

# YSO : kieltä ja kulttuuria

4.12.2013

[susanna.nykyri@helsinki.fi](mailto:susanna.nykyri@helsinki.fi)

# Yleinen suomalainen ontologia

- Yleisyys?
- Suomalaisuus?

Kuitenkin kielinä myös ruotsi + englanti

# Monikielisyys vs. monikulttuurisuus

- Esimerkkinä nationalismi YSAssa, HASSETissa ja UNESCO-tesauruksessa

# YSA

Käytettävä asiasana: **nationalismi**

Laajemmat termit: **ideologiat**

Rinnakkaistermit: **etnosentrismi**  
**fennomania**  
**isänmaallisuus**  
**kansallisuusaate**  
**kansallisuuskysymykset**  
**panslavismi**  
**ruotsinmielisyys**  
**slavofiilit**  
**suomalaisuus**

Korvaa termit: **kansallismielisyys**

Kuuluu ryhmiin: [65] Valtio-oppi. Poliitikka. Kansainvälinen politiikka  
Ruotsinkielinen asiasana: nationalism

# HASSET

Current term		
NATIONALISM		
Broader terms	<input type="checkbox"/> Narrower terms	<input type="checkbox"/> Related terms
<ul style="list-style-type: none"><li>▶ <a href="#">IDEOLOGIES</a></li><li>▶ <a href="#">POLITICAL DOCTRINES</a></li><li>▶ <a href="#">POLITICS</a></li></ul>	<ul style="list-style-type: none"><li>▶ <a href="#">CULTURAL NATIONALISM</a></li><li>▶ <a href="#">SCOTTISH NATIONALISM</a></li><li>▶ <a href="#">ZIONISM</a></li></ul>	<ul style="list-style-type: none"><li>▶ <a href="#">ETHNOCENTRISM</a></li><li>▶ <a href="#">FASCISM</a></li><li>▶ <a href="#">NATIONAL PRESTIGE</a></li><li>▶ <a href="#">NATIONAL SOCIALISM</a></li><li>▶ <a href="#">NEO-NAZISM</a></li><li>▶ <a href="#">PATRIOTISM</a></li><li>▶ <a href="#">PLAID CYMRU</a></li></ul>

# UNESCO

Nationalism

MT 6.15 Politics and government

FR Nationalisme

SP Nacionalismo

BT1 Political doctrines

RT Cultural nationalism

RT Fascism

RT Nations

RT Nazism

RT Regionalism

RT Separatism

# Kielet kielten sisällä – diskurssit ja odotukset

- Asiasanoitustyön joukkoistaminen webissä?
- Sisällönkuvailun funktio?
- Esimerkkeinä päiväkirjatyypiset blogitekstit

# ***Päiväkirjablogien metadatan funktioista***

***Susanna Nykyri, Informationsvetenskap, Åbo Akademi***

***&***

***Risto Niemi-Pynttäri, Taiteiden ja kulttuurin tutkimuksen laitos,  
Jyväskylän yliopisto***



- Blogit – muoto vs. sisältö
- Subjektiiiviset tekstimuotoiset päiväkirjatyyppiset blogit

-> **tekstityyppinä** edustavat **verkkoproosaa**

<- luovaa, dialogista, puheenomaista, mutta tyyliteltyä kirjoitusta (Niemi-Pynttäri 2007)

- **Asiasanoitus** (indeksointi, dokumentaatio) ymmärretään alan standardeissa (ISO 5963-1985) tekona, jossa kuvaillaan tai identifioidaan dokumentti sen asiasisällön perusteella. **Asiasanat** ovat metadataa eli tietoa tiedosta, ja mikäli ne on valittu hyvin, ne kuvaavat viittaamansa dokumentin sisällön keskeiset aiheet (Milstead & Feldman 1999).

# Tags

” A **tag** is a keyword which describes all or part of a Post. Think of it like a Category, but smaller in scope. A post may have several tags, many of which relate to it only peripherally. Like Categories, Tags are usually linked to a page which shows all posts having the same tag. Unlike Categories, Tags can be created on-the-fly, by simply typing them into the tag field.

Tags can also be displayed in "clouds" which show large numbers of Tags in various sizes, colors, etc. This allows for a sort of total perspective on the blog, allowing people to see the sort of things your blog is about most.

