

MOLTO-konekäännösprojekti ja Grammatical Framework monikielisissä sovelluksissa

KITES-aamiaisseminaari

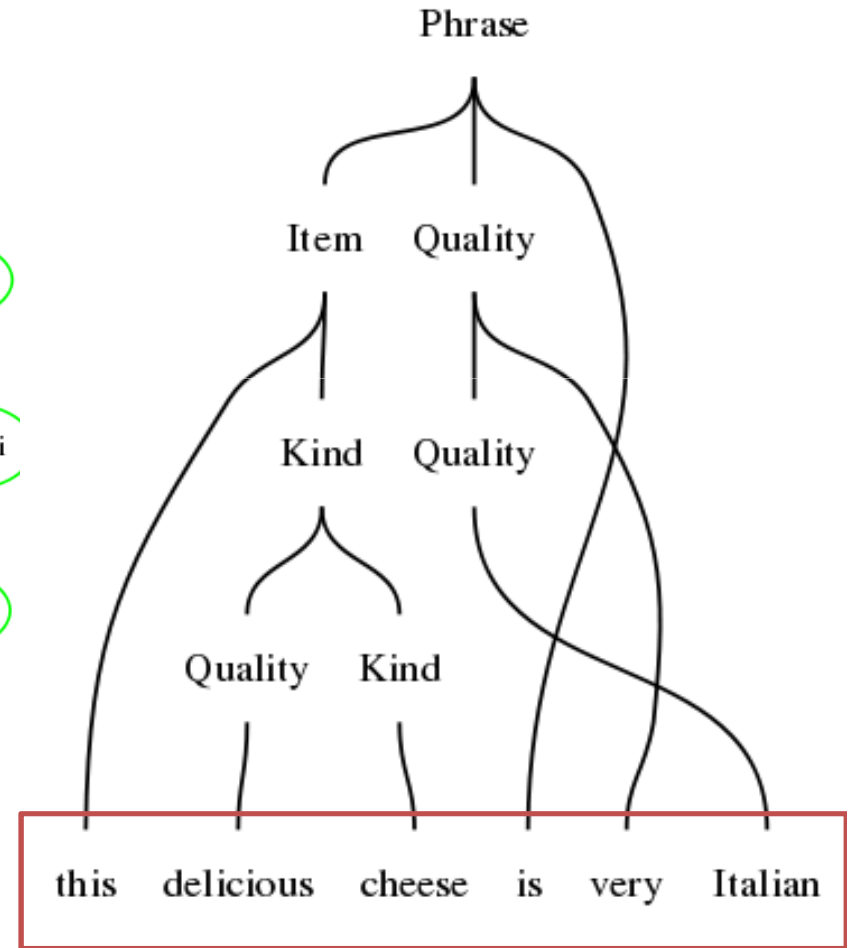
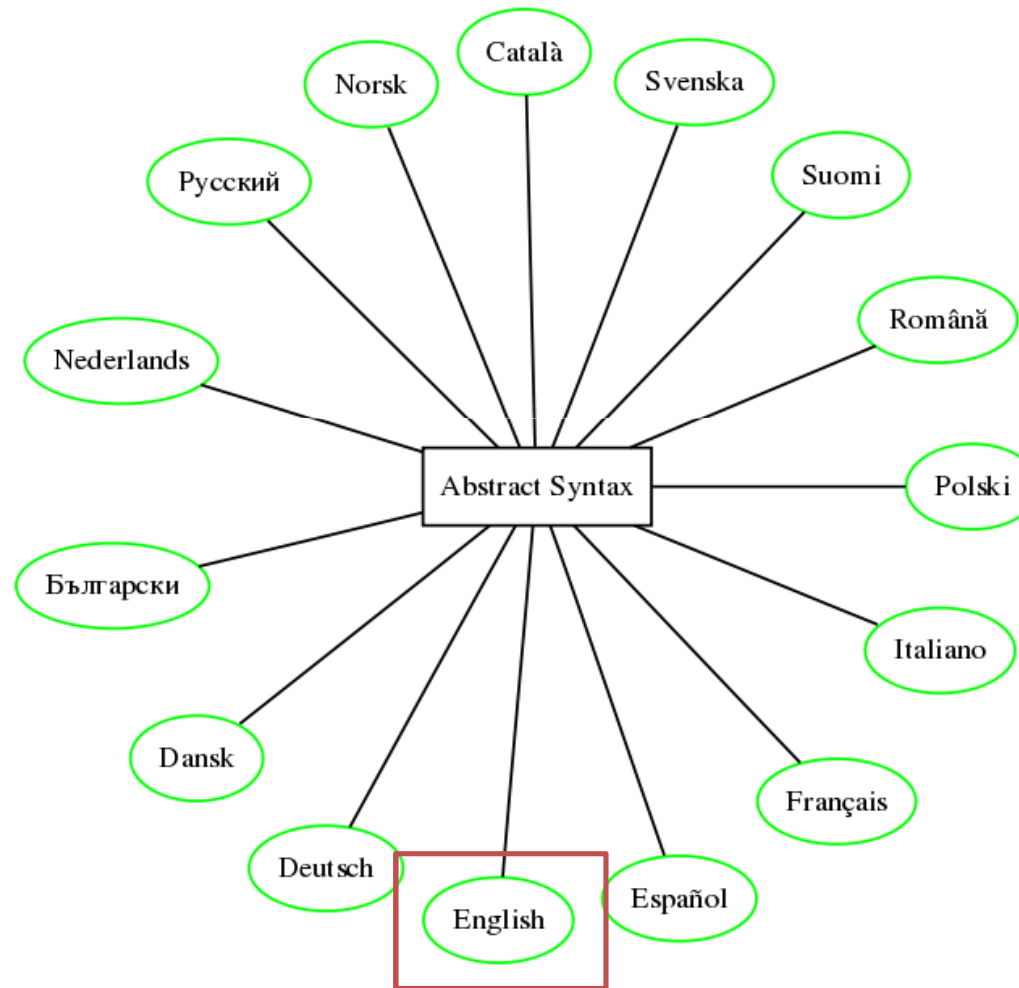
29. marraskuuta 2012

