

Lemmatisering/normalisering: diskussionspunkter

Helle Degnbol og Alex S. Kjeldsen

20. november 2009, revidert 24. november 2009

1 Teoretisk diskussion → praktisk udkomme: retningslinjer og redskaber

- teoretisk
 - hvorfor normaliserer man? for hvem? hvordan? til glæde for hvem? til besvær for hvem?
 - hensyn til: normalisatør, læser (forskellige behov), databank
 - lemmatisering vs. normalisering
- * normalisering af lemmata
- * normalisering af løbende tekst
- praktisk
 - hvilke papirer og redskaber eksisterer?
 - hvilke vil vi arbejde videre med?
 - hvilke problemtyper/problemer?
 - fordeling af arbejdsopgaver

2 Afgrænsning

- vestnordisk vs. østnordisk
- prosa vs. poesi
 - spec. krav/ønsker til normalisering af poesi?
 - eddadigte vs. skjaldedigte?
 - i prosimetrum: tekstens enhed?

3 Lemmatisering

- dynamiske datasamlinger, fluktuerende data
 - mulighed for opdatering af lemmata i Menotas tekster¹
 - hvordan kommunikerer ONP og Menota ved ændringer? (logfil? versionskontrol?)
 - markering af særlig vanskelige ord/ordgrupper? (fx *ór-/ør-*)
- disambiguering: ord + homografnummer (à la ONP?)
- harmonisering mht. grammatisk bestemmelse (à la ONP?)
- diverse problemtyper (herunder problemer ved opmærkning):
 - et eller flere ord: *tryggva, tryggja, kveykva, kveikja*
 - ordforbindelse eller ord: *alls staðar, í hjá, upp á, framan af*
 - *øllu* (adj. dat. sg. brugt som adv.)
 - *handan* (adv. brugt som præp.)

¹ Jf. Odd Einars kommentar i forbindelse med referat fra sidst rådsmøde: “Vi tenkjer oss definitivt at tekstar kan oppdaterast, både ved at utgjevaren sjølv rettar feil, eller ved at andre supplerer teksten med t.d. normalisering, lemmatisering e.l. Dei enkelte bidragsytarane vil bli krediterte i header til teksten.”

- præet. part.: *bráddr* under *bráða* vb., *saddr* under *seðja* vb. vs. *aldráðr* adj. (kun præet. part. belagt)
- grammatisk grundform: *lög* under *lag*, *danir* under *danr*, *páskar/páskir* under *páskr* (undtagelsesvis plurale tantum: *dyrr*, stednavne)

4 Løbende tekst: Menota-normalisering vs. individuel normalisering (håndskriftintern, national, moderne, poetisk...)

- hvilke muligheder hos Menota: flere norm-niveauer?

1. menotisk normalisering (ONP)
2. individuel normalisering (med beskrivelse)
3. begge dele?²

- rigiditet

- ortografisk (svarende til ONP; men hvad med sideformer?)³

* *glík-/lík-*

* *nekkverr/nøkkurr*

- morfologisk: med ensretning à la Wimmers læsebog eller med respekt for sprogets/tekstens grammatiske finurligheder?

* *mundi/myndi*

* *eigu/eiga*

* *fjórðu/firði*

5 Mulige arbejdsopgaver

- håndbog i normalisering inkl. grammatik

- praktisk orienteret mod lemmatisering/normalisering, en guide for lemmatisør/normalisatør med udgangspunkt i håndbogens lemmatiseringskapitel (grammatikken ideelt set genereret fra Menotas normaliserede tekster); publiceres som *work in progress*, brugerne skal lettest muligt kunne stille spørgsmål og levere kommentarer til gavn for den fortsatte redigering af grammatikken

- diskussion af syntaktisk analyse

- fx. ved præp./adv. (jf. ONP 1-3: Nøgle, 32-35)

- fx. ved verber: løs/fast sammensætning, simplex/partikel+verbum

- forholdet til Dag Haugs syntaktiske projekt

- normalisering uden for murene: (løselig, oversigtlig) beskrivelse af diverse allerede foreliggende normaliseringer (inkl. Menotas tekster og ordbøgers lemmata)

- særlige forholdsregler ved poesi: muligheder for samarbejde med andre projekter, fx skjaldeprojektet.

² Jf. Odd Einars kommentar i forbindelse med referat fra sidst rådsmøde: “Derimot er det ikkje enkelt å tenkje seg alternative normaliseringar. Her trur eg at ein må leve med at det er ein viss variasjon. Derfor er det desto viktigare at lemmatiseringa er koordinert.”

³ http://onp/red_onporto-2009-september1.pdf