, @SEMPREP. Målsetningen er at representasjonen av klokkeslettsuttrykk i PP skal være mest mulig lik enhver annen temporal relasjon som uttrykkes med PP som toppnode. For setningene

(70) Jeg kommer like før middag.

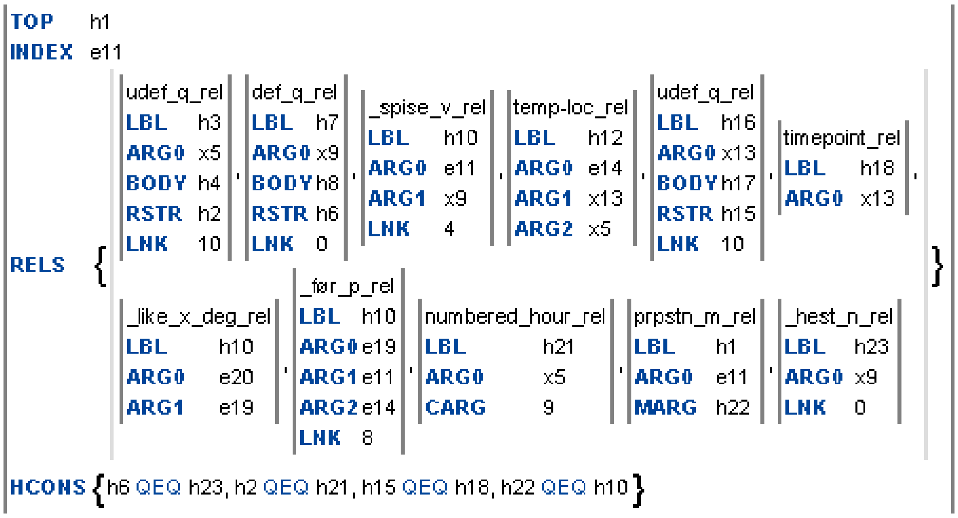
og

(71) Hesten spiser like før ni.

ønsker vi at MRS-representasjonene for preposisjonsfrasen skal bygges opp på samme måte. *Like* er et gradsadverb som her er en temporal modifikator som står som spesifikator for preposisjonen *før*. At (71) er et klokkeslettsuttrykk spiller her ingen rolle for representasjonen av *før*s spesifikator. NorGrams representasjon av MRS-strukturen til (70) gjengis i figur 4.9, og strukturen for (71) illustreres i figur 4.10.



Figur 4.9: MRS-representasjon for *Jeg kommer like før middag*

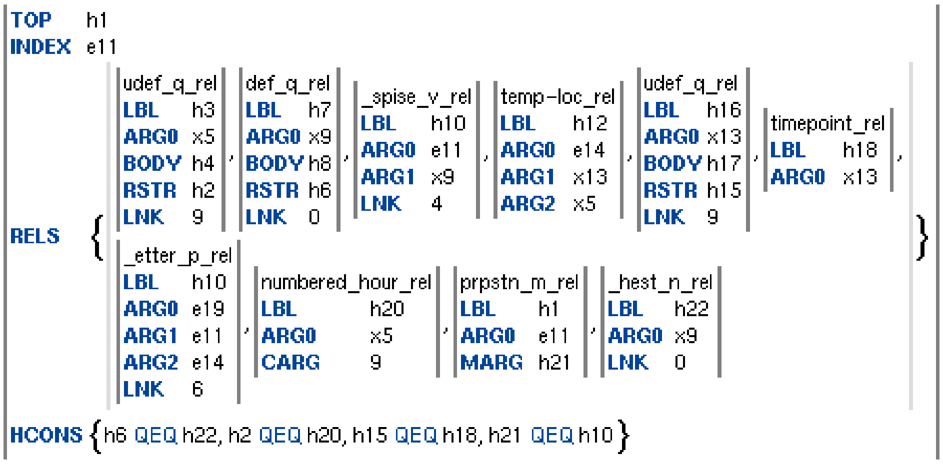


Figur 4.10: MRS-representasjon for *Hesten spiser like før ni*

I figur 4.10 innføres EPen '_før_p_rel' av @SEMPREP. Dens ARG0-event er koindeksert med ARG1 for spesifikatorens EP. På samme måte som past/ before-EPene, vil ARG0 til preposisjons-EPen i PP referere til den temporale modifikasjonen som foreligger. I setningen

(72) Hesten spiser etter ni.

finnes ingen spesifikator, og event-verdien for ARG0 er unik (se figur 4.11).



Figur 4.11: MRS-representasjon for *Hesten spiser etter ni*

@SEMPREP-templatene introduserer ikke EPen 'timepoint_rel', så disse EPene forekommer aldri for tidsrelasjoner som uttrykkes utenfor HOURP og PP-TIME. Imidlertid vil en tidsreferanse uttrykt i spesifikatoren til PP stå som en measure-relasjon dersom spesifikatoren er en nomen- eller kvantorfrase.

Rekursive klokkeslettsuttrykk

I motsetningen til klokkeslettsuttrykk hvor toppnoden er PP-TIME eller HOURP, kan klokkeslettsuttrykkene med PP som toppnode være uendelig lange (se XLE-regelen til PP, avsnitt 4.4.1). Dette er på grunn av at en PP kan ta en ny PP som sitt objekt. Setningen

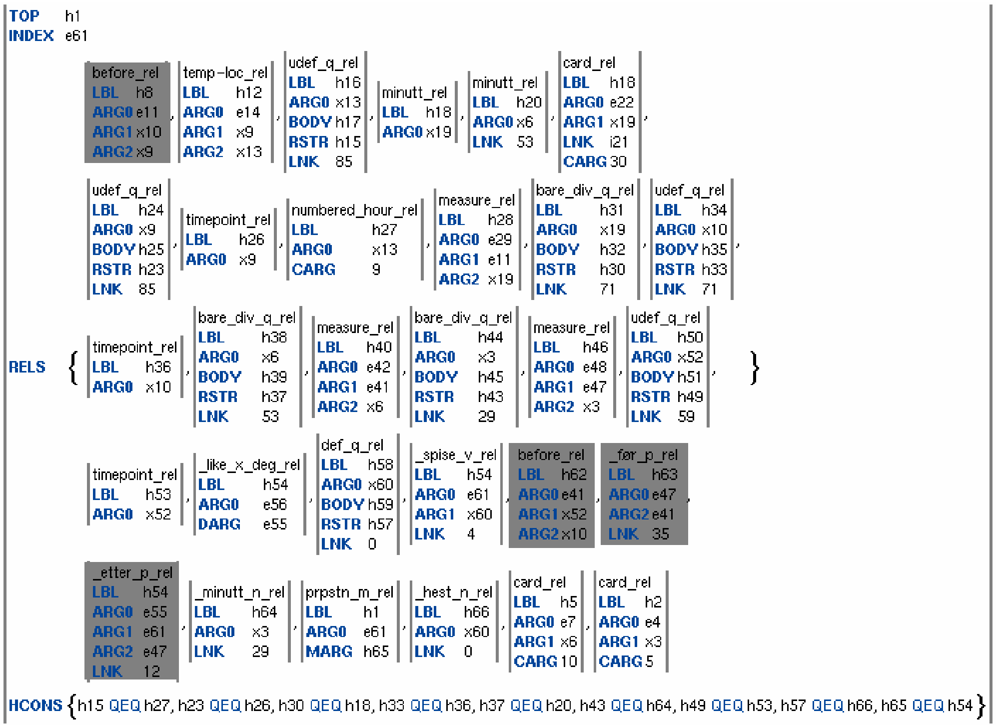
(73) Hesten spiser like etter fem minutter før ti på halv ni.

vil få en komplisert MRS-representasjon som er vanskelig å lese. I forhold til den hele klokketimen *ni* innføres fire modifikasjoner, altså relasjoner som spesifiserer at et tidsrom skal legges til eller trekkes fra klokketimen. Disse fire tidsrommene er:

1. halv (tretti minutter)
2. ti (ti minutter)
3. fem minutter

4. like

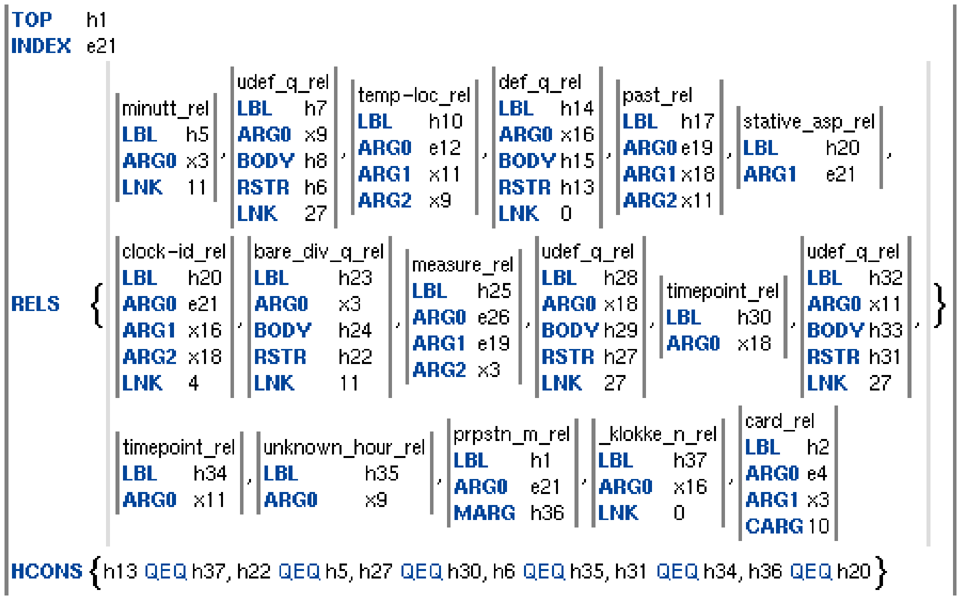
De fire tidsrommene vil bli representert med en EP som uttrykker en before/past-relasjon internt i PP-TIME/HOURP, og vanlige preposisjons-EPer i PP. Disse EPene referere igjen til 'measure'-relasjoner, unntatt *etter*-EPen som refererer til gradsadverbialet *like*s EP. I (73) uttrykkes tre tidspunkter innefor PP-TIME, og disse tre tidspunktene får hver sin timepoint-EP. MRS-representasjonen for (73) illustreres i figur 4.12, og de fire EPene som uttrykker før/etter-relasjonene er uthevet.



Figur 4.12: MRS-representasjon for *Hesten spiser like etter fem minutter før ti på halv ni*

4.7.4 Uuttrykt klokketime

Dersom klokketimen i klokkeslettsuttrykket ikke uttrykkes (se avsnitt 4.5), innføres EPen 'unknown_hour_rel'. I MRS-representasjonen oppfører denne seg nøyaktig slik 'numbered_hour_rel' gjør, bortsett fra at den ikke har en CARG-verdi hvor klokketimen angis. I figur 4.13 vises MRS-projeksjonen for *Klokka er ti over*.