Seppo . Nyrkko (ät) helsinki .fi

Grammatical Framework

- Haskell-pohjainen, funktionaalinen ohjelmointikieli erityisesti NLP-ohjelmointiin
- Jäsennin + Generoija = Konekäännösympäristö
- Resurssikirjasto (Resource Grammar Library)
 - 26 kieltä (myös kiina)
 - Taivutusparadigmat, syntaktisia sääntöjä
 - Mallileksikko (sanasto)
 - Joitakin lauserakenteita

GF: Kielimalli



Työkaluja

- Eclipse plugin kielimallin kehitykseen
- Portable Grammar Format
 - ajonaikaiseen käyttöön
 - Haskell
 - Java, in particular the Android platform
 - JavaScript
 - PGF runtime API for C (vrt <> Gettext-kirjasto)
- Binäärikääntäjä

GF: Saatavuus ja yhteensopivuus

- GF is **open-source**, licensed under [GPL](#) (the program) and [LGPL](#) and [BSD](#) (the libraries). It is available for
 - Linux
 - Mac OS X
 - Windows
 - Android mobile platform (via Java; runtime)
 - via compilation to JavaScript / web browser (runtime)

Sovelluskohteita GF-kielimalleille

- Luonnollinen kieli (teksti) käyttöliittymässä
- Tiedon syöttö luonnollisella kielellä
- Puheentunnistuksen kielimallin tuottaminen
- Kannettavat "low cpu" laitteet
- Upotettavat "gadgetit"
- Kielen validointi

Luonnollinen kieli käyttöliittymässä

- Lähetetty henkilö(i)lle Sami
- Päivänä 29. marraskuuta
- Liitetiedostoja 1
- Siirrä kohde Ladattu kansio kohteeseen Väliaikaiset tiedostot
- Lähetetty Samille
- Marraskuun 29:na
- Yksi liitetiedosto
- Siirrä ladattu kansio väliaikaisiin tiedostoihin

GF: Resource Language demo

```
snyrkko@ws992:~/GF$ gf
```

```
      * * *
    *       *
  *         *
 *          *
*           *
*      * * * * *
*      *       *
 *      * * * * *
  *      *       *
    *      *       *
      * * *

```

This is GF version 3.3.3-darcs.

11 recorded changes since SNAPSHOT-20121022

Last recorded change: Thu Oct 25 01:40:18 EEST 2012 hallgren@chalmers.se

Built on linux/x86_64 with ghc-7.4, flags: interrupt server

License: see help -license.