Many people confuse Tags and Categories, but the difference is easy: Categories generally don't change often, while your Tags usually change with every Post.”  
(*WordPress Glossary*)

**-Huom. Tagittamisen spontaanisuus ja dynaamisuus!**

# LABELS

” Labels are handy if you blog about a lot of different subjects, since readers can then choose to view all of your posts on, say, dancing at once. Or politics, or knitting, or whatever you happen to be writing about. If you have a team blog, you could also give each member their own label, so you could easily read all the posts by any one of them. Feel free to come up with other creative new uses as well!”

*(How do I label my posts? – Blogger help, URL:  
<http://www.google.com/support/blogger/bin/answer.py?hl=en&answer=44498>)*

**Huom. Tunnisteiden luova käyttö!**

# Esim. 1

”Taisin edellisessä postauksessa tai sen kommentteissa väittää, ettei tässä vieroituksessa voi tulla ylläreitä. Olin väärässä. Ja reilusti. Nyt on nimittäin käynyt sillä tavalla, etten olekaan kipeä. Alle kaksi viikkoa mennyt ja pystyn toimimaan melkein normaalisti melkein kaikessa. Okei, nukun huonosti, mutta nukun kuitenkin. Kylmä-kuuma-oloja tulee muutaman kerran päivässä ja niihin auttaa yksi britlo. Jalat on vähän jäykät, mutta ei ole tietoaakaan siitä surraavasta hermokivusta, jota aina refloissa pelkään ihan hysteerisesti. Myöskään käsivarsissa ei ole kipuja. Maha on vähän sekaisin, mutta kertaakaan ei ole tullut sellaista fiilistä, että pärähtää lahkeeseen, etten voi lähteä ovesta ulos. Haukottelen aika paljon ja pupillit on välillä ihan tajuttoman kokoiset, mutta kumpikaan oireista ei varsinaisesti haittaa minua mitenkään. Okei, olen välillä niin väsynyt, että hyvä kun en ala itkeä kun pitäisi lähteä töihin tai koiran kanssa ulos, mutta reflat alkaa olla melkein ohi! Tämä on aika helvetin käsittämätöntä! Syön paljon Tramalia, koska sitä on ja koska en todellakaan kaipaa mitään kipuja tai ripaskaa, mutta tänään söin aamulla 300mg Tramalia ja yhden britlon, sekä puoliltapäivin taas britlon ja olen aivan fine. - - - ”

\* Tunnisteet: **ei pysty ymmärtämään, elämä, enkä muuten nuku, huumeet, katko, kipu ja turska, ylläri**

# Esim. 2

” Huikeat syysääät. Aurinko häikäisee aamusta asti, katoilla on yön jäljiltä kuuraa. Tiaiset kiipeilevät amppeleissa ja ikkunalaudoilla. On perusruoan aika. Pikkulinnuille siemeniä ja pähkinöitä. Meille ihmisille lapsuusruokaa: maksalaatikkoa ja puolukkasurvosta, hernekeittoa ja näkkileipää, jauhelihakastiketta ja punajuuria, keitettyä kukkakaalia, päälle iso nokare voita.

En murehdi. Nukun yöt hyvin, aamut kirjoitan. On levollinen olo. Yhdeltä haen tyttären, pyöräilemme syysmaisemassa, kotona syömme kanelipullaa.”

\* Tunnisteet: **arki, kirjoittaminen, linnut, ruoka, syksy, tytär**

# Sosiaalisen median diskursseista...

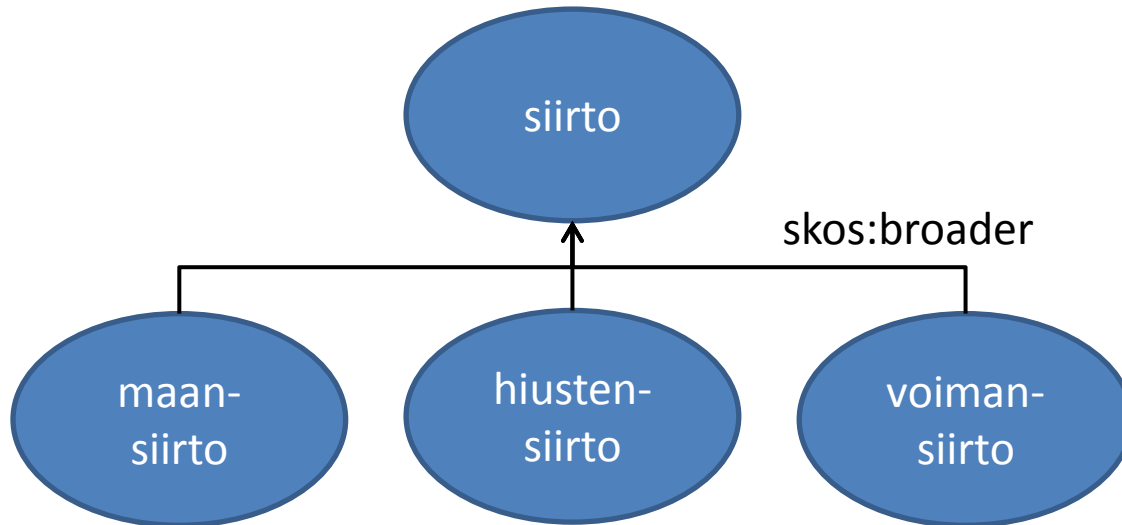
- **Asiasanat ≠ Tagit / Labelit / Kategoriat**
- **Asiasanat edustavat metadataa, mutta blogien sisällönkuvailu on usein osa verkkoproosaa**
- **Interaktiivinen ja diskursiivinen verkkoproosa ja metadata**
- **Sisällönkuvailussa ja tiedonhaussa blogien sisäinen koherenssi vs. blogien välinen koherenssi**

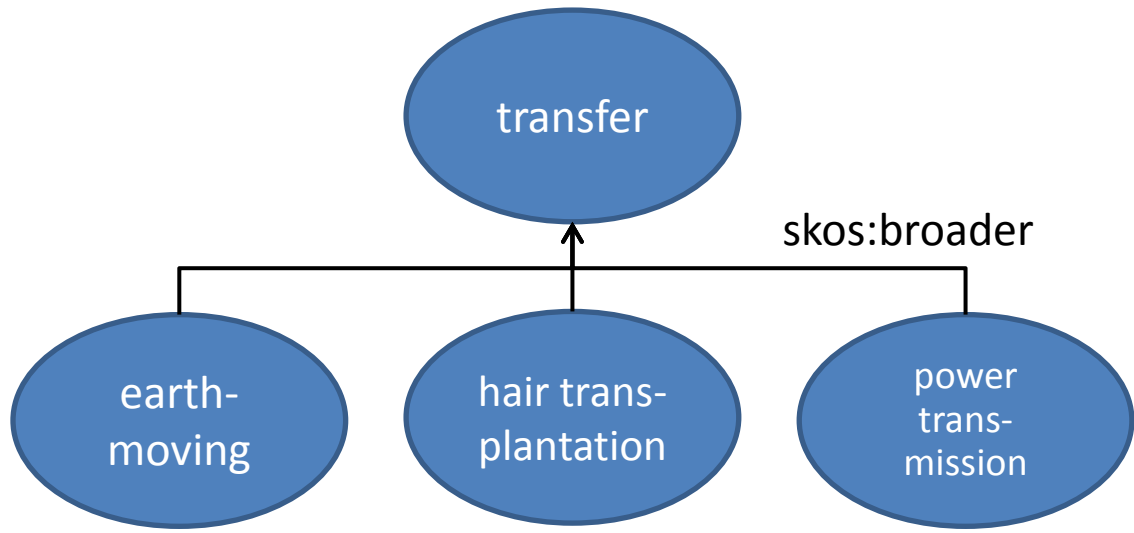
# Ontologia?

- Mitä käännetään? Käsitteitä vai termejä?
- - > nykyisessä YSO-työssä fokus siirtynyt tiedonhakijaan, joten mietittävä aiempaa enemmän käsitetasoa, intuitiivista ja temaattista etenemistä, ja tiedonhakukyselyjen laajentamista/supistamista
- Esimerkkinä siirto (transfer)



# Kieli vaikuttaa hierarkiaan





# Johtopäätökset

- Yksi suomalainen ontologia.
- Monikielisyys osaltaan paljastaa käsitetason epäloogisuudet, ja haastaa ilmaisemaan tarkemmin ja sijoittamaan käsitteet niiden tosiasiallisesti keskeisimmän merkityssisällön mukaan.

**Kiitos!**

# ARBETET MED EN FLERSPRÅKIG ONTOLOGI – DEN SVENSKA VERSIONEN

Maria Forsén

---

# UTMANINGEN MED FLERSPRÅKIGHET I ONTOLOGIER

- Utmaningen är att skapa ontologier på olika språk som till innehåll och form motsvarar varandra, men samtidigt även fungerar skilt på vardera språket och följer det egna språkets villkor
- Tekniska lösningar finns, men betraktat ur filosofisk och språklig , samt användarvänlig synvinkel uppstår många problem
- Den preliminära översättningen av ontologin, på svenska hittills kallad ALLSO har utvecklats i enlighet med YSO och samtidigt på basen av ALLÄRS
  - I den finns motsvarande problem samt därtill sådana brister som har att göra med flerspråkigheten

# TIDIGARE RIKTLINJER

---

- Från tidigare finns standarder och regler för flerspråkiga *tesaurer*
  - Documentation – Guidelines for the Establishment and Development of Multilingual Thesauri (ISO 5964-1985)
  - IFLA Guidelines for Multilingual Thesauri (cop. 2009)
- **Ingen standard för flerspråkiga *ontologier***, men delar av dessa kan fungera som stöd och ge riktlinjer
  - I vilken mån kan och bör vi följa tidigare normer?

# PRINCIPIELLA BESLUT

---

- Grunden för arbetet med en flerspråkig ontologi är principiella beslut om hurdana språkversioner av ontologin man eftersträvar
  - **En flerspråkig ontologi som beskriver en kultur eller en flerspråkig ontologi som gäller flera olika kulturer**
  - **En ontologi uppbyggd utifrån ett språk och enbart översatt till ett annat eller en ontologi som byggs upp och utvecklas utifrån flera språk**
- Målet är nu att skapa en finländsk ontologi med i princip samma innehåll på tre språk
  - Den svenska versionen bör inte avvika från den finska (såsom Allärs kan ha gjort), utan varje term bör ha en och endast en svensk motsvarighet samt vara relaterade på samma sätt som de finska termerna
  - Utvecklingen kräver intensifierat samarbete mellan olika språkversioner



# UTMANINGAR I ARBETET MED FLERSPRÅKIGA ONTOLOGIER: EKVIVALENS

---

Ekvivalens (motsvarighet, likvärdighet)

- Ekvivalens-problematiken blir i en flerspråkig ontologi dubbel
  - Förutom den vanliga ekvivalensen mellan termer inom samma språk tillkommer ekvivalensen mellan termer på olika språk
- Ekvivalens mellan termer på olika språk har semantiska, kulturella och strukturella aspekter
- Ekvivalens mellan termerna återfinns i olika grad på en skala mellan exakt, oexakt eller nära, partiell, en-till-flera och icke ekvivalens
  - Exempel
    - bibliotek = kirjastot
    - omstridda fenomen = rajatieto
    - siare, spåkvinnor/män = ennustajat
    - redaktörer + journalister + reportrar = toimittajat
    - ? = hyppyherätys

# UTMANINGAR I ARBETET MED FLERSPRÅKIGA ONTOLOGIER: STRUKTUR

---

Strukturella problem i vokabulären gäller relationerna mellan termerna

- Ekvivalenta och associativa relationer utgör sällan något problem - såvida huvudtermerna är korrekta, men de hierarkiska relationerna bör utredas närmare även på målspråket
- Strukturen bör byggas upp så att hierarkin och associationerna blir sammanhängande på samtliga språk
  - Exempel
    - astronomi, ersätter begrepp -  
= tähtitiede, korvaa termit: astronomia
    - brottsanmälningar, överordnade begrepp: annonser (!)  
= rikosilmoitukset, yläkäsite: ilmoitukset
    - kronologi, relaterat begrepp: tideräkning  
= kronologia, rinnakkaiskäsite: ajanlasku