Figur 4.13: MRS-representasjonen for *Klokka er ti over*

4.7.5 Digital representasjon

En naturlig målsetning for MRS-representasjonene er at de direkte skal danne grunnlag for å definere et digitalt klokkeslettsuttrykk. Dersom klokkeslettsuttrykket bare inneholder én modifikator i tillegg til den hele klokketimen (som i uttrykket *fem på åtte*) vil dette bli representert med en before-relasjon (h24), en EP for klokketimen (h23) og en EP som definerer de antall minutter som skal trekkes fra klokketimen (h2). Tidsuttrykket bindes sammen med 'measure_rel' og 'temp-loc_rel', se figur 4.14.

$$\begin{bmatrix} \texttt{before\_rel} & \\ \text{LBL} & \text{h17} \\ \text{ARG0} & \text{e19} \\ \text{ARG1} & \text{x18} \\ \text{ARG2} & \text{x8} \end{bmatrix} \begin{bmatrix} \texttt{numbered\_hour\_rel} & \\ \text{LBL} & \text{h31} \\ \text{ARG0} & \text{x7} \\ \text{CARG} & \text{8} \end{bmatrix} \begin{bmatrix} \texttt{card\_rel} & \\ \text{LBL} & \text{h2} \\ \text{ARG0} & \text{e4} \\ \text{ARG1} & \text{x3} \\ \text{CARG} & \text{5} \end{bmatrix}$$

$$\begin{bmatrix} \texttt{measure\_rel} & \\ \text{LBL} & \text{h25} \\ \text{ARG0} & \text{e26} \\ \text{ARG1} & \text{e19} \\ \text{ARG2} & \text{x3} \end{bmatrix} \begin{bmatrix} \texttt{temp-loc\_rel} & \\ \text{LBL} & \text{h6} \\ \text{ARG0} & \text{e9} \\ \text{ARG1} & \text{x8} \\ \text{ARG2} & \text{x7} \end{bmatrix}$$

Figur 4.14: De sentrale EPene for *fem på åtte*

MRS-representasjonen i 4.14 korresponderer med en digital representasjon tilsvarende *07.55*, og en slik digital representasjon vil være en ønskelig omformulering av *fem på åtte*. Dersom det foreligger flere modifikatorer i klokkeslettsuttrykket vil det derimot ikke være så elementært. I setningen *Hesten spiser like etter fem minutter før ti på halv ni* (figur 4.12) foreligger det fire modifikasjoner i forhold til klokketimen. Det ville vært ønskelig om alle de nøyaktige modifikasjonene ble slått sammen til én EP som uttrykker minutt-relasjonen, fordi vi ønsker å kunne omformulere klokkeslettsuttrykket til *like etter 8.15*. Alle vage tidsrom må uttrykkes med egne EP-relasjoner, men alle de nøyaktige kan slås sammen til én 'measure_rel'-relasjon som definerer antall minutter/sekunder som skal trekkes fra den hele klokketimen.[13] En tenkt gullstandard for klokkeslettsuttrykket *fem minutter før ti på halv ni* demonstreres i figur 4.15.

$$
\begin{bmatrix} \texttt{before\_rel} & \\ \text{LBL} & \text{h16} \\ \text{ARG0} & \text{e18} \\ \text{ARG1} & \text{x??} \\ \text{ARG2} & \text{x??} \end{bmatrix}
\begin{bmatrix} \texttt{numbered\_hour\_rel} & \\ \text{LBL} & \text{h??} \\ \text{ARG0} & \text{x??} \\ \text{CARG} & 9 \end{bmatrix}
\begin{bmatrix} \texttt{measure\_rel} & \\ \text{LBL} & \text{h24} \\ \text{ARG0} & \text{e25} \\ \text{ARG1} & \text{e18} \\ \text{ARG2} & \text{x10} \end{bmatrix}
$$

$$
\begin{bmatrix} \texttt{card\_rel} & \\ \text{LBL} & \text{h9} \\ \text{ARG0} & \text{e??} \\ \text{ARG1} & \text{x10} \\ \text{CARG} & 45 \end{bmatrix}
\begin{bmatrix} \texttt{minutt\_rel} & \\ \text{LBL} & \text{h9} \\ \text{ARG2} & \text{x10} \end{bmatrix}
$$

Figur 4.15: De sentrale EP-ene i en gullstandard for *fem minutter før ti på halv ni*

I figur 4.15 ser vi at alle de nøyaktige modifikasjonene blir slått sammen til en measure-relasjon. Dersom vi skal oppnå en MRS-representasjon lik den i figur 4.15 må det bygges en automat som slår sammen disse modifikasjonene, og dette arbeidet burde ikke være altfor krevende.

4.8 Oppsummering og diskusjon

I dette kapitlet har jeg gått igjennom selve implementeringen av klokkeslettsgrammatikken i XLE. Jeg har basert meg på det teoretiske grunnlaget jeg presenterte i kapittel 3. I avsnitt 4.2 til avsnitt 4.6 er XLE-grammatikken som forholder seg til c- og f-strukturen drøftet. Jeg har ivaretatt de restriksjonene jeg pekte på i tabell 3.1 (side 44) i forhold til preposisjonene som kan stå som hode i klokkeslettsuttrykk. I avsnitt 4.7 har jeg gjennomgått hvordan MRS-representasjonen for de forskjellige analysene ser ut. I til-

[13] Dersom det foreligger modifikasjoner spesifisert både i minutter og sekunder, må imidlertid to measure-relasjoner innføres.

legg til at jeg har forholdt meg til restriksjonene jeg skisserte i tabell 3.1, har jeg også tillatt den type topikalisering som beskrives i avsnitt 3.2.3.

Topikalisering og framflytting

Når klokkeslettsgrammatikken inkorporeres med NorGram, må topikaliseringen tilpasses mer komplekse setningstyper dersom dette er ønskelig.[14]

(74) Halv åtte sa Anne at klokken var fem minutter over.

Dersom (74) skal parses i NorGram, må det åpnes for at vi ikke vet hvor mange ledd som kan stå mellom den topikaliserte delen av klokkeslettsuttrykket, og hodet til dette leddet. Dette kalles funksjonell usikkerhet. Ved å tillate dette for klokkeslettsuttrykk, vil (74) parses som grammatisk på lik linje med en setning som:

(75) Avisen sa Kari til Per at hunden hadde spist.

Restriksjoner preposisjonen stiller

XLE-grammatikken ivaretar ikke alle de kravene jeg har pekt på at de forskjellige preposisjonene stiller til den omkringliggende syntaksen. De restriksjonene preposisjonene stiller til verbets tempus i setningen, er ikke inkorporert i min XLE-grammatikk. Jeg følger her NorGrams analyser, hvor det tillates at for eksempel preposisjonen *siden* kan innlede tidsadverbialer uten å stille restriksjoner til det finitte verbets tempus:

(76) *Hesten spiser middag siden halv åtte.

Setning (76) tillates i min grammatikk, på samme måte som (77) tillates i NorGram-grammatikken:

(77) *Jeg vasker opp siden jul.

Å definere disse tidadverbialene slik at de krever at verbets tempus er perfektum vil være et enkelt tillegg i min minigrammatikk, men jeg vet ikke hvordan dette vil gripe inn i NorGram-grammatikken, og har latt være å legge denne restriksjonen inn.

[14] Det må også tas stilling til om man *ønsker* å åpne for en slik topikalisering, forekomster som i (74) vil utvilsomt være svært sjeldne.

Kapittel 5

Evaluering og konklusjon

I dette kapitlet vurderer jeg det ferdige resultatet av klokkeslettsgrammatikken og antyder mulige utvidelser og forbedringer.

5.1 Pragmatiske valg

De mest optimale valgene fra et lingvistisk ståsted kan komme i konflikt med de reglene som er praktiske for en automatisk parser. I særlig grad gjelder dette balansegangen mellom hvor fri grammatikken skrives, i forhold til hvor mange potensielle løsninger dette kan generere. I avsnittet *Subjekt som hendelse* (4.2.2) diskuterte jeg restriksjonen som krever at dersom subjektet *klokke* viser til en hendelse, må CLOCK-OCCURRENCE +_ instansieres for at subjektet skal tolkes dithen. Dette innebar at setning (1) ikke vil få den intenderte analysen:[1]

(1) Auksjonen på bildet er klokka tre, sofaen er halv fire, og klokka er halv fem.

Dette er et typisk eksempel på vurderinger som kan tas for å få ned antall mulige analyser, og dermed uønsket flertydighet. Lingvistisk robusthet blir ofret fordi vi ikke ønsker to mulige analyser av de sannsynligvis mest frekvente av alle klokkeslettsuttrykk, nemlig de tilfeller hvor tidspunktet på dagen uttrykkes som i setningen:

(2) Klokka er tre.

Problemer knyttet til flertydighet vil jeg drøfte videre i påfølgende avsnittet.

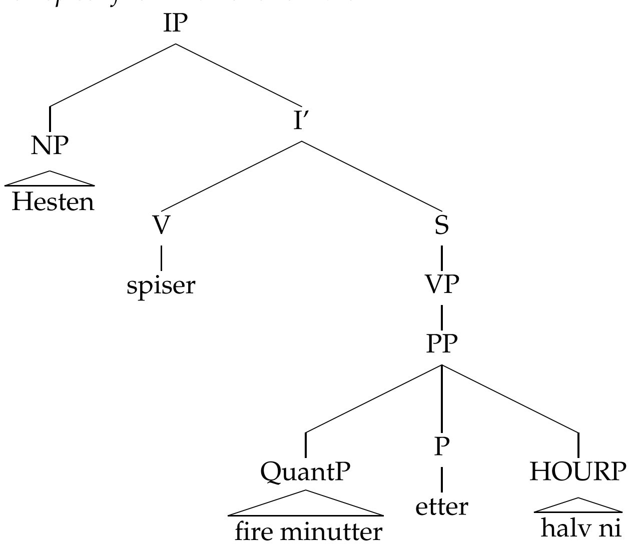
[1] CLOCK-OCCURRENCE +_ instansieres ikke siden *klokka/klokken* ikke står innad i klokkeslettsuttrykket, men bare på subjektsplass.

5.1.1 Flertydighet

Som i all parsing av naturlige språk, får klokkeslettsuttrykkene ofte uønskede analyser i tillegg til den intenderte. Det vil alltid være et mål å holde flertydigheten så langt nede som mulig, men de virkemidlene som kan tas i bruk for å redusere antall mulige løsninger vil også medføre at robustheten til parseren går ned. Dette kan igjen føre til at mer uvanlige uttrykksmåter ikke blir parset, selv om de er grammatiske.

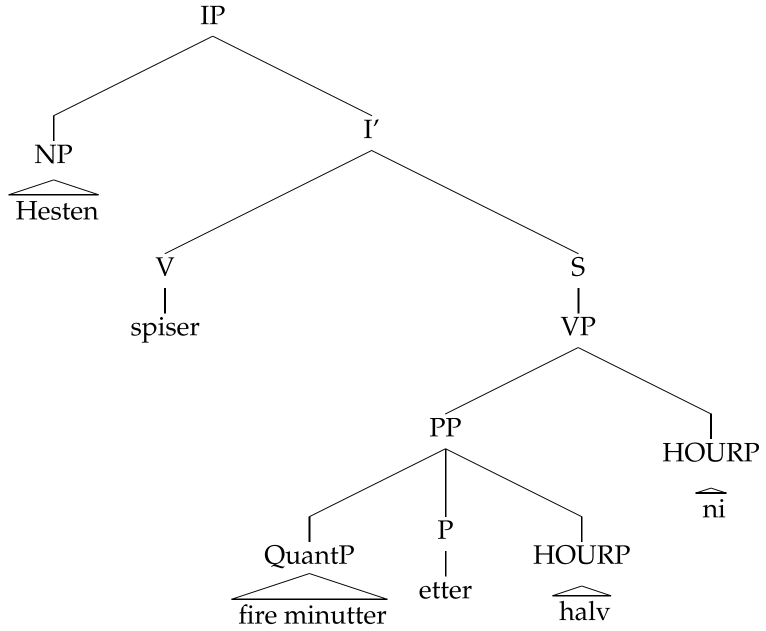
Det viktigste elementet som skaper flertydighet i min grammatikk, er at ett klokkeslettsuttrykk tolkes som flere forskjellige klokkeslett. Dette kan være et uttrykk som i setningen *Hesten spiser fire minutter etter halv ni*. I min grammatikk vil denne setningen få tre forskjellige løsninger. Den ønskede løsningen illustreres i (3).

(3) *Hesten spiser fire minutter etter halv ni*

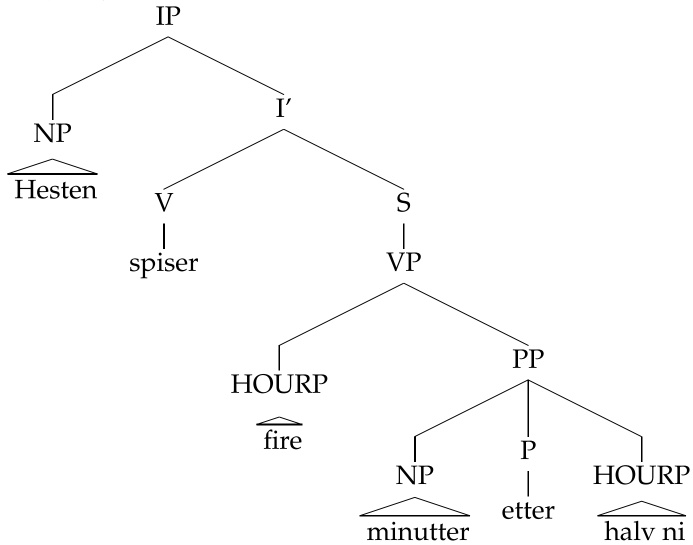


Siden et tallord i seg selv kan angi et klokkeslett og være et selvstendig klokkeslettsuttrykk, vil de to tallordene i (3) også kunne tolkes som et klokkeslett, i tillegg til den øvrige delen av klokkeslettsuttrykket *fire minutter etter halv ni*. Derfor vil vi også få (4) og (5) som analyse av setningen i (3).

(4) *Hesten spiser fire minutter etter halv ni*



(5) *Hesten spiser fire minutter etter halv ni*



Det er selvfølgelig mulig å legge inn restriksjoner slik at man unngår å få (4) og (5) som mulige analyser. For eksempel kan det kreves at i den grammatiske funksjonen ADJUNCT kan et tallord i seg selv ikke definere et klokkeslett dersom ikke *klokka/klokken* står foran dette tallordet, som i (6). Dermed vil heller ikke tallordet i (7) bli parset som et klokkeslettsuttrykk.

(6) Hesten spiser klokka ni.

(7) Hesten spiser ni.

Dette kan imidlertid være en ulempe i andre sammenhenger, i og med at et klokkeslett særlig i muntlige og uformelle sammenhenger ofte uttrykkes nettopp som i (7).

Dersom parsingresultatene skal rangeres i forhold til hvor sannsynlige de er, burde det la seg gjøre å implementere en automat som rangerer de lengste klokkeslettsuttrykkene (målt i antall ord) foran de resultatene som inneholder flere klokkeslettsuttrykk med færre ord. Dette vil gjøre at de mest sannsynlige parsingresultatene blir foretrukket. For min grammatikk må dette i så fall gjøres etter at mitt tillegg er tilpasset NorGrams øvrige grammatikk.

5.2 Utvidelser

5.2.1 Digitale representasjoner

Målet for masteroppgaven var å bygge en grammatikk som skulle analysere vanlige norske klokkeslettsuttrykk. Siden mitt utgangspunkt var de muntlige uttrykksformene, har jeg ikke gjort forsøk på å inkorporere digitale uttrykk, altså uttrykk som anvender tall. Dette kan være uttrykk hvor hele klokkeslettet er skrevet med tall (8), eller bare deler av det (9).

(8) 11.30

(9) klokka halv 12

Den digitale representasjonen i (8) virker å være den mest vanlige måten å uttrykke et klokkeslett på i korpusene jeg har undersøkt. Selve syntaksen kan variere, men dette beror hovedsaklig på hvilke tegnsettinger som foretrekkes, eller om *klokka/klokken* blir forkortet eller ikke:

(10)
- kl. 8.30
- klokka 8.30
- klokka 0830
- kl. 8:30
- osv.

Å lage en automat som gjenkjenner digitale klokkeslettsuttrykk burde ikke by på særlig lingvistiske utfordringer, men vil hovedsaklig være et arbeid som består av å undersøke hvilke variasjoner som benyttes innen digitale representasjoner, og lage en algoritme som gjenkjenner disse. Deretter bør uttrykket få en MRS-representasjon som følger prinsippene som benyttes for de bokstavelige klokkeslettsuttrykkene i denne oppgaven.

5.2.2 Andre grammatiske funksjoner

Klokkeslettsuttrykkene i min grammatikk kan bare stå i den grammatiske funksjonen PREDLINK eller ADJUNCT. Dette betyr at en setning som

(11) Halv åtte er en fin tid å stå opp på.

ikke kan parses slik min grammatikk ser ut i dag, siden klokkeslettet i (11) innehar den grammatiske funksjonen SUBJECT. Dersom det skal gjøres utvidelser hvor setninger som (11) blir parset, kan det tas utgangspunkt i min grammatikk. Det største problemet med å tillate at klokkeslettsuttrykk også står som subjekt, vil utvilsomt være at flertydigheten går i været. Alle enkeltstående tallord fra ett til tolv vil da få en analyse hvor de tolkes som et klokkeslettsuttrykk:

(12) Sju er et lykketall.

Dersom klokkeslettsuttrykk skal tillates på subjektsplass, kan mine regler for klokkeslettsuttrykkene gjenbrukes. Visse modifikasjoner må nok påregnes, siden det sikkert vil være noen syntaktiske forskjeller mellom klokkeslettsuttrykk i ADJUNCT og klokkeslettsuttrykk på subjektsplass.

5.2.3 Sammenslåing av kategorier

Spesifikatorene til PP

Når klokkeslettsgrammatikken blir integrert i NorGram, vil det være mulig å slå sammen kategorier som fungerer som spesifikatorer i PP. Dette gjelder de spesifikatorene som kalles av variabelen PP-TIME-MOD[2]. Både gradsadverbial og kvantorfraser tillates som spesifikator til en preposisjonsfrase i NorGram, mens variabelen PP-MOD-TIME kun tillater et subsett av disse spesifikatorene å stå internt i preposisjonsfrasen. I og med at PP-MOD-TIME-variabelen bare kan appliseres dersom PP er et klokkeslettsuttrykk, vil ikke dette være en kilde til flertydighet for preposisjonsfraser som ikke er et klokkeslettsuttrykk. For et uttrykk som:

(13) Påsken kommer noen måneder etter jul.

vil ikke PP-MOD-TIME-variabelen benyttes, siden denne krever at trekket CLOCK-EXPRESSION + er til stede i f-strukturen. Derimot kan klokkeslettsuttrykkene i noen tilfeller få to løsninger, hvor spesifikatoren til PP kan velges enten fra PP-MOD-TIME-disjunksjonen, eller fra en av reglene som allerede foreligger i PP til NorGram. Dette vil gjelde uttrykket:

(14) Hesten spiser middag like etter halv ni.

[2]Se regel (34), kapittel 4.

Klokkeslettsuttrykket i (14) vil komme til å få to analyser dersom min grammatikk uten videre legges til NorGram. *Like* vil både kunne analyseres av PP-MOD-TIME og av en generell NorGram-regel som tillater gradsadverbial å stå som spesifikator for PP. Ideelt sett bør PP-MOD-TIME omarbeides til å inngå i en større generell regel som definerer hvilke spesifikatorer PP kan ta. Eventuelt kan det skilles ut en egen variabel som definerer temporale spesifikatorer, slik at vi får en sammenslåing av spesifikator-regelen som kaller *noen timer* i (15) og (16):

(15) Jeg kommer noen timer etter tolv.

(16) Jeg kommer noen timer etter middag.

Arbeidet med å slå sammen disse kategoriene må jeg imidlertid overlate til den som tar på seg oppgaven å integrere min grammatikk i NorGrams, i og med at det er vanskelig for meg å vite nøyaktig hvordan den overordnede NorGram-grammatikken vil reagere på slike sammenslåinger.

PP-TIME og PP

Mye av fundamentet for min klokkeslettsgrammatikk ligger i bruken av PP-TIME-frasen. Denne frasen definerer det idiomatiske skjellettet for klokkeslettsuttrykk hvor *på* (i betydningen *før*) og *over* (i betydningen *etter*) står som hode. Preposisjonen *på* forandrer meningsinnhold internt i PP-TIME i forhold til i en vanlig PP, mens *over* også kan bety *etter* i PP. I PP-TIME får selve uttrykket en idiomatisk karakter, særlig tydelig er dette i et klokkeslettsuttrykk som *fem på ni*, hvor spesifikatoren her skal tolkes som *fem minutter*.

For å fange inn idiomatiske uttrykksformer knyttet til klokkeslettsuttrykk med *over* og *på*, har det vært hensiktsmessig å skille disse uttrykkene fra konvensjonelle preposisjonsfraser der det ikke foreligger noen idiomatiske konstruksjoner, som for klokkeslettsuttrykket (17).

(17) ti minutter etter tolv

Det vil i (17) være likegyldig om *tolv* byttes ut med *møtet*, noe som ikke er tilfelle i uttrykk (18).

(18) ti over tolv

Det kan imidlertid diskuteres om skillet mellom PP og PP-TIME er ønskelig fra en rent lingvistisk synsvinkel. Syntaktisk følger PP-TIME konvensjonene til en klassisk PP ved at den tar et objekt og spesifikator. Skillet mellom PP-TIME og PP er snarere semantisk motivert, i og med at *på* og *over* her opptrer som hode i mer eller mindre idiomatiske konstruksjoner.

De syntaktiske forskjellene mellom PP-TIME og PP går først og fremst ut på at PP-TIME er langt mer kresen på hvilke spesifikatorer som tillates, og at objektet må være en kategori fra HOUREXP-disjunksjonen (se regel (24), kapittel 4) eller være fortrengt. Ut ifra dette kan det argumenteres for at PP og PP-TIME kan slås sammen, og dagens PP-TIME-analyser kan utgjøre et subsett av de analysene PP vil gjøre. Et klart argumentet for en sammenslåing av PP-TIME og PP ble illustrert i avsnitt 4.4.2, hvor jeg diskuterte bakgrunnen for at klokkeslettsuttrykket *ti minutter over åtte* får to analyser. Og ut ifra min grammatikk blir det umulig å avgjøre hvilken av de to løsningene som er den mest korrekte eller ønskelige. Denne problemstillingen hadde vi sluppet dersom hele PP-TIME-frasen var integrert i PP.

Etter mitt syn gir oppdelingen i PP-TIME og PP et bedre grunnlag for å forstå syntaksen til klokkeslettsuttrykk. Det er en viktig forskjell mellom uttrykkene i PP-TIME (som eksempel (18)) og generelle PP-fraser som tar et klokkeslettsuttrykk som objekt (17). Særlig siden mitt arbeid bare er en isolert minigrammatikk som senere skal integreres i større grammatikker, føler jeg det er viktig å være tydelig på det klare skillet som må trekkes mellom egenskapene til klokkeslettsuttrykk med PP-TIME som toppnode i forhold til klokkeslettsuttrykk i en vanlig PP.

Selv om jeg har antydet at det rent lingvistisk kan være å foretrekke at PP-TIME og PP slås sammen, kan det være datalingvistiske argumenter som taler mot en slik sammenslåing. PP-regelen vil få en enorm mengde restriksjoner dersom den skal parse like mange typer klokkeslettsuttrykk som min grammatikk i dag, og samtidig være i stand til å skille ut uønskede klokkeslettslignende uttrykk som ikke har korrekt syntaks. Dette vil være komplisert å implementere, og det vil sannsynligvis også gjøre PP til en større flaskehals enn det den allerede er i NLP[3]. Ut ifra dette vil jeg ikke si at en sammenslåing av PP-TIME og PP vil være den beste løsningen for NorGram og LOGON-prosjektet, men det vil absolutt være interessant å se hvordan PP-regelen vil reagere dersom alle klokkeslettsuttrykk med preposisjon vil bli integrert i denne. PP-TIMEs skjebne bør altså avgjøres først når det er kartlagt hvordan en sammenslåing av PP og PP-TIME vil innvirke på preposisjonsfrasen generelt.

5.3 Konklusjon

Problemstillingen for denne masteroppgaven var å lage en selvstendig XLE-grammatikk som parset bokstavelige klokkeslettsuttrykk[4]. I dette arbeidet har jeg sett på den generelle syntaksen for klokkeslettsuttrykk og belyst faktorer som virker inn i denne. Dette kan være inherente egenskaper hos

[3] Natural Language Processing, *Prosessering av naturlig språk.*

[4] Klokkeslettsuttrykk som transkriberes ord for ord tilsvarende den muntlige uttrykksformen.

preposisjonene, eller klokkeslettsuttrykkets grammatiske funksjon i setningen.

Klokkeslettsuttrykkene har fått en MRS-representasjon som gir et godt utgangspunkt for det videre arbeidet med oversettelsen til engelsk. MRSen som min grammatikk produserer, bør postprosesseres av en automat slik at den endelige MRS-projeksjonen ligner den gullstandarden jeg skisserte i avsnitt 4.7.5. Når dette skjer, vil transferdelen til LOGON-prosjektet motta en innmating som ikke burde være vanskelig å oversette til den ønskede engelske representasjonen av klokkeslettet. I og med at MRS-representasjonen ligger nært opptil de prinsippene LinGO ERG-grammatikken (CSLI LinGO Lab, URL) benytter for klokkeslettsrepresentasjoner, burde arbeidet med transferreglene med være uproblematisk.

Det største problemet ved parsing av kompliserte klokkeslettsuttrykk er flertydigheten. Denne utfordringen er det største hinderet for all NLP i dag. Jeg har i avsnitt 5.1.1 skissert mulige modifiseringer som kan redusere antall analyseresultater, men dette vil gå utover robustheten til grammatikken. De mest frekvente klokkeslettsuttrykkene vil være de enkleste og mest ukompliserte, og for disse uttrykkene vil ikke problemet med flertydighet være stort. De største utfordringene med hensyn til flertydighet vil imidlertid komme etter at klokkegrammatikken er integrert i NorGram.

Mitt viktigste bidrag for å redusere flertydigheten har vært å gjøre parseren dyktig til å ekskludere uttrykk som ikke er grammatiske klokkesletts-uttrykk, på tross av at de har en lignende syntaks. Dette gjelder for eksempel setning (19) og (20), hvor bare (20) tolkes som et klokkeslettsuttrykk.

(19) fem fra ni

(20) fem over ni

Minigrammatikken jeg har laget er i stand til å parse de aller fleste former for bokstavelige klokkeslettsuttrykk, samtidig som den ikke parser klokkeslettlignende uttrykk som ikke har korrekt syntaks. Grammatikken vil være enkel å modifisere med tanke på eventuelle endringer, eller implementering i andre grammatikker. Ved implementering i NorGram, bør også en automat for gjenkjenning av digitale klokkeslettsuttrykk legges til. Dersom dette blir gjort, vil LOGONs parser være meget godt rustet for å parse alle typer komplekse og enkle klokkeslettsuttrykk.

Referanser

Erik Andersson. *Verbfrasens struktur i svenskan*. Åbo Akademi, Åbo, 1977.

Emily M. Bender, Ivan A. Sag, og Thomas Wasow. *Syntactic Theory: A Formal Introduction*. CSLI Publications, Stanford, California, andre utgave, 2003. Tilgjengelig på: `http://hpsg.stanford.edu/book/`.

Joan Bresnan. *Lexical-Functional Syntax*. Blackwell, Oxford, England, 2001.

Joan Bresnan og Ronald M. Kaplan. Lexical-functional grammar: A formal system for grammatical representation. I J. Bresnan, redaktør, *The Mental Representation of Grammatical Relations*, side 173–281. MIT Press, Cambridge, Massachusetts, 1982.

Miriam Butt, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, og Christian Rohrer. The Parallel Grammar Project. 2002. Tilgjengelig på: `http://www2.parc.com/istl/groups/nltt/pargram/buttetal-coling02.pdf`.

Miriam Butt, Tracy Holloway King, Marma-Eugenia Nino, og Fridirique Segond. *A Grammar Writer's Cookbook*. Center for the Study of Language and Inf, Stanford, California, 1999.

John Carroll, Ann Copestake, Dan Flickinger, og Victor Poznanski. An efficient chart generator for (semi-)lexicalist grammars. I *Proceedings of the 7th European Workshop on Natural Language Generation (EWNLG'99)*, side 86–95, Toulouse, Frankrike, 1999.

Noam Chomsky. *Syntactic Structures*. Mouton and Co, The Hague, 1957.

Noam Chomsky. Remarks on nominalization. I Roderick Jacobs og Peter Rosenbaum, redaktører, *Readings in English transformational grammar*, side 184–221. Blaisdell, Waltham (Mass.), 1970.

Ann Copestake, Dan Flickinger, Ivan Sag, og Carl Pollard. Minimal recursion semantics: An introduction. Tilgjengelig på: `www-csli.stanford.edu/~aac/papers/newmrs.pdf`. 2003.

Dick Crouch, Mary Dalrymple, Ron Kaplan, Tracy King, John Maxwell, og Paula Newman. *XLE Documentation*. Palo Alto Research Center [oppsøkt 18.12.2005], 2002-2005. Tilgjengelig på: `http://www2.parc.com/istl/groups/nltt/xle/doc/xle_toc.html`.

CSLI LinGO Lab. *LinGO English Resource Grammar*. CSLI LinGO Lab [oppsøkt 20.12.2005]. Tilgjengelig på: `http://lingo.stanford.edu:8000/erg`.

Mary Dalrymple. *Lexical-Functional Grammar (Syntax and Semantics, Volume 34) (Syntax and Semantics)*. Academic Press, London, England, 2001.

Helge Dyvik. Norsk komputasjonell grammatikk. Prosjektbeskrivelse. 1999. Tilgjengelig på: `http://www.hf.uib.no/i/LiLi/SLF/Dyvik/norgram/norgram-beskrivelse.html`.

Helge Dyvik. From LFG-lexicon to MRS-relations. Tilgjengelig på: `http://share.emmtee.net/bscw/ bscw.cgi/d4032-2/*/*/From(internt dokument)`. 2003.

Jan Terje Faarlund, Svein Lie, og Kjell Ivar Vannebo. *Norsk referansegrammatikk*. Universitetsforlaget, Oslo, tredje utgave, 1997.

Yehuda N. Falk. *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. CSLI Publications, Stanford, California, 2001.

Dan Flickinger. On building a more efficient grammar by exploiting types. I *Natural Language Engineering, 6 (1) (Special Issue on Efficent Processing with HPSG*, side 15–28, 2000.

Anette Frank. From parallel grammar development towards machine translation. I *Proceedings of MT Summit VII*, 1999.

Ray Jackendoff. *X′ Syntax: A Study of Phrase Structure*. MIT Press, Cambridge, Massachusetts, 1977.

Marit Ingebjørg Landrø og Boye Wangensteen, redaktører. *Bokmålsordboka*. Kunnskapsforlagets blå språk- og ordboktjenester [oppsøkt 20.12.2005], Oslo, Norge. Tilgjengelig på: `http://norskordbok.no`.

LOGON Consortium. Logon workplan, 2005-2006. Tilgjengelig på: `http://share.emmtee.net/bscw/bscw.cgi/d35361/plan.dec-04.pdf (internt dokument)`. 2004.

John Maxwell. *XLE*. Palo Alto Research Center [oppsøkt 18.12.2005], 2002-2005. Tilgjengelig på: `http://www2.parc.com/istl/groups/nltt/xle/`.

NA. Norsk aviskorpus. Tilgjengelig på: `http://avis.uib.no/`.

Stephan Oepen, Helge Dyvik, Jan Tore Lønning, Erik Velldal, Dorothee Beermann, John Carroll, Dan Flickinger, Lars Hellan, Janne Bondi Johannessen, Paul Meurer, Torbjørn Nordgård, og Victoria Rosén. Som å kapp-ete med trollet? Towards MRS-based Norwegian-English Machine Translation. I *TMI:04*, Baltimore, MD, October 2004. Tilgjengelig på: `http://share.emmtee.net/pub/bscw.cgi/d23044/tmi04.pdf`.

OK. Oslo-korpuset av taggede norske tekster. Tilgjengelig på: `http://avis.uib.no/`.

John I. Saeed. *Semantics*. Blackwell Publishing, Oxford, England, andre utgave, 1997.

Robert D. Van Valin. *An Introduction to Syntax*. Cambridge University Press, Cambridge, England, 2001.

Kjell Ivar Vannebo. *Tempus og tidsreferanse : tidsdeiksis i norsk*. Novus Forlag, Oslo, Norge, 1979.

Zeno Vendler. *Linguistics in Philosophy*. Cornell University Press, Ithaca, New York, 1967.

Wolfgang Wahlster, redaktør. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer, Berlin, 2000.

Tillegg A

Trekkene i XLE-grammatikken

Her følger en oversikt over hvilke trekk min grammatikk innfører i f-strukturen. Alle trekkene vil bli innført på et nivå i f-strukturen til klokkeslettsuttrykket, unntatt CLOCK-EVENT +, som legges til i f-stukturen til subjektet i setningen. Trekkene blir anvendt som restriksjoner og krav i frasestrukturreglene. Hvordan de konkret griper inn i grammatikken, er gjennomgått i kapittel 4.

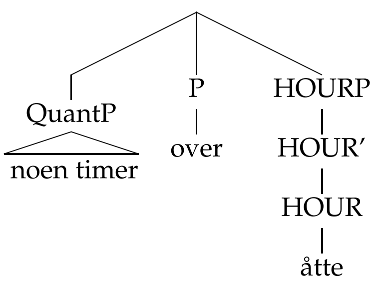
Trekkene er sortert etter opphavsstedet for hvor de introduseres til f-strukturen. Dette er enten fra leksikon eller frasestrukturreglene.

A.1 Leksikon

NTYPE CLOCK-MODIFIER +

Templatet @CLOCK-MODIFIER legger dette trekket til i nomenets f-struktur. Alle nomen med dette trekket kan opptre i disjunksjonen PP-TIME-MOD som spesifikator i PP.

(1) *noen timer over åtte*



(1)
$$
\left[\begin{array}{ll}
\text{PRED} & \text{'over}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\
\text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{'hour-8'}\\ \text{CLOCK-EXPRESSION} & +\end{array}\right] \\
\text{PTYPE-CLOCK} & \left[\text{TEMP} +\right] \\
\text{CLOCK-EXPRESSION} & + \\
\text{ADJUNCT} & \left\{\left[\begin{array}{l} \text{PRED 'time'} \\ \text{SPEC} \left[\text{QUANT}\left[\text{PRED 'noen'}\right]\right] \\ \text{TEMP-SPEC} + \\ \left[\text{NTYPE}\left[\text{CLOCK-MODIFIER} +\right]\right]\end{array}\right]\right\}
\end{array}\right]
$$

PTYPE-CLOCK TEMP +

Alle preposisjoner som kan være i hode i et klokkeslettsuttrykk i PP har dette trekket. Dette gjelder: *siden, innen, til, fra, etter, før* og *over*.

PTYPE-CLOCK SUBTEMP +

Alle preposisjoner med subjunksjonsegenskaper (se avsnitt 3.3.2) har dette trekket i f-strukturen i tillegg til PTYPE TEMP +. Dette gjelder: *fra, innen, siden* og *til*.

PTYPE-CLOCK TEMPFINALE +

Preposisjonene som kan stå som hode i PP-TIME har dette trekket. Dette er preposisjonene *på* og *over*.

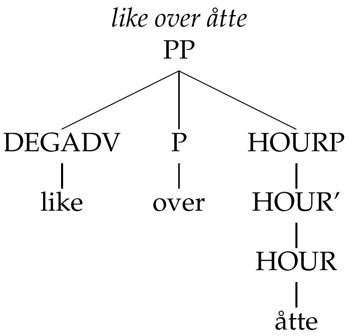
CLOCK-OCCURRENCE +_

Dette er et instansiert trekk. Innen f-strukturen til klokkeslettsuttrykket vil trekket sipre til alle nivå dersom det blir introdusert. Når det instansieres én gang, kan det ikke instansieres på nytt. Derfor regulerer restriksjonen forekomsten av ordet *klokka/klokken* og sørger for at det ikke forekommer flere ganger enn det er ønskelig (null eller én gang). Trekket legges inn fra leksikonoppslagene til N-CLOCK-kategorien *klokka/klokken*.

ATYPE CLOCK +

Dette trekket markerer at gradsadverbialet kan stå som klokkemodifikator i disjunksjonen PP-TIME-MOD. Gradadverbialet vil stå som spesifikator i PP i uttrykk som (2).

(2)



$$(2)\ \left[\begin{array}{ll}
\text{PRED} & \text{'over}\left\langle\left(\uparrow \text{OBJ}\right)\right\rangle\text{'} \\
\text{OBJ} & \left[\text{PRED 'hour-8'}\right] \\
\text{PTYPE-CLOCK} & \left[\text{TEMP +}\right] \\
\text{CLOCK-EXPRESSION +} & \\
\text{ADJUNCT} & \left\{\left[\begin{array}{l}\text{PRED 'like'} \\ \text{ATYPE}\left[\text{CLOCK +}\right] \\ \text{TEMP-SPEC +}\end{array}\right]\right\}
\end{array}\right]$$

A.2 Frasestrukturreglene

CLOCK-EVENT +

Dette trekket markerer at setningens subjekt viser til en hendelse (event). Trekket vil ligge i f-strukturen til subjektet (3).

(3) Kampen er åtte.

$$(3)\ \left[\begin{array}{ll}
\text{PRED} & \text{'være}\left\langle\left(\uparrow \text{SUB}\right)\left(\uparrow \text{PREDLINK}\right)\right\rangle\text{'} \\
\text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{'kamp'} \\ \text{CLOCK-EVENT} & +\end{array}\right] \\
\text{PREDLINK} & \left[\begin{array}{ll}\text{PRED} & \text{'hour-8'} \\ \text{CLOCK-EXPRESSION} & +\end{array}\right]
\end{array}\right]$$

CLOCK-PREDLINK +

I setninger hvor subjektet er *klokka/klokken/den* og klokkeslettsuttrykket har den grammatiske funksjonen PREDLINK, kan dette trekket legges til på øverste nivå i f-strukturen til klokkeslettsuttrykket. Dette indikerer at subjektet ikke refererer til en hendelse, men at hele uttrykket beskriver et klokkeslett (4).

(4) Klokka er åtte.

$$
(4)\quad \begin{bmatrix}
\text{PRED} & \text{'være}\left\langle\left(\uparrow\text{SUB}\right)\left(\uparrow\text{PREDLINK}\right)\right\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix}\text{PRED 'klokke'}\end{bmatrix} \\
\text{PREDLINK} & \begin{bmatrix}\text{PRED 'hour-8'} \\ \text{CLOCK-EXPRESSION +} \\ \text{CLOCK-OCCURRENCE -} \\ \text{CLOCK-PREDLINK +}\end{bmatrix}
\end{bmatrix}
$$

CLOCK-OCCURRENCE -

Disjunkt én og tre i TIMEEXP-variabelen[1] innfører trekket CLOCK-OCCURRENCE -. Når dette trekket innføres kan ikke *klokka/klokken* stå inne i et klokkeslettsuttrykk på grunn av en restriksjon i leksikonet til *klokka/klokken*. Trekket siprer til alle nivå i f-strukturen til klokkeslettsuttrykket når det blir instansisert, på linje med CLOCK-OCCURRENCE +_.

Hensikten med restriksjonen er å ivareta regelen om at *klokka/klokken* ikke blir anvendt internt i klokkeslettsuttrykket dersom hele setningen skal uttrykke et klokkeslett. I setning (5) kan subjektet *klokka* bare referere til en hendelse, og setningen vil på grunn av forekomsten *klokken* internt i klokkeslettsuttrykket ikke bli tolket som en referanse til hva klokken er i ytringstidspunktet.

(5) Klokka er klokken ni.

CLOCK-EXPRESSION +

Dette trekket blir introdusert av disjunksjonen TIMES, og siprer til alle nivå i klokkeslettsuttrykkets f-struktur. Alle klokkeslettsuttrykk har dette trekket, og det brukes for å markere hva som defineres som klokkeslettsuttrykk.

TEMP-SPEC +

Trekket legges til av disjunksjonen PP-TIME-MOD, og indikerer at en spesifikator er til stede foran hodet i PP. Dette for å vise at en temporal modifikator står foran hodet i PP. Trekket er blant annet avgjørende for å kunne parse et uttrykk som *Kampen er ti minutter fra halv ni* for ugrammatisk. Bruken av dette trekket demonstreres i eksempel (1).

[1]Se regel (12) og (50) i kapittel 4.

Tillegg B

Frasestrukturregler i XLE

Dette er et utvalg av frasestrukturreglene som brukes av XLE for å parse klokkeslettsuttrykkene. Jeg har utelatt de subleksikalske regelene, og fjernet annen informasjon som er unødvendig for å lese reglene.

```
----

TEST NORSK RULES (1.0)


IP --> { { NP: (^ SUBJ)=!
          | PRON: (^ SUBJ)=!}
         |PP*: @ADJUNCTPP
         |TIMEEXP: (! CLOCK-EXPRESSION)=c +
        | HOUR': (^ PREDLINK OBJ)=!
         |HALF':(^ PREDLINK OBJ)=!}
        I': @MRS-PRPSTN.

I' --> V: ^=!
       (! VFORM)=c fin;
       (S: ^=!).

S --> ({ NP: (^ SUBJ)=!
          | PRON: (^ SUBJ)=!}),
        (SADV: ! $ (^ ADJUNCT))
        (VP).

NP --> (D)
        A*:!$(^ ADJUNCT);
        N: ^=!;
        PP*: @ADJUNCTPP
```

```
            ~(! CLOCK-EXPRESSION)=+. "Can only be under VP."

"Unless fragments or the entire clock expression is topicalised,
the clock expression will appear as the daughter of VP."
"Clock expressions whose top node is a PP, and whose syntactic
function is ADJUNCT, will appear under PP"

VP --> (V: ^=!)
       (SADV: ! $ (^ ADJUNCT))
       ({ NP: (^ OBJ)=!
        | PRON: (^ OBJ)=!})
       PP*: @ADJUNCTPP,
            (TIMEEXP: (! CLOCK-EXPRESSION)=c +).

"All clock expressions which cannot be analysed under
PP*: !$(^ ADJUNCT) , will be analysed as one of these four
disjunction rules in TIMEEXP:"

TIMEEXP =
       { PP: { (^ SUBJ PRED FN)=c klokke
                (^ SUBJ DEF)=c +
              | (^ SUBJ PRON-FORM)=c den
                (^ SUBJ GEND NEUT)=c -}
                (^ PREDLINK)=!
                (! CLOCK-PREDLINK)=+
                (! CLOCK-OCCURRENCE)=-
                (! PTYPE-CLOCK TEMP)=c +
               ~(! PTYPE-CLOCK SUBTEMP)=+

      "For expressions where the subject refers to an event."
      | PP: (^ PREDLINK)=!
            (^ SUBJ CLOCK-EVENT)=+
            (! PTYPE-CLOCK TEMP)=c +
            { ~(^ SUBJ PRED FN)= klokke
             |(^ SUBJ PRED FN)= klokke
              (! CLOCK-OCCURRENCE)=c +_}

      | TIMES: { !$(^ ADJUNCT)
                 @MRS-CLOCKADJUNCT
                |(^ PREDLINK)=!
                 { (^ SUBJ PRED FN)=c klokke
                   (^ SUBJ DEF)=c +
                 | (^ SUBJ PRON-FORM)=c den
                    (^ SUBJ GEND NEUT)=c -}
```

```
                  (! CLOCK-PREDLINK)=+
                  (! CLOCK-OCCURRENCE)=-

                "For expressions where the subject refers
                 to an event."
                | (^ PREDLINK)=!
                  (^ SUBJ CLOCK-EVENT)=+
               {  ~(^ SUBJ PRED FN)= klokke
                | (^ SUBJ PRED FN)= klokke
                  (! CLOCK-OCCURRENCE)=c +_}}}.

"The disjunction category for the specifiers that can precede
the head of PP when PP is a clock expression."
PP-TIME-MOD =
       { DEGADV: !$(^ ADJUNCT)
                 "</mrs>" @MRS-DEGADVSYNT "<mrs/>"
                 (! ATYPE CLOCK)=c +

       | QuantP: !$(^ ADJUNCT)
                 @MRS-MINUTEMEASURE
                 (! NTYPE CLOCK-MODIFIER)=c +

       | NP: !$(^ ADJUNCT)
             @MRS-MINUTEMEASURE
             (! NUM)=c pl
             (! CHECK _DEF-MORPH)=c -
             (! NTYPE CLOCK-MODIFIER)=c +}.

PP-CLOCK =
       { PP | TIMES}.


TIMES = { PP-TIME: (! CLOCK-EXPRESSION)=+
          | HOURP: (! CLOCK-EXPRESSION)=+ }.


"PP-TIME is a preposition phrase that can only occur in clock
expressions. It is none-recursive."

PP-TIME -->
    (N-CLOCK: (^ CLOCK-OCCURRENCE)=c +_)
     TIME-MOD: (^ SPEC)=!
               @MRS-MINUTEMEASURE;
     P'-TIME: ^=!
```

```
              (^ PTYPE-CLOCK TEMPFINALE)=c +.

"The disjunction category for the specifiers that can precede
the head of PP-TIME."

TIME-MOD = { QuantPmin: { (! PRED FN)=c minutt
                        | (! PRED FN)=c sekund
                        | (! PRED FN)=c pro }
                  | QUART
                  | N: { (! PRED FN)=c minutt
                  | (! PRED FN)=c sekund }
                     m::!=%OBJECT
                     @(MRS-COVQUANT _få_q 3 pl) }.


"Only the prepositons 'på' and 'over' have the feature required
by P'-TIME."
P'-TIME -->
       P-TIME: ^=!
               (! PTYPE-CLOCK TEMPFINALE)=c +;
               { (HOUREXP:
                 (^ OBJ)=!)
               |e: (^ OBJ PRED)='pro'
                   m::(^ OBJ)=%PREP
                   @MRS-EMPTYHOUR}.

HOURP -->
       (N-CLOCK: (^ CLOCK-OCCURRENCE)=c +_)
        HOUREXP.

HALF' --> HALF: ^=!;
       { HOUR': (^ OBJ)=!
       | e: (^ OBJ PRED)='pro'
            m::(^ OBJ)=%PREP
             @MRS-EMPTYHOUR}.

HOUR' --> HOUR.

HOUREXP = { HOUR' | HALF' }.

QuantPmin --> { { Q: ^=!
                     (^ CHECK _PREDET)=+
         | NUMdig: ^=!
                  (^ CHECK _PREDET)=+
```

```
         | NUMP: (^ SPEC NUMBER)=!
                 @LITNUM
                 (^ CHECK _PREDET)=+ }
        NP: ~(^ DEF)=+
       |{ NUMdig: ^=!
                  (^ PRED)='pro'
                  "</mrs>" @(MRS-OBJECTGVN minutt)"<mrs/>"
       | NUMP: (^ PRED)='pro'
               (^ SPEC NUMBER)=!
               @LITNUM
               (s::M* arg1 ind given)=+
                "</mrs>"@(MRS-OBJECTGVN minutt)"<mrs/>"}}.

 NUMEXP = { NUM1|NUM11}.

"NORGRAM RULES WITH MODIFICATIONS:"

"The first two optional categories in PP are added in order
to parse clock expressions."

PP --> {
         (N-CLOCK: (^ CLOCK-OCCURRENCE)=c +_
                   (^ ADJUNCT TEMP-SPEC)=c +
                   (^ CLOCK-EXPRESSION)=c +)
         (PP-TIME-MOD: (^ CLOCK-EXPRESSION)=c +_
                      ~(^ PTYPE-CLOCK SUBTEMP)= +
                       (^ PTYPE-CLOCK TEMP)=c +
                       (! TEMP-SPEC)=+)
           |
             ( QuantP: !$(^ ADJUNCT)
                      ~(! DEF)=+
                       @COMPLETECHECKFEATS
                      { (! NTYPE NSEM TIME TEMPNOUN)=c +
                      | ~(! NTYPE NSEM TIME TEMPNOUN)=c +
                        (! NTYPE NSEM MEASURENOUN)=c + }
                      (^ CHECK _MSRBL)=c +
                      "</mrs>" @MRS-TEMPNP "<mrs/>")}
       { P: ^=!
            (^ CLOCK-EXPRESSION)=(^ OBJ CLOCK-EXPRESSION)
            (^ PTYPE)=c sem
            @OBJCASE
            s::M*=s::*
            s::M*=s::^
           (s::M* arg2)=s::(^ OBJ)
```

```
           @(COMPLETE (^ OBJ CHECK _PREPEXISTS))
           @(COMPLETE (^ OBJ CHECK _P-SELFORM))
           @(COMPLETE (^ PSEM))
           @(COMPLETE (^ PTYPE))
       | Pprd: ^=! "For the special case of PPs with PREDLINK"
              (^ PTYPE)=c sem
              @OBJCASE
              s::M*=s::*
              s::M*=s::^
             (s::M* arg2)=s::(^ OBJ)
             @(COMPLETE (^ OBJ CHECK _PREPEXISTS))
             @(COMPLETE (^ OBJ CHECK _P-SELFORM))
             @(COMPLETE (^ PSEM))
             @(COMPLETE (^ PTYPE)) }
       { NP: (^ OBJ)=!
             @COMPLETECHECKFEATS
            ~(! CHECK _P-SELFORM)
            { ~(! PRON-TYPE)=refl
            | (! PRON-TYPE)=c refl
              (^ CHECK _REFL)=c +}
            { ~(! PRON-TYPE)=refrefl
            | (! PRON-TYPE)=c refrefl
              @PPREFREFLEXIVE }
             @REFERENTIAL?
             @PPPERCOLATEANT
             "@REFERENTIAL"
     | PP: (^ OBJ)=!
           "(^ CHECK _MSRBL)=c +"
           @PPPERCOLATEANT
           "@REFERENTIAL"
     | e: (^ CHECK _MOVED-OBJ)=+
         ~(^ OBJ TOPCP)=-
         ~(^ OBJ CHECK _TOPVP)=-

"This disjunction will provide the OBJECT in a clock expression."
     | PP-CLOCK: (^ OBJ)=!
                 (^ CLOCK-EXPRESSION)=c +
                 (^ PTYPE-CLOCK TEMP)=c +
                 (^ CLOCK-OCCURRENCE)=(^ OBJ CLOCK-OCCURRENCE)}

       "To make sure that the object CP or VP' is not
        just a sister of PP under VP (in the case of PCOMP and PXCOMP)"
      ({ AP: (^ PREDLINK)=!
             (^ PREDLINK GEND)=(^ OBJ GEND)
```

```
            (^ PREDLINK NUM)=(^ OBJ NUM)
           ~(! CHECK _ADVERBIAL)=+
     | PP: (^ PREDLINK)=!
            { (! PFORM)=c på
            | (! PFORM)=c i }
            @EMBEDANTECEDENTFEATS}).

QuantP --> { { QUANTF: @DEFIFPREDEF "LOGONs QuantP-rule"
                        (^ CHECK _PREDET)=+
           | NUMP: (^ SPEC NUMBER)=!
                   @DEFIFPREDEF
                   @LITNUM
                   (^ CHECK _PREDET)=+ }
            " QUANTCOMP"
           NP: { (^ DEF)=c + s::M*=s::* "&&&Originally QUANTCOMP&&"
               | ~(^ DEF)=+ (s::M* arg1)=s::*
                 (s::M* sel)=(s::* sel) } ;"Make sure that selectional
           restrictions are available at top level"
           (CONJPcomp: @QUANTCOMPCLAUSE)
          | { QUANTF: (^ PRED)='pro'
                       @DEFIFPREDEF
                      (s::M* arg1 ind given)=+
                      "</mrs>" @MRS-PROOBJECT"<mrs/>"
            | NUMP: (^ PRED)='pro'
                    (^ SPEC NUMBER)=!
                    @DEFIFPREDEF
                    @LITNUM
                   (s::M* arg1 ind given)=+
                   "</mrs>"@(MRS-OBJECTGVN generic-nom)"<mrs/>"}
            (PPtil: @POSSPP2)
            (PPCMPL: @ADJUNCTPP
                      @(OT-MARK Mark6))
            ({ CPrel: @RESTREL
              |CPnullc: @RESTREL}
             (COMMA))
             (CONJPcomp: @QUANTCOMPCLAUSE)
           | { QUANTF: @DEFIFPREDEF
             | NUMP: ^=!
                    (^ REF)=+
                    @DEFIFPREDEF
                    @MRS-LITNUM }
             PP: @PARTITIVEPP1;
             (CONJPcomp: @QUANTPARTCOMPCLAUSE)}.
```

```
"NORGRAM RULES WITHOUT MODIFICATIONS:"



QP --> ( DEGADV: !$(^ SPEC AQUANT ADJUNCT)
                 (^ SPEC AQUANT DEGREE)=(! DEGREE)
                 (^ SPEC AQUANT DEG-DIM)=(! DEG-DIM)
                 "</mrs>" @MRS-DEGADVSYNT "<mrs/>")
       Q: s::M*=s::*.

   QUANTF = { QP|NUM|ART|NUMdigP }.

NUMP --> (DEGADVnum: ! $ (^ ADJUNCT)
          "</mrs>" @MRS-DEGADVnumSYNT "<mrs/>")
       NUMEXP: ^=!.

NUMdigP --> (DEGADVnum: ! $ (^ ADJUNCT)
                        "</mrs>" @MRS-DEGADVdigSYNT "<mrs/>")
        NUMdig: ^=!
                @(COMPLETE (^ CHECK _ARITH)).
```

Tillegg C

Leksikonet i XLE

Her følger et utvalg av leksikonet som benyttes av XLE. Oppslagsordene er sortert etter kategori.

```
TEST NORSK LEXICON (1.0)

[...]

"PRONOUNS:"

den     PRON * @PRONOUN
               @SG
               (^ PERS)=3
               @COMMON
               (s::M* sel human)=-; ETC.

det     PRON * { (^ PRON-FORM)=det
                 (^ PRON-TYPE)=expl_
                 (^ REF)=-
                 (^ NUM)=sg
                 (^ PERS)=3
                 @NEUT
                 ~(^ PRED) "To avoid postverbal SUBJ as well."
                 @COMPLETENOMFEATS
               | @PRONOUN
                 @SG
                 (^ PERS)=3
                 @NEUT
                (s::M* sel human)=-}; ETC.

"NOUNS:"
```

```
hest      N XLE @(COUNTNOUN hest horse).

kamp      N XLE @(COUNTNOUN kamp fight).

klokke    N XLE @(COUNTNOUN klokke clock); ETC.

middag    N XLE @(COUNTNOUN middag dinner).

"The nouns that follow here, are the nouns that may occur within
the phrase rule PP-TIME-MOD and TIME-MOD. These nouns have the
ability to single-handedly or as the head of a quantifier phrase
modify the clock expression. The nouns refer to a period of time.
The template @CLOCK-MODIFIER adds the feature NTYPE CLOCK-MODIFIER +
to the f-structure."


dag       N XLE @(COUNTNOUN dag day)
                @CLOCK-MODIFIER; ETC.

kvarter   N XLE @(COUNTNOUN kvarter quarter)
                @CLOCK-MODIFIER; ETC.

minutt    N XLE @(COUNTNOUN minutt minute)
                @CLOCK-MODIFIER; ETC.

sekund    N XLE @(COUNTNOUN sekund second)
                @CLOCK-MODIFIER; ETC.

stund     N XLE @(COUNTNOUN stund stund)
                @CLOCK-MODIFIER; ETC.

time      N XLE @(COUNTNOUN time hour)
                @CLOCK-MODIFIER; ETC.

uke       N XLE @(COUNTNOUN uke week)
                @CLOCK-MODIFIER; ETC.

"VERBS:"

spise     V XLE { @(V-SUBJ spise eat)
                | @(V-SUBJ-OBJ spise eat) }; ETC.

være      V XLE { @(COPULA være)
```

```
                  @DURATIVE
                  { @ACOPULA
                  | @ADVCOPULA
                  | @NCOPULA
                  | @COMPCOPULA
                  | @VPREDCOPULA
                  | @CLOCKCOPULA }
                | @PERFAUX
                  (^ XCOMP CHECK _AUX-SELECT)=c be
                | @(V-SUBJexpl-NCOMP-COMP være) "Clefts"
                  @DURATIVE
                | @(V-SUBJexpl-ACOMPorNCOMP-XCOMPorCOMP være)"Extrapositions"
                  @DURATIVE
                | @(V-SUBJexpl-ACOMP være)"Impers. adjectives"
                  @DURATIVE
                | @(V-SUBJ exist exist)
                 ~(^ TNS-ASP MOOD)=imperative
                  (^ VTYPE)=main
                  (^ PRESENTATIVE)=c +
                  (^ ADJUNCT)
                | (^ PRED)='pass<(^ XCOMP)> (^ SUBJ)'
                 ~(^ TNS-ASP MOOD)=imperative
                  (^ SUBJ)=(^ XCOMP SUBJ)
                  (^ XCOMP PASSIVE)=+
                  (^ XCOMP PASS-TYPE)=periph
                  (^ XCOMP VFORM)=c pastpart
                  (^ XCOMP CHECK _SUPINE)=+
                 ~(^ XCOMP NUM)=pl
                 ~(^ XCOMP DEF)=+
                  (^ VTYPE)=c aux
                  s::M*=s::(^ XCOMP)
                  @UNACCUSATIVE
                  "</mrs>"m::^ = m::(^ XCOMP)"<mrs/>"}
                  "</mrs>"@MRS-STATE"<mrs/>"
                  ~(^ FREEPRED); ETC.

"CATEGORIES USED ONLY BY CLOCK-EXPRESSIONS:"

halv      HALF * (^ PRED)='halv<(^ OBJ)>'
                 %ARG2=m::(^ OBJ)
                 @MRS-HALV.

klokka     N-CLOCK * (^ CLOCK-OCCURRENCE)=+_
                    ~(^ CLOCK-OCCURRENCE)=-.
```

```
klokken   N-CLOCK * (^ CLOCK-OCCURRENCE)=+_.

kvart       QUART * (^ PRED)='kvart'
                    @MRS-KVART.


"PREPOSITIONS:"

"The clock expression prepositions can be tagged with three
different features: PTYPE-CLOCK TEMP +, PTYPE-CLOCK SUBTEMP +
and PTYPE-CLOCK TEMPFINALE + "

etter     P * @(SEMPREP etter past)
              (^ PTYPE-CLOCK TEMP)=+.

fra       P * @(SEMPREP fra from)
              @MEASURABLEPREP
              (^ PTYPE-CLOCK TEMP)=+
              (^ PTYPE-CLOCK SUBTEMP)=+; ETC.

før       P * @(SEMPREP før before)
               (^ PTYPE-CLOCK TEMP)=+.

innen     P * @(SEMPREP innen until)
              (^ PTYPE-CLOCK TEMP)=+
              (^ PTYPE-CLOCK SUBTEMP)=+.

siden     P * @(SEMPREP siden since)
              (^ PTYPE-CLOCK TEMP)=+
              (^ PTYPE-CLOCK SUBTEMP)=+.

til       P * @(SEMPREP til to)
              (^ PTYPE-CLOCK TEMP)=+
              (^ PTYPE-CLOCK SUBTEMP)=+.

"The P-TIME-category can only by used within clock-expressions."

over      P-TIME * (^ PRED)='over<(^ OBJ)>'
                   (^ PTYPE-CLOCK TEMPFINALE)=+
                   @MRS-PAST;
          P * @(SEMPREP over over)
               { (^ PTYPE-CLOCK TEMP)=+
                 (^ CLOCK-EXPRESSION)=c +
```

```
               { (^ CLOCK-PREDLINK)=c +
               |
                 (^ ADJUNCT TEMP-SPEC)=c +}
               |
                ~(^ CLOCK-EXPRESSION)=+}.

på        P-TIME * (^ PRED)='på<(^ OBJ)>'
                   (^ PTYPE-CLOCK TEMPFINALE)=+
                   (^ CLOCK-EXPRESSION)=c +
                   @MRS-BEFORE;
          P * @(SEMPREP på på)
              ~(^ CLOCK-EXPRESSION)=+.

"ADVERB:"

ikke      SADV XLE (^ PRED)='neg'.

"Degree adverbs. All degree adverbs must be tagged
ATYPE CLOCK + in order to operate within clock expressions."

langt     DEGADV * @(DEGADV langt langt)
                   (^ ATYPE CLOCK)=+.

lenge     DEGADV * @(DEGADV lenge lenge)
                   (^ ATYPE CLOCK)=+.

lengre    DEGADV * (^ PRED)='lengre'.

like      DEGADV * @(DEGADV like like)
                   (^ ATYPE CLOCK)=+.

litt      DEGADV * @(DEGADV litt litt)
                  (^ ATYPE CLOCK)=+.

rett      DEGADV * @(DEGADV rett rett)
                   (^ ATYPE CLOCK)=+.

straks    DEGADV * @(DEGADV straks straks)
                   (^ ATYPE CLOCK)=+.

"QUANTIFIERS:"

flere     Q * (^ REF)=+
              (^ NTYPE NSEM COMMON)=count
```

```
              (^ PERS)=3
              (^ NUM)=pl
              (s::M* relation base)=many
              (s::M* relation degree)=more-than
              { (^ PRED)='mange<(^ OBL)>'
                (^ DEG-DIM)=pos
                (^ OBL OBJ NUM)=c pl
                { (%ADJCT PRED)='mer'
                  "</mrs>" @MRS-COMPARATIVE1ARG "<mrs/>"
              | (%ADJCT PRED)='mer<(%ADJCT OBL-COMPAR)>'
                "</mrs>" @MRS-COMPARATIVE2ARG "<mrs/>"}
              %ADJCT $ (^ ADJUNCT)
              "</mrs>"@(MRS-PARTQUANT noen_q)
               @(MRS-AQUANT mange)"<mrs/>"
             | (^ SPEC AQUANT PRED)='mange'
               (^ SPEC AQUANT DEG-DIM)=pos
               ~(^ SPEC QUANT)
               ~(^ CHECK _PREDEF)=c +
               (^ SPEC AQUANT DEGREE)=comparative
               { (%ADJCT PRED)='mer'
                 "</mrs>" @MRS-AQUANTCOMPARATIVE1ARG "<mrs/>"
               | (%ADJCT PRED)='mer<(%ADJCT OBL-COMPAR)>'
                 "</mrs>" @MRS-AQUANTCOMPARATIVE2ARG "<mrs/>"}
               %ADJCT $ (^ SPEC AQUANT ADJUNCT)
               "</mrs>" @(MRS-QUANT noen_q)
               @(MRS-AQUANT mange) "<mrs/>"}; ETC.

noen      Q * @NOEN
              { (^ PRED)='noen<(^ OBL)>'
                (^ OBL OBJ NUM)=c pl
(^ NUM)=pl
"</mrs>"@(MRS-PARTQUANT noen_q)"<mrs/>"
      | (^ SPEC QUANT QUANT-TYPE)=existential
(^ SPEC QUANT PRED)='noen'
"</mrs>"@(MRS-QUANT noen_q)"<mrs/>"
{ @PL
| @SG
  @COMMON }
(^ DEF)= - }; ETC.

"NUMERALS:"

ett       HOUR * @(NUMBEREDHOUR hour-1 1); ETC.
```

```
to        NUM1 * @(NUMFEATS 2);
          HOUR * @(NUMBEREDHOUR hour-2 2); ETC.

tre       NUM1 * @(NUMFEATS 3);
          HOUR * @(NUMBEREDHOUR hour-3 3); ETC.

fire      NUM1 * @(NUMFEATS 4);
          HOUR * @(NUMBEREDHOUR hour-4 4); ETC.

fem       NUM1 * @(NUMFEATS 5);
          HOUR * @(NUMBEREDHOUR hour-5 5); ETC.

seks      NUM1 * @(NUMFEATS 6);
          HOUR * @(NUMBEREDHOUR hour-6 6); ETC.

sju       NUM1 * @(NUMFEATS 7);
          HOUR * @(NUMBEREDHOUR hour-7 7); ETC.

syv       NUM1 * @(NUMFEATS 7);
          HOUR * @(NUMBEREDHOUR hour-7 7); ETC.

åtte      NUM1 * @(NUMFEATS 8);
          HOUR * @(NUMBEREDHOUR hour-8 8); ETC.

ni        NUM1 * @(NUMFEATS 9);
          HOUR * @(NUMBEREDHOUR hour-9 9); ETC.

ti        NUM11 * @(NUMFEATS 10);
          HOUR * @(NUMBEREDHOUR hour-10 10); ETC.

elleve    NUM11 * @(NUMFEATS 11);
          HOUR * @(NUMBEREDHOUR hour-11 11); ETC.

tolv      NUM11 * @(NUMFEATS 12);
          HOUR * @(NUMBEREDHOUR hour-12 12); ETC.

tretti    NUM11 * @(NUMFEATS 30).

----
```

Tillegg D

Templatene i XLE

Her følger et utvalg av de templatene som benyttes i XLE-koden. Disse templatene blir kalt i frasestrukturregelene eller fra leksikon. Alle templatene som er laget i forbindelse med min minigrammatikk, er gjengitt her. I tillegg inkluderer jeg også en del av NorGram-templatene som min grammatikk anvender.

```
BOKMAL NORSK TEMPLATES (1.0)

[...]

"Existing NorGram-templates that are used in this grammar."

"Adjunct PPs with OBJ-taking Ps:"

ADJUNCTPP =
          ! $ (^ ADJUNCT)
          ~(^ CASE)=gen
          (! PTYPE)=c sem
          ~(! PSEM)=part
          { (s::M* ind)
            (s::* arg1)=(s::M* ind)
          | ~(s::M* ind)
            (s::* arg1)=sit }
            s::* $ (s::M* loc)
          @ANTECEDENTFEATS
          Mark5 $ o::*
          "Efficiency: Selected-PP readings whenever possible"
          "</mrs>"@MRS-ADJUNCTPP "<mrs/>".

COUNTNOUN (P S) =
```

```
          "{" (^ PRED)='P'
          (^ NTYPE NSEM COMMON)=count
          (^ NTYPE NSYN)=common
          (s::* relation)=S
          (s::* arg1 ind)=(s::M* ind)
          s::* $ (s::M* cond)
          "</mrs>" @(MRS-COUNTNOUN P) "<mrs/>"
          "| @(MASSNOUNGRIND P S)}".
          "Remove grinder in clock version of templates."

DEFIFPREDEF =
          { (^ DEF)=c +
            (^ CHECK _PREDEF)=c +
            s::* $ (s::M* cond)
            (s::* arg1)=(s::M* ind)
          | (^ DEF)=-
            { (^ CHECK _PREQ)=c +
            | ~(^ CHECK _PREQ)=+
              s::M*=s::^ "See comment in DP rule"
              @(COMPLETENOMFEATS) }
            s::M*=s::* }.

DEGADV (P S) =
          { (^ PRED)='P'
          | (^ PRED)='P<(^ OBL-COMPAR)>' }
          (s::M* rel-mod)=S.

LITNUM =
          (^ REF)=+
          { (! AGRNUM)=c(! HEADNUM)
            (^ NUM)=(! AGRNUM)
          | ~(! AGRNUM)=(! HEADNUM)
             { (^ NUM)=(! AGRNUM)
              ~(^ NUM)=(! HEADNUM)
             | (^ NUM)=(! HEADNUM) }}
          "</mrs>"@MRS-LITNUM"<mrs/>".

MEASURABLEPREP =
          (^ CHECK _MSRBL)=+.

MRS-ADJUNCTPP =
          (m::! RELS_EL LBL)=(m::^ RELS_EL LBL)
          (m::! RELS_EL ARG1)=(m::^ RELS_EL ARG0)
          "@(IF ~[(m::^ RELS_EL ARG0 type)=ref-ind]
```

```
        (m::! RELS_EL ARG0)=(m::^ RELS_EL ARG0))"
        (m::! RELS_EL ARG0 TENSE)=notense
        (m::! RELS_EL ARG0 PERF)=- "Sjekk!"
        @(RELSAPPEND1 (m::! RELS))
        @(H-CONSAPPEND1 (m::! H-CONS)).

MRS-COUNTNOUN (S) =
        (m::^ RELS_EL _CAT)=n
        { (^ NUM)=c sg
          (m::^ RELS_EL ARG0 DIV)=-
        | ~(^ NUM)=sg
        (m::^ RELS_EL ARG0 DIV)=+ }
        (m::^ RELS_EL ARG0 NUM)=(^ NUM)
        @(IF ~(^ CHECK _PREDICATIVE)=+
        (m::^ RELS_EL ARG0 PERS)=(^ PERS))
        @(MRS-COUNTNOUNCOMMON S).

MRS-COVQUANT (S pers num) =
        (%QUANT RELS_EL ARG0 PERS)=pers
        (%QUANT RELS_EL ARG0 NUM)=num
        %QUANT=(m::^ _QUANT)
        (%QUANT RELS_EL LNK)=(^ PRED SFID)
        @(MRS-COVQUANTCOMMON S).

 MRS-DEGADVSYNT =
        { (! ADJUNCT-TYPE)=c degree
          (! DEGREE)=c comparative
          (^ DEGREE)=(! DEGREE)
          (^ DEG-DIM)=(! DEG-DIM)
          (^ CHECK _DEG-MORPH)=c positive
          (%COMPAR RELS_EL relation)='comp'
          (%COMPAR RELS_EL _CAT)=(^ DEG-DIM)
          (%COMPAR RELS_EL LBL type)=handle
          (%COMPAR RELS_EL LBL)=(m::^ RELS_EL LBL)
          (%COMPAR RELS_EL ARG0 type)=event
          (%COMPAR RELS_EL ARG0)=(m::^ RELS_EL ARG0)
          (%COMPAR RELS_EL ARG1)=(m::^ RELS_EL ARG1)
          @(RELSLIST1 (%COMPAR RELS_EL))
          @(IF (! OBL-COMPAR)
          [  (%COMPAR RELS_EL ARG2)=(m::(! OBL-COMPAR OBJ) RELS_EL ARG0)
             (%COMPAR RELS_EL LBL)=(m::(! OBL-COMPAR) RELS_EL LBL)
             @(RELSAPPEND1 (m::(! OBL-COMPAR) RELS))
             @(H-CONSAPPEND1 (m::(! OBL-COMPAR) H-CONS))])
          | (! ADJUNCT-TYPE)=c degree
```

```
              (! DEGREE)=c superlative
              (^ DEGREE)=(! DEGREE)
              (^ DEG-DIM)=(! DEG-DIM)
              (^ CHECK _DEG-MORPH)=c positive
              (%COMPAR RELS_EL relation)='superl'
              (%COMPAR RELS_EL _CAT)=(^ DEG-DIM)
              (%COMPAR RELS_EL LBL type)=handle
              (%COMPAR RELS_EL LBL)=(m::^ RELS_EL LBL)
              (%COMPAR RELS_EL ARG0 type)=event
              (%COMPAR RELS_EL ARG0)=(m::^ RELS_EL ARG0)
              (%COMPAR RELS_EL ARG1)=(m::^ RELS_EL ARG1)
              @(RELSLIST1 (%COMPAR RELS_EL))
            | (! ADJUNCT-TYPE)=c degree
              (! DEGREE)=c positive
              (^ DEGREE)=(! DEGREE)
              (^ DEG-DIM)=(! DEG-DIM)
              (^ CHECK _DEG-MORPH)=c positive
              (%COMPAR RELS_EL relation)='comp'
              (%COMPAR RELS_EL _CAT)=(^ DEG-DIM)
              (%COMPAR RELS_EL LBL type)=handle
              (%COMPAR RELS_EL LBL)=(m::^ RELS_EL LBL)
              (%COMPAR RELS_EL ARG0 type)=event
              (%COMPAR RELS_EL ARG0)=(m::^ RELS_EL ARG0)
              (%COMPAR RELS_EL ARG1)=(m::^ RELS_EL ARG1)
              @(RELSLIST1 (%COMPAR RELS_EL))
              @(IF (! OBL-COMPAR)
              [  (%COMPAR RELS_EL ARG2)=(m::(! OBL-COMPAR OBJ) RELS_EL ARG0)
                 (%COMPAR RELS_EL LBL)=(m::(! OBL-COMPAR) RELS_EL LBL)
                @(RELSAPPEND1 (m::(! OBL-COMPAR) RELS))
                @(H-CONSAPPEND1 (m::(! OBL-COMPAR) H-CONS))])
            | ~(! ADJUNCT-TYPE)=degree
              (%COMPAR RELS_EL relation)=(! PRED FN)
              (%COMPAR RELS_EL _CAT)=x_deg
              (%COMPAR RELS_EL LBL)=(m::^ RELS_EL LBL)
              (%COMPAR RELS_EL ARG0 type)=event
              "(%COMPAR RELS_EL ARG0)=(m::^ RELS_EL ARG0)"
              "(%COMPAR RELS_EL DARG)=(m::^ RELS_EL relation)"
              (%COMPAR RELS_EL ARG1)=(m::^ RELS_EL ARG0) "More ERG-like??"
              @(RELSLIST1 (%COMPAR RELS_EL))}.

MRS-LITNUM =
          (m::! _CARD ARG1)=(%QUANT RELS_EL ARG0)
          (m::! _TOPHNDL)=(%OBJECT RELS_EL LBL)
          "@(RELSLIST1 (m::! RELS_EL))"
```

```
          { ((NUMBER !)QUANT)
           ~((SPEC NUMBER !)CHECK _PREDEF)
           |((SPEC NUMBER !)CHECK _PREDEF)
           |~((NUMBER !)QUANT)
            ~((SPEC NUMBER !)CHECK _PREDEF)
            { ~(^ NUM)=sg
              @(MRS-BAREDIVQUANTCOMMONREST)
            | (^ NUM)=c sg
              @(MRS-COVQUANTCOMMON udef_q) }}
            { ~(^ OBL)
            m::^ = %OBJECT
          | (^ OBL)
            m::(^ OBL) = %OBJECT }.


MRS-NUMFEATS (dig) =
          (m::^ _CARD relation)='card'
          (m::^ _CARD LBL type)=handle
          (m::^ _CARD ARG0 type)=event
          (m::^ _CARD ARG1 type)=ref-ind
          (m::^ _CARD CARG)=dig
          @(RELSLIST1 (m::^ _CARD)).

MRS-OBJECT (rel pers num) =
          (%OBJECT TOP type)=handle
          (%OBJECT RELS_EL relation)='rel'
          (%OBJECT RELS_EL LBL type)=handle
          (%OBJECT RELS_EL ARG0 type)=ref-ind
          (%OBJECT RELS_EL ARG0 PERS)=pers
          (%OBJECT RELS_EL ARG0 NUM)=num
          (%OBJECT RELS_EL LNK)=(^ PRED SFID)
          @(RELSLIST1 (%OBJECT RELS_EL)).

MRS-OBJECTGVN (rel) =
          m::^ = %OBJECT
          @(MRS-OBJECT rel 3 (^ NUM)).


MRS-PRPSTN =
          (m::^ INDEX type)=event
          { (m::^ INDEX TENSE)=(^ TNS-ASP TENSE)
          | (m::^ INDEX TENSE)=c fut }
          (m::^ INDEX)=(m::^ RELS_EL ARG0)
          (m::^ _MSG relation)='prpstn_m_'
```

```
          (m::^ _MSG LBL type)=handle
          (m::^ _MSG MARG type)=handle
          (m::^ _MSG ARG0)=(m::^ INDEX)
          (m::^ TOP)=(m::^ _MSG LBL)
          @(RELSLIST1 (m::^ _MSG))
          (%local2 relation)=qeq
          (%local2 SC_ARG)=(m::^ _MSG MARG)
          (%local2 OUTSCPD)=(m::^ RELS_EL LBL)
           @(H-CONSLIST1 %local2).

MRS-QUANT (S) =
          m::^ = %OBJECT
          (%QUANT RELS_EL ARG0 PERS)=(^ PERS)
          (%QUANT RELS_EL ARG0 NUM)=(^ NUM)
          %QUANT=(m::^ _QUANT)
          (%QUANT RELS_EL LNK)=(%QUANT RELS_EL relation SFID)
          @(MRS-COVQUANTCOMMON S).

MRS-SEMPREP (P) =
         (m::^ RELS_EL _CAT)=p
          @(MRS-COVPREP P). "See explanation under MRS-COVQUANT."

NOEN =
         (^ REF)=+
         (s::M* relation)=some
         (^ PERS)=3.

"The PRED-introducing reading of prepositions:"

SEMPREP (P S) =
      (^ PRED)='P<(^ OBJ)>'
      (^ PTYPE)=sem
      (^ PFORM)=P
      (s::M* relation)=S
      "</mrs>"%ARG2=m::(^ OBJ)
      @(MRS-SEMPREP P) "<mrs/>".

"Templates made for this thesis:"

CLOCKCOPULA =
          ~(^ PREDLINK NTYPE)
          ~(^ PREDLINK ATYPE)
          ~(^ PREDLINK VTYPE)
          ~(^ PREDLINK PRON-TYPE)
```

```
        ~(^ PREDLINK ADV-TYPE)
         (^ PREDLINK CLOCK-EXPRESSION)=c +
        "(^ SUBJ PRED FN)=c klokke"
        %ARG1=m::(^ SUBJ)
        @(MRS-1ARGCOMMON clock-id)
        (m::^ RELS_EL ARG2 type)=ref-ind
        (m::^ RELS_EL ARG2)=(m::(^ PREDLINK) RELS_EL ARG1)
        @(RELSAPPEND1 (m::(^ PREDLINK) RELS))
        @(H-CONSAPPEND1 (m::(^ PREDLINK) H-CONS)).

CLOCK-MODIFIER =
        (^ NTYPE CLOCK-MODIFIER)=+.

MRS-BEFORE =
        (m::^ RELS_EL relation) ='before'
        (m::^ RELS_EL LBL type)=handle
        (m::^ RELS_EL ARG0 type)=event
        (m::^ RELS_EL ARG1 type)=ref-ind
        @(MRS-TIMEPOINT (m::^ RELS_EL ARG1))
        (m::^ RELS_EL ARG2)=(m::(^OBJ) RELS_EL ARG1)
        @(RELSLIST1 (m::^ RELS_EL))
        @(RELSAPPEND1 (m::(^ OBJ) RELS))
        @(H-CONSAPPEND1 (m::(^ OBJ) H-CONS)).

MRS-CLOCKADJUNCT =
        (%PREP RELS_EL relation)='unspec_loc'
        (%PREP RELS_EL LBL type)=handle
        (%PREP RELS_EL LBL)=(m::^ RELS_EL LBL)
        (%PREP RELS_EL ARG0 type)=event
        (%PREP RELS_EL ARG1)=(m::^ RELS_EL ARG0)
        (%PREP RELS_EL ARG2)=(m::! RELS_EL ARG1)
        (%PREP RELS_EL LNK)=(m::! PRED SFID)
        @(RELSPUSH (m::! RELS)(%PREP RELS_EL))
        @(H-CONSAPPEND1 (m::! H-CONS)).

MRS-EMPTYHOUR =
        (%HOUR RELS_EL relation)='unknown_hour'
        (%HOUR RELS_EL LBL type)=handle
        (%HOUR RELS_EL ARG0 type)=ref-ind
        @(RELSLIST1 (%HOUR RELS_EL))
        (%PREP RELS_EL relation) ='temp-loc'
        (%PREP RELS_EL LBL type)=handle
        (%PREP RELS_EL ARG0 type)=event
        (%PREP RELS_EL ARG1 type)=ref-ind
```

```
          @(MRS-TIMEPOINT (%PREP RELS_EL ARG1))
          (%PREP RELS_EL ARG2)=(%HOUR RELS_EL ARG0)
          @(RELSLIST1 (%PREP RELS_EL))
          %HOUR=%OBJECT
          @(MRS-COVQUANT udef_q 3 sg).


MRS-HALV =
          @MRS-BEFORE
          (m::^ RELS_EL ARG2)=(%ARG2 RELS_EL ARG1)
          (m::^ _CARD relation)='card'
          "(%NUM TOP type)=handle" "Necessary??"
          (m::^ _CARD LBL type)=handle
          (m::^ _CARD ARG0 type)=event
          (m::^ _CARD ARG1 type)=ref-ind
          (m::^ _CARD CARG)=30
          (m::^ _CARD LNK)=(^ SPEC NUMBER PRED SFID)
          @(RELSLIST1 (m::^ _CARD))
          (m::^ _CARD ARG1)=(%QUANT RELS_EL ARG0)
          (m::^ _CARD LBL)=(%OBJECT RELS_EL LBL)
          { ~(^ NUM)=sg
            @(MRS-BAREDIVQUANT)
           | (^ NUM)=c sg
            @(MRS-QUANT udef_q) }
          (%OBJECT RELS_EL relation)='minutt'
          (%OBJECT RELS_EL LBL type)=handle
          (%OBJECT RELS_EL ARG0 type)=ref-ind
          @(RELSLIST1 (%OBJECT RELS_EL))
          (%MEAS RELS_EL relation)='measure'
          (%MEAS RELS_EL LBL type)=handle
          (%MEAS RELS_EL ARG0 type)=event
          (%MEAS RELS_EL ARG1)=(m::^ RELS_EL ARG0)
          (%MEAS RELS_EL ARG2)=(%OBJECT RELS_EL ARG0)
          @(RELSLIST1 (%MEAS RELS_EL)).

MRS-HOUR (dig) =
        (%HOUR RELS_EL relation)='numbered_hour'
        (%HOUR RELS_EL LBL type)=handle
        (%HOUR RELS_EL ARG0 type)=ref-ind
        (%HOUR RELS_EL CARG)=dig
        @(RELSLIST1 (%HOUR RELS_EL))
        (m::^ RELS_EL relation) ='temp-loc'
        (m::^ RELS_EL LBL type)=handle
        (m::^ RELS_EL ARG0 type)=event
```

```
        (m::^ RELS_EL ARG1 type)=ref-ind
        (m::^ RELS_EL ARG2)=(%HOUR RELS_EL ARG0)
        @(RELSLIST1 (m::^ RELS_EL))
        %HOUR=%OBJECT
        @(MRS-COVQUANT udef_q 3 sg).

MRS-KVART =
        @MRS-MINUTE
        (m::^ _CARD relation)='card'
        (m::^ _CARD LBL type)=handle
        (m::^ _CARD ARG0 type)=event
        (m::^ _CARD ARG1 type)=ref-ind
        (m::^ _CARD CARG)=15
        (m::^ _CARD LNK)=(^ SPEC NUMBER PRED SFID)
        @(RELSLIST1 (m::^ _CARD))
        (m::^ _CARD ARG1)=(%QUANT RELS_EL ARG0)
        (m::^ _CARD LBL)=(%OBJECT RELS_EL LBL)
         m::^=%OBJECT
        @(MRS-BAREDIVQUANT).

MRS-MINUTE =
        (m::^ RELS_EL relation)='minute'
        (m::^ RELS_EL LBL type)=handle
        (m::^ RELS_EL ARG0 type)=ref-ind
        @(RELSLIST1 (m::^ RELS_EL)).

MRS-MINUTEMEASURE =
          (%MEAS RELS_EL relation)='measure'
          (%MEAS RELS_EL LBL type)=handle
          (%MEAS RELS_EL ARG0 type)=event
          (%MEAS RELS_EL ARG1)=(m::^ RELS_EL ARG0)
          (%MEAS RELS_EL ARG2)=(m::! RELS_EL ARG0)
          @(RELSLIST1 (%MEAS RELS_EL))
          @(RELSAPPEND1 (m::! RELS))
          @(H-CONSAPPEND1 (m::! H-CONS)).

 MRS-PAST =
          (m::^ RELS_EL relation) ='past'
          (m::^ RELS_EL LBL type)=handle
          (m::^ RELS_EL ARG0 type)=event
          (m::^ RELS_EL ARG1 type)=ref-ind
          @(MRS-TIMEPOINT (m::^ RELS_EL ARG1))
          (m::^ RELS_EL ARG2)=(m::(^ OBJ) RELS_EL ARG1)
          @(RELSLIST1 (m::^ RELS_EL))
```

```
          @(RELSAPPEND1 (m::(^ OBJ) RELS))
          @(H-CONSAPPEND1 (m::(^ OBJ) H-CONS)).

MRS-TIMEPOINT (var) =
          (%TIMEPT RELS_EL relation)='timepoint'
          (%TIMEPT RELS_EL LBL type)=handle
          (%TIMEPT RELS_EL ARG0)=var
          @(RELSLIST1 (%TIMEPT RELS_EL))
          %TIMEPT=%OBJECT2
          @(MRS-COVQUANT2 udef_q 3 sg).

NUMBEREDHOUR (P dig) =
          (^ PRED)='P'
          @(MRS-HOUR dig).

----
```