Bug reports: <http://code.google.com/p/grammatical-framework/issues/list>

Languages:

```
> i present/LangEng.gfo
```

```
linking ... OK
```

Languages: LangEng

```
404 msec
```

```
Lang> i present/LangSwe.gfo
```

```
linking ... OK
```

Languages: LangEng LangSwe

```
708 msec
```


GF: Resource Language demo

Parsing

```
Lang> p -lang=Eng "there are cats in the boat"
```

```
PhrUtt NoPConj (UttS (UseCl (TTAnt TPres ASimul) PPos (ExistNP  
  (AdvNP (DetCN (DetQuant IndefArt NumPl) (UseN cat_N)) (PrepNP  
    in_Prep (DetCN (DetQuant DefArt NumSg) (UseN boat_N))))))  
  NoVoc
```

Parsing and Linearizing

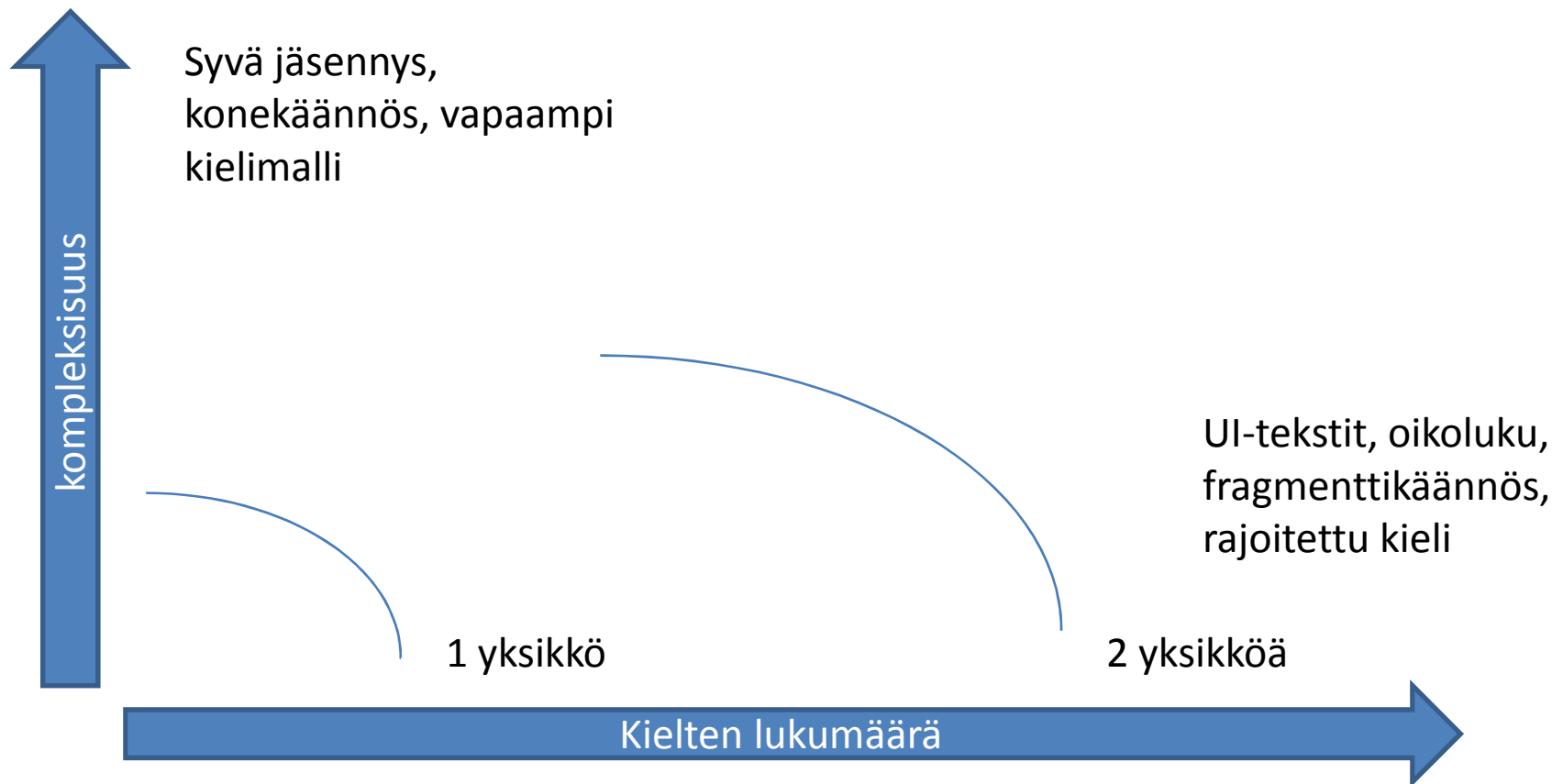
```
Lang> p -lang=Eng "there are cats in the boat" | 1
```

```
there are cats in the boat
```

```
det finns katter i båten
```

GF: Kielimallin kehittäminen

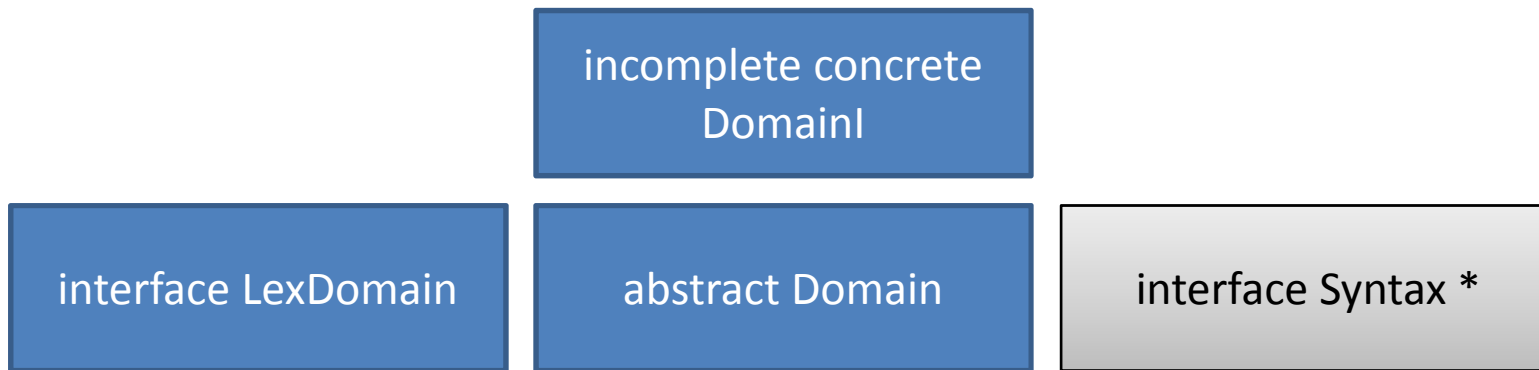
Työmäärän parametrit



GF: Kielimallin kehittäminen

“Design pattern”

Domain-spesifi kielimalli, abstrakti osa



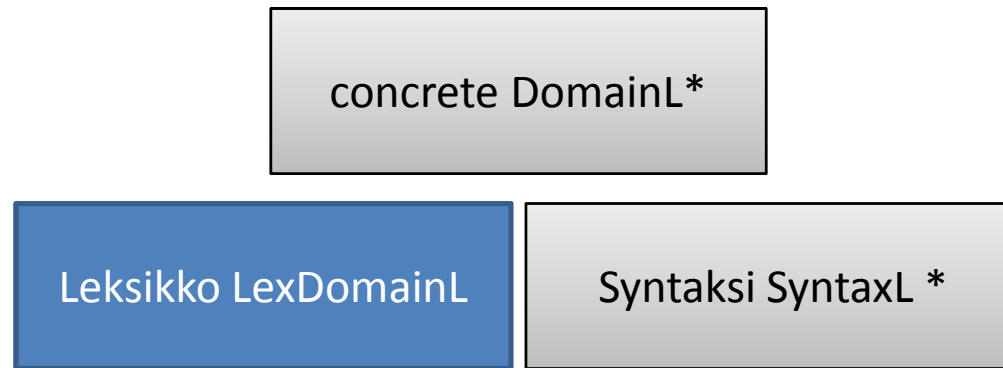
* valmiina saatava osa / “helppo toteuttaa”

Käsin kirjoitettava osuus

GF: Kielimallin kehittäminen

“Design pattern”

Kielispesifi osa: $L = \{\text{Fin, Swe, Eng, ...}\}$



* valmiina saