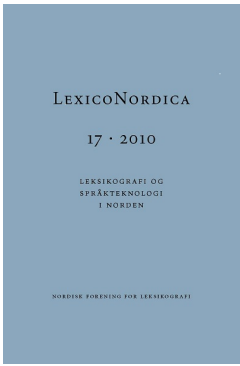


LexicoNordica

Titel:	Felles leksikalske ressursar for språkteknologi og leksikografi	
Forfatter:	Trond Trosterud	
Kilde:	LexicoNordica 17, 2010, s.211-223	
URL:	http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive	

© LexicoNordica og forfatterne

Betingelser for bruk af denne artikel

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

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

Søgbarhed

Artiklerne i de ældre LexicoNordica (1-16) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Felles leksikalske ressursar for språkteknologi og leksikografi

Trond Trosterud

Lexicography and language technology have traditionally been seen as two distinct disciplines. In addition to pointing at the common need to manage huge lemma lists in one way or another, the present paper argues for conducting lexicographical and language technology projects in an integrated way, with the same lemma list for both dictionary and grammatical analyser/generator. This is especially important for small language communities, which lack the resources to uphold several parallel lexical databases.

1. Innleiing

La oss byrje med å sjå verda frå språkteknologen og leksikografen sin ståstad. Språkteknologen kan analysere laupande tekst, og vil gje analysen kunnskap om orda. Leksikografen vil vite kva eigenskapar lemmaforrådet har, og presentere dette for ordboksbrukaren, for eitt og eitt lemma om gangen. Som regel er det lite kontakt desse miljøa imellom. For små språksamfunn er det problematisk. Større språksamfunn har ressursar til å vedlikehalde store grammatiske og leksikalske ressursar i fleire miljø, små språksamfunn kan ikkje det.

Ein integrasjon mellom leksikografi og språkteknologi kan t.d. innebere at leksikonet i den morfologiske analysatoren blir erstatta med ordboka si lemmaliste, slik at den informasjonen som analysatoren treng, også blir lagt til basen som er grunnlaget for ordboka. Den morfologiske analysatoren treng lemma, stamme og informasjon om kontinuasjonleksikon (bøyingsklasse). Den syn-

taktiske analysatoren treng informasjon om syntaktiske funksjonar og valens, og informasjon om semantiske roller.

Denne artikkelen vil argumentere for å integrere delar av det leksikografiske og språkteknologiske arbeidet, der same lemmaliste inneheld både leksikografisk og språkteknologisk informasjon. I tillegg til å spare arbeid til vedlikehald av store lemmalister, vil det også gje meir avanserte ordbøker og analyseprogram.

2. Leksikografi og språkteknologi som separate prosjekt

Det er lett å innsjå det problematiske ved å vedlikehalde fleire parallelle lemmalister, særleg for språksamfunn med avgrensa ressursar. Korfor er den vanlegaste arbeidsmåten likevel at leksikografane og språkteknologane arbeider kvar for seg?

Det viktigaste argumentet for å operere med leksikalske basar som berre tener eitt formål kvar, er at sjølve basen, eller kjeldekoden, blir enklare. Viss kvart lemma får knytt til seg berre ein handfull typar av data, kan basen bli lagra som ei kommaseparert liste (t.d. i form av eit rekneark) med berre dei opplysningane som trengst til eitt spesifikt prosjekt. Ein slik framgangsmåte kan vere kostnadseffektivt for akkurat det relevante formålet. Det er også lettare å finne hyllevareprogramvare til eitt veldefinert formål. Ein ordboksbase til ei ordbok med eitt formål er lettare å tilpasse til eit ordboksredigeringsprogram enn ein base som skal brukast til mange formål.

Argument mot å la same leksikalske base ha mange formål, er at det blir stort og komplisert. Prosjekta risikerer å velte på lange diskusjonar om leksikonstruktur; innafor språkteknologien er dette er ei velprøvd tue som har velta mange lass. Slike fleirbruksbasar stiller også større krav til datakompetansen til brukarane. Det er med andre ord problematisk å lage fleirbruksbasar, og det er enklare på kort sikt å halde basar for ulike føremål kvar for seg.