# ANDRA FRÅGOR I ARBETET MED FLERSPRÅKIGA ONTOLOGIER

---

## Termernas utformning

- Språkens struktur och konventioner gällande användningen av olika ordtyper och -former varierar, t.ex. i finskan förekommer ofta nominalisering, passiv- och agentivhärledning
- Exempel
  - bli vän med = ystäväystyminen
  - förflutna (det) = menneisyys
  - hjärtat = sydän
  - lokal anknytning = paikallisuus
  - mässor = messut, mässa (gudstjänst) = messu

## Ord med samma form som har flera betydelser (homografer, polysemer)

- Ord som skrivs likadant men har olika betydelser förekommer både inom språk och mellan språk
- Om inte andra synonyma termer kan användas i stället bör deras mening klargöras med parenteser
- Exempel
  - val (politik) = vaalit, val (selektion) = valinta, valar = valaat

# ANDRA FRÅGOR I ARBETET MED FLERSPRÅKIGA ONTOLOGIER

## FORTS.

---

### Lånade termer (från andra språk)

- Bör godkännas som huvudtermer endast om ingen annan tillfredställande lösning finns, såsom i fall där
  - Översättningen av en term annars skulle kräva en lång definition eller förklaring som inte bör användas som indexterm
  - Termen hänvisar till ett begrepp som är "infött" hos användarna av källspråket och detta begrepp osannolikt kommer att uppstå hos målspråkets användare
  - Exempel
    - herännäisyys (finländsk pietism)
    - leasing
- Ibland kan en lånad term och en förmodad översättning existera samtidigt

### Nyskapade termer (neologismer)

- Bör godkännas som huvudtermer endast om ingen annan tillfredsställande lösning finns
- Bör användas endast om det är sannolikt att de blir allmänt accepterade och efter konsultation med indexerare, språkvetare och ämnesexperter
- Exempel
  - alkioism = alkiolaisuus

# ANDRA FRÅGOR I ARBETET MED FLERSPRÅKIGA ONTOLOGIER FORTS.

---

## Termer som ändrats p.g.a. förändringar i språkbruk

- En del termer ändras med tiden på ett språk, men inte nödvändigtvis på det andra
  - Exempel
    - andelskassor -> andelsbanker = osuuskassat -> osuuspankit
    - elektroniska publikationer = elektroniset julkaisut -> sähköiset julkaisut

## Rikssvenska termer

En del begrepp benämns på olika sätt i Sverige och svenska Finland

- I vissa fall är den finlandssvenska termen mycket vedertagen och i dessa fall kan den rikssvenska termen bli ersatt
- Exempel
  - glansbilder / bokmärken (SE) = kiiltokuvat
  - postlådor / brevlådor (SE) = postilaatikot

# ANDRA FRÅGOR I ARBETET MED FLERSPRÅKIGA ONTOLOGIER FORTS.

---

## Användningen av parenteser och anmärkningar

- Parenteser och anmärkningar är viktiga och vanligt förekommande i det flerspråkiga ontologi-arbetet
  - För att förklara och definiera termer och deras betydelse eller sammanhang
  - För att beskriva användningen av termer
  - För att klargöra skillnader mellan olika språk
- För att uppnå konsekvens och enhetlighet bör man bestämma hur parenteser och anmärkningar används och hur de utformas
- Exempel
  - tornugglor (familj)
  - skrivare (personer)
  - reduktion (kemi)
  - deglaciation : Isavsmältning, frismältning av landområden genom minskning av inlandsisens eller andra glaciärers utbredning

# KONTAKT

---

- Kommentarer, frågor, förslag m.m. som gäller den svenskspråkiga tesauren eller ontologin kan skickas till **maria.forsen@abo.fi**
- Tack!

# ARBETSGRUPPENS FÖR AUKTORISERADE SVENSKA INDEXERINGSTERMER FÖRSLAG TILL SVENSKSPRÅKIGT NAMN PÅ DEN ALLMÄNNA FINLÄNDSKA ONTOLOGIN

Den svenska versionen av ontologin Yleinen suomalainen ontologia - YSO har tidigare kallats Allmän svensk ontologi - ALLSO. Detta kan inte anses korrekt med tanke på sammanhanget och omständigheterna kring ontologin. För det första kan det missförstås så att ontologin har något att göra med Sverige och rikssvenska förhållanden, vilket inte är fallet. Det finska namnet hänvisar till Finland och den finländska kulturen (i vilken den finlandssvenska ingår) och ontologins innehåll skapas i det finländska sammanhanget och avspeglar denna miljö. I ontologin är också förhållandet mellan den finska och den svenska delen ett annat än det varit i tesaurerna. Ontologin är i princip en enda och strävan är att innehållet ska vara det samma på begreppsnivå och presenteras med termer på tre språk, vilket gör att även namnen på de olika språken till sin betydelse bör vara de samma. Med tanke på dessa faktiska omständigheter vore den korrekta översättningen av namnet till svenska Allmän finländsk ontologi. Akronymen AFO är redan upptagen för en ontologi i ONKI-systemet så denna är utesluten som namnförkortning. En naturlig förkortning vore i stället ALLFO, i vilken den tidigare delen kan anknytas till orden allmän och allt och med vilken den allmänna tesaurusens namn ALLÄRS lätt associeras. Arbetsgruppen för auktoriserade svenska indexeringstermer föreslår därmed att den svenskspråkiga versionen av ontologin får namnet Allmän finländsk ontologi med förkortningen ALLFO.



# YSO: kääntäminen ja siltaus

Satu Niininen 4.12.2013

# FI-EN kääntäminen

# Kääntämisen taustoista

- Tällä hetkellä YSO sisältää noin 26 100 käsitettä, joista n. 23 000 käännetty englanniksi.
- YSOon lisättiin 2007 Helsingin kaupunginkirjaston luokitusjärjestelmästä (HKLJ) saadut YSO:n käsitteitä vastaavat englanninkieliset vastineet.
- Vuonna 2009 käännettiin puuttuvat 6 000 käsitettä sekä 2 000 käsitettä, jotka olivat tulleet YSOon YSA-päivitysten kautta.

# Kääntämisestä 2

- HKLJ-käännösten yhteydessä ei huomioitu ontologian rakennetta eikä sanamuoto-ohjeistusta  
→ näiden tarkistus kokonaan tekemättä.
- 2007 ja 2009 käännösten yhdenmukaisuus tarkistamatta.
- Yli 3 000 käsitettä kokonaan ilman englanninkielistä vastinetta

# Haasteita

**Käännösstrategian valinta: sanasanainen käännös, selittävä käännös, kääntämättä jättäminen?**

Eläintarhan ajot, Eläintarha race, Eläintarhan ajot race, Helsinki Zoo race?

**Suomalaisen kulttuuripiirin erikoisterminologia**

erillistalot, iltalukiot

**Kulttuuriset jäsenyserot ja sanojen monimerkityksisyys**

esim. kasvattajat, kalusto/laitteisto/varusteet

heimosodat - tribal wars vai Kindred Nations Wars

# Siltaus

# Siltauksen työprosessi

- Testisiltaus 80 satunnaisesti valitulla YSO-käsitteellä (Wordnet 37 % LCSH 60 %, Freebase 61 %)
- Kääntämisen yhteydessä siltaus Library of Congress Subject Headingsiin (LCSH)
- Kun käännös on valmis, automaattinen siltaus Freebaseen (tarkistus ihmistyönä)
- Luodaan kriteeristö ekvivalenssin määrittämiseen
  - Siltaussuhteet closeMatch ja exactMatch
  - Millaisten kriteerien on täytyttävä, jotta vastaavuuden voidaan sanoa olevan close tai exact?

# Siltaushaasteita

Enemmistö käsitteistä tulee todennäköisesti jäämään ilman siltauskohdetta

→ Aitoon kirkastusjuhlat , ilmakiväärit, kaneli

Käsitteen merkityksen ja käyttökontekstin symmetrisyys

→ armeliaisuus – clemency ; huovat – blankets

Ratkaisujen yhtenäisyys

→ Tarvitaan toimivat perusperiaatteet, joita voidaan soveltaa eri tapauksissa



# Kiitos!

satu.niinen@helsinki.fi



KANSALLISKIRJASTO