I den leksikografiske kvardagen ser vi at akademiske ordboksprosjekt ofte er store prosjekt med lange tradisjonar, og ikkje knytte til språkteknologi. Kommersielle ordboksprosjekt på si side opererer under strenge økonomiske vilkår. Einforfattarprosjekt vil ofte heller ikkje ha kunnskap til eit vidare prosjekt. Alt dette talar for å oppretthalde status quo, og la ulike prosjekt arbeide på kvar sine leksika, heller enn å la same leksikon tene både leksikografiske og språkteknologiske føremål.

I den språkteknologiske kvardagen er akademiske språkteknologiske prosjekt sjeldan innretta på å analysere laupande tekst på ein måte som er robust nok til faktisk å gje gode resultat for autentiske tekstar. Prosjekt for semantisk annotering er som regel ikkje integrert med andre språkteknologiske eller leksikografiske komponentar. Kommersielle språkteknologiske prosjekt opererer, på same måten som kommersielle ordboksprosjekt, under strenge økonomiske vilkår. Maskinomsetjing var tidlegare eit viktig felt for datamaskinell leksikografi, ved utarbeiding av avanserte transferleksika, dvs. ordbøker til bruk for maskinomsetjingsprogramma. Frå og med midten av 1990-talet kom det eit paradigmeskifte innafør maskinomsetjingsfeltet, fokus vart flytt frå grammatisk til statistisk basert maskinomsetjing, og maskinomsetjingsfeltet har seinare vore lite interessert i transferleksika med semantisk informasjon (om valens o.l.). Alt dette gjev dårlege vilkår for kontakt mellom leksikografi og språkteknologi. Også innafør språkteknologien er det altså fleire faktorar som dreg i retning av status quo.

3. Integrasjon mellom leksikografi og språkteknologi

Ei felles lemmaliste og ein felles leksikalsk database for både ordbøker og språkteknologiske applikasjonar er lettare å vedlikehalde enn å ha parallelle leksikalske ressursar. Men det er også andre fordelar ved å ha ein felles database.

Der lemmaforrådet er direkte knytt til ein morfologisk analysator, har ordboksbrukaren tilgang ikkje berre til oppslagsforma, men til alle ordformene i bøyingsparadigmet. Dette er spesielt viktig for elektroniske resepsjonsordbøker. Ein morfologisk analysator knytt til ordboka gjer dette mogleg. Dei kommersielt interessante språka er for morfologifattige til å tvinge fram integrerte grammatiske analysar i dei elektroniske ordbøkene. Det vesle som finst av morfologi i t.d. engelsk og tysk kan leggjast til manuelt eller ved enkle prosessar, og løysingar for morfologirike språk blir ikkje utvikla.

Ved å integrere leksikon med morfologisk analyse vil den grammatiske analysatoren ha tilgang til informasjon frå ordboka. Viktig her er valens og semantikk. For å få ein vellukka grammatisk analyse er det viktig å ha t.d. valensinformasjon for verb. Ein NP i oblik kasus kan bli analysert som oblikt objekt eller laust adverbial, alt etter eigenskapane til verbet. Omvendt vil også aktiv bruk av den grammatiske informasjonen i ordboka setje denne informasjonen på prøve: Viss t.d. valensramma seier at eit visst verb ikkje skal ta akkusativobjekt, men vi ved å kombinere leksikon med ein analysator og ein tekst finn setningar der verbet faktisk tar akkusativobjekt likevel, blir vi nøydde til å revidere ordboka på det punktet.

4. Korleis integrere

Eit eksempel på integrerte løysingar er arbeidet med leksikografi og språkteknologi for komi som blir gjort i Helsingfors og Tromsø. Komi er eit finskugrisk språk i Nordvest-Russland. Arbeidet har ikkje kome langt, men integreringa av leksikografi og språkteknologi er på plass. Bakgrunnen for arbeidet er Jack Rueters *Komi-English-Finnish Dictionary* (KEFD) og analysatoren for komi (KOMFST). Poenget med eksempelet er altså verken ordboka eller

analyseprogrammet, men måten dei er integrert med kvarandre på.

Figur 1 viser ein tilfeldig ordboksartikkel for ordboka komi-engelsk/finsk, verbet аддзёдчыны 'addzödtsjyny' ”å møte”. Relevant her er, i tillegg til lemma, også stamme (<stem>), ordklasse (<pos>), og kontinuasjonsleksikon (dvs. ein peikar til kor i automaten vi skal gå for å lese neste tilstand (<contlex>)).

```

<entry>
  <lemma>аддзёдчыны</lemma>
  <stem>аддзёдчы</stem>
  <contlex>Verb3</contlex>
  <pos>V</pos>
  <article>
    <sem>RECIP</sem>
    <syn>SVP/-кбд</syn>
    <eng>
      <choice>
        <variant>meet</variant>
      </choice>
    </eng>
    <fin>
      <choice>
        <variant>tavata</variant>
      </choice>
    </fin>
  </article>
</entry>

```

Figur 1: Komi-leksikon i XML-format

Ordboka er integrert med kjeldekoden til ein endeleg tilstandsautomat.¹ Kvar ordklasse genererer ei leksikonfil i automaten, som

1 Ein endeleg tilstandsautomat (*finite-state automaton*) er ein modell som inneheld eit endeleg sett av *tilstandar* og overgangar frå tilstand til tilstand, der *tilstand* er definert som eit unikt sett av informasjon. Ein automat for ordet *båt* vil bestå av fire tilstandar: ein starttilstand, ein tilstand for *b*, ein for *bå*, og ein slutttilstand for *båt*. Ein spesiell type tilstandsautomatar er transdusarar, som ikkje berre les symbol, men også endrar dei, slik at strengen *båtane* kan bli endra til strengen *båt +N +Sg +Def* og omvendt. For ei innføring i bruk av endelege tilstandsautomatar i språkteknologi, sjå Beesley og Karttunen (2003).

par av lemma (til venstre for kolon) og stamme (til høyre for kolon), som vist i Figur 2. Frå denne representasjonen peiker kontinuasjonsleksikonet (her: Verb3) vidare i automaten. For å gjere leksikonet lettare å lese er det også teke med ei omsetjing, men berre som kommentarfelt til internt bruk.

```

абутомасьны:абутомась Verb2 "feign poverty" ;
адавны:адал Verb1 "gobble up" ;
аддзавны:аддзал Verb1 "find" ;
аддзбдчыны:аддзбдчы Verb3 "meet" ;
аддзбдлыны:аддзбдлы Verb3 "see" ;
аддзыны:аддзы Verb3 "see" ;
аддзывны:аддзыл Verb3 "experience" ;
аддзысьны:аддзысь Verb2 "be found" ;

```

Figur 2: Leksikonfil i automaten for komi

I verbmorfologifila i Figur 3 får verbet infinitivsending, og det blir sendt vidare til leksikonet Finiteforms, der det får tempus, før det blir sendt vidare til leksikona for personendingar. Resultatet er ein morfologisk generator/analysator for komi.

```

LEXICON Verb3 ! ярмыны. Ending in -ыны.
+V: VerbConj ;
+V+Inf:Жыны K ;

LEXICON VerbConj
Finiteforms ;
Non-finiteforms ;

LEXICON Finiteforms ! Gives linking vowels for 3 tenses
+Ind+Fut:Жа PresPret1 ; ! 1,2 Future
+Ind+Prs:Жа PresPret1 ; ! 1,2 Present
+Ind+Prt1:Жа PresPret1 ; ! 1,2 Preterite1. i и variation

LEXICON PresPret1 ! First and second person
+Sg1:Жа K ;
+Sg2:Жн K ;
+Pl1:Жм K ;
+Pl1:Жмб K ;
+Pl1:Жмбй K ;
+Pl2:Жннд K ;
+Pl2:Жд K ;

```

Figur 3: Morfologisk generator/analysator for Komi (utdrag)

Med utgangspunkt i denne automaten er det mogleg å analysere tekst på komi. Figur 4 viser ein analyse av presentasjonsteksten på framsida til komiutgåva av Wikipedia. Komi-analysatoren er uferdig, og nye lånord og namn er ikkje med.

"<Тайо>"	"тайо" Pron Dem	"<Кони>"	"кони" N Sg Nom
"<субдоменсö>"	"субдоменсö" ?	"<кыв>"	"кыв" N Sg Acc
"<видзэны>"	"видзэны" V Ind Prs Pl3	"<кыв>"	"кыв" N Sg Nom
"<,>"	"," CLB	"<кыв>"	"кыв" N Sg Acc
"<недэйт>"	"недэйт" CC	"<кыв>"	"кыв" N Sg Nom
"<вочны>"	"вочны" V Inf	"<кыв>"	"кыв" N Sg Acc
"<Википедия>"	"Википедия" ?	"<вэлын>"	"вэлын" Adv
"<гытэжэнь>"	"гытэжэнь" ?	"<вэв>"	"вэв" N Sg Ine
		"<гыжюдъяс>"	"гыжюд" N Pl Nom
		"<.>"	"гыжюд" N Pl Acc
			"," CLB

Figur 4: Analysert tekst frå komi Wikipedia, med manglande dekning for lånord som субдоменсö 'subdomene' og Википедия 'Wikipedija'

I tillegg til å danne kjernen i ein morfologisk analysator, kan den leksikalske databasen sjølvstøtt også brukast til å lage ordbøker, t.d. e-ordbøker, som i Figur 5. Her er på ny ordet for å møte, *аддзöдчыны*, og same lemmaartikkel er no presentert i eit ordboksgrensensnitt. Ordforma er ei infinitivform, jf. endinga *-ыны*, men frå denne ordboka sitt synspunkt er ordforma eit uanalysert heile.



Figur 5: Databasen til automaten fungerer også som input for ei komi e-ordbok

Så lenge den morfologiske analysatoren er kompatibel med informasjonen om bøyingsklasse og stamme som blir oppgjeven i ordboka, er det mogleg å bruke same morfologiske analysator til fleire ordbøker. Eit døme på det er færøysk.

Fróðskaparsetur Føroya har gjeve ut ei svært god einspråkleg ordbok, *Føroysk orðabók* (FO). Dei har deretter gjort lemmaliste og bøyingskode frå ordboka tilgjengeleg for språkteknologiske føremål. Denne lista utgjer leksikon for analysatoren for færøysk utarbeidd ved Universitetet i Tromsø (FAOFST, Trosterud 2009). Analysatoren FAOFST bruker nøyaktig same bøyingskoder som FO, og kan med hjelp av denne informasjonen generere alle paradigma for alle oppslagsorda i ordboka. Dette inneber at det også er mogleg å kombinere nye lemmalister, t.d. nye, reviderte utgåver av ordboka, eller den færøyske delen av eventuelle tospråklege ordbøker, med analysatoren, så lenge desse andre ordbøkene også bruker FO sitt bøyingskodesystem. Eit døme på ein slik situasjon har vi for finsk, der *Suomen kielen perussanakirja* (SKP) sitt bøyingsklassesystem har vorte ein de facto standard for fleirspråklege ordbøker, t.d. til marisk og norsk (jf. Trosterud 2003:13ff for ei drøfting). Det same bøyingsklassesystemet ligg til grunn for ein ope tilgjengeleg morfologisk analysator for finsk (OMORFI, Pirinen 2008).

5. Bruk av språkteknologi i leksikografien

Ein føresetnad for at det skal vere mogleg å gjere bruk av automatiske analysemodellar i leksikografien, er at dei leksikografiske databasane er maskinleselege. Dei ulike delane av lemmaartikkelen må formaterast semantisk, og ikkje visuelt, dvs. etter kva dei ulike felte inneheld og ikkje etter korleis dei ser ut. Innhaldet i felte for grammatisk informasjon må også vere utvitydig maskinleseleg. T.d. er det ikkje nok å gje eit fullt bøyingsmønster (med

fleirtalsformer inkludert) for eit abstrakt substantiv som berre optrer i eintal. For menneskelege lesarar er det mogleg å stole på at dei forstår at substantivet det gjeld, ikkje blir brukt i fleirtal, og dermed heller ikkje har fleirtalsformer, men slike subtile kombinasjonar av direkte formalisert og indirekte implisitt informasjon er for vanskelege for maskinell prosessering. Skal det vere mogleg å representere den informasjonen som potensielt ligg i ordboka, må han bli gjort eksplisitt. Det må, for kvart av leksema utan fleirtalsformer, faktisk bli brukt ein annan kode enn for elles identiske leksema som altså har fleirtalsformer.

Ein måte å bruke språkteknologi i leksikografien er ved utarbeiding av elektroniske resepsjonsordbøker. Som vist i Antonsen m.fl. 2009, vil ei resepsjonsordbok basert berre på oppslagsformene i ei mellomstor ordbok (testen vart utført med lemmaforrådet i *Bokmålsordboka* (BO), SKP og *Sámi-suoma sátnegirji* (SSS)) kjenne att berre 30,5 % av orda i laupande tekst for norsk, 10 % for finsk, og 7,9 % for samisk (jf. Antonsen m.fl. 2009). Ein stor del av dei ordformene som ordbøkene ikkje kjenner att, vil ha regelrett morfologi og dermed vere lett atkjennelege for lesarar, men så lenge ordboka ikkje kjenner dei att, vil det vere umogleg å klikke på desse og få omsetjingar direkte i lesing av tekst på skjerm.

Neste steg for den elektroniske resepsjonsordboka er disambiguering og leksikalsk seleksjon. Med tilgang til heile setninga vil det for ordboka vere mogleg å finne rett ordklasse og rett homonym, og å prioritere mellom ulike tydingar. Det syntaktiske rammeverket som har vist seg robust nok til å analysere laupande tekst, er *føringsgrammatikk* (*Constraint Grammar*, jf. t.d. Karlsson et al 1995). Føringsgrammatikalske analyseprogram tar setningar (eller større einingar) som er morfologisk analyserte som i Figur 4, som input, vel rett morfologisk analyse og legg til syntaktiske funksjonar. I ei setning som *Vi skal finne mat* vil eit sett føringsgrammatiske reglar velje infinitivslesinga for *finne* pga. modalverbet til venstre, og forkaste imperativslesinga av *mat* pga. det transi-

tive verbet til venstre. Korrekt grammatisk analyse vil dermed (for ei norsk–engelsk e-ordbok) unngå omsetjingsframlegg som *Finn* og *feed*, til fordel for *find* og *food*. Same type reglar kan også skilje mellom ulike tydingar av same ord, som i setningane *Toget er i rute* og *Toget hadde overraskande mange deltakarar*, der substantiva til slutt i setningane peiker på omsetjingar som *train* og *march*.

Det å setje saman leksikografi og språkteknologi opnar også nye perspektiv for vurdering av lemmaforrådet generelt, og dermed også for ordbokskritikken. Eit godt døme på det er arbeidet til Hurskainen (2004), ein ordbokskritikk som kviler på eit noko uvanleg forarbeid. For å kunne evaluere fem ulike ordbøker med swahili som L1 har Hurskainen fått tilgjenge til elektroniske versjonar av alle fem, og deretter kopla dei alle til ein morfologisk analysator over swahili. Deretter har han analysert eit balansert tekstkorpus på omtrent 4 millionar ord. På den måten har han vore i stand til å vurdere dei ulike ordbøkene etter to ulike kriterium: I kor stor grad dei dekkjer ordforrådet i korpuset, og i kor stor grad dei inneheld ord som ikkje finst i korpuset.

Det første kriteriet tilsvare ein vanleg framgangsmåte ved ordbokskritikk: Å undersøke i kor stor grad ordforrådet i eit visst tekstkorpus er dekt av ordboka. Skilnaden er at der vi normalt er nøydd til å ta stikkprøver, kan Hurskainen faktisk analysere heile tekstsamlingar. Med det andre kriteriet undersøker Hurskainen i kor stor grad ordboka inneheld såkalla ordboksord, ord som berre finst i ordbøker, men aldri i aktuell språkbruk. Viss vi finn ord som verkeleg aldri opptrer utanfor ordboka, må dei sjølvstakt ut. Men manglande treff i korpuset for ord som er ein del av det mentale leksikonet vårt, kan også bli brukt til å evaluere korpusmaterialet vårt. Dersom vi t.d. kan vise at det i ordboka finst ord som ikkje er belagde i store korpus på fleire hundre millionar ord (eller ikkje på sjølve Internett), og det viser seg at desse orda faktisk inngår i språkkunnskapen til morsmålsinformantar, kan denne metoden vise systematiske slagsider ved korpussamlingane. Arvi Hurskai-

nen sitt poeng er likevel å vise kva av ordbøkene hans som på den mest økonomiske måten (med færrest moglege lemma) er i stand til å dekkje størst mogleg del av testkorpuset.

Føresetnaden for å kunne gjere nytte av språketechnologi i leksikografisk basert programvare er at det leksikografiske grunnlagsmaterialet i seg sjølv er strukturert på ein maskinleseleg måte.

6. Bruk av leksikografisk informasjon i språktechnologien

For språktechnologien er lemmalista leksikografanes viktigaste bidrag. Utan lemmaliste blir det ingen analysator. Deretter gjev ordboka bøyingsklasseinformasjonen for kvart einskilt lemma, og ein god ordboksgrammatikk knytt til ordboka og bøyingsklasseinformasjonen vil vere betre enn ein allmenn grammatikk, i og med at ordboksgrammatikken er laga for å gje informasjon om bøyinga til kvart einskilt lemma.

Mange ordbøker har valensinformasjon, enten i form av eksempelsetningar eller i form av kodar for ulike valensklasser. Brukarar og maskiner har ulike preferansar; der maskinene må ha eintydige kategoriar, må brukarane ha døme eller ordformer det er lett å hugse tydinga av. Databasen kan sjølvsgt innehalde begge delar. Viss denne informasjonen blir formalisert, kan han også bidra til språktechnologisk bruk, som t.d. grammatisk disambiguering og syntaktisk analyse.

7. Konklusjon

Leksikografien er framleis langt frå å utnytte potensialet som ligg i å integrere ordbøker og språktechnologi. Dette er viktig både for utvikling av leksikografien, for effektiv evaluering av ordbøker, og

for brukaren, som får tilgang til nye bruksområde for ordbøkene. Språkteknologien på si side har no robuste analysatorar og kan utnytte den semantiske kunnskapen leksikografane har skaffa fram.

På denne måten vil den integrerte leksikografien og språkteknologien oppfylle grammatikaren sin draum om å modellere menneskets språkkunnskap: kunnskap om alle orda – og om korleis dei kan bli brukt.

Litteratur

Ordbøker

BO = Boye Wangensteen (red.) 2005: *Bokmålsordboka. Definisjons- og rettskrivningsordbok*. 3. utgave. Oslo: Kunnskapsforlaget.

FO = Jóhan Hendrik W. Poulsen, Marjun Simonsen, Jógvan í Lon Jacobsen, Anfinnur Johansen og Zakaris Svabo Hansen 1998: *Føroysk orðabók*. Band 1–2. Tórshavn: Føroya Fróðskaparfelag.

SKP = Risto Haarala m.fl. 2001: *Suomen kielen perussanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 55. Helsinki.

SSS = Pekka Sammallahti 1989: *Sámi-suoma sátnegirji = Saamelais-suomalainen sanakirja*. Ohcejohka: Jorgaleaddji.

Annan litteratur

Antonsen, Lene/Trond Trosterud/Ciprian-Virgil Gerstenberger/Sjur Nørstebø Moshagen 2009: Ei intelligent ordbok for samisk. I: *LexicoNordica* 16, 271–283.

Beesley, Kenneth R. og Lauri Karttunen 2003: *Finite State Morphology*. Palo Alto, CA: CSLI Publications.

Hurskainen, Arvi 2004: Computational testing of five Swahili dictionaries. I: *20th Scandinavian Conference of Linguistics*. <http://www.ling.helsinki.fi/kielitiede/20scl/Hurskainen.pdf>

- Karlsson, Fred, Atro Voutilainen, Juha Heikkilä og Arto Anttila (red.) 1995: *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Natural Language Processing, No 4. Berlin/New York: Mouton de Gruyter.
- Pirinen, Tommi 2008: Suomen kielen äärellistilainen automaattinen morfologinen analyysi avoimen lähdekoodin menetelmin, pro gradu -tutkielma. <http://www.helsinki.fi/~tapirine/gradu/Pirinen2008.pdf>
- Trosterud, Trond 2003: Ordbokskritikk. I: *LexicoNordica* 10, 65–88.
- Trosterud, Trond 2009: A constraint grammar for Faroese. I: *NE-ALT Proceedings Series* 2009; Vol. 8. 1–7. <http://dspace.utlib.ee/dspace/handle/10062/14285>

Internettreferansar

- FAOFST = Trond Trosterud 2009: *Færøysk morfologisk analyseprogram*. <http://giellatekno.uit.no/cgi/index.fao.nno.html>
- KEFD = Jack Rueter/Ciprian Gerstenberger/Trond Trosterud 2010: *Komi–English–Finnish Dictionary*. https://victorio.uit.no/langtech/trunk/kt/kom/src/working_files/
- KOMFST = Trond Trosterud/Ciprian Gerstenberger/Jack Rueter 2010: *Komi morfologisk analyseprogram*. <http://giellatekno.uit.no/cgi/index.kom.nno.html>

Trond Trosterud
 førsteamanuensis, ph.d.
 Fakultet for humaniora, samfunns-
 vitenskap og lærerutdanning
 Universitetet i Tromsø
 NO-9037 Tromsø
 trond.trosterud@uit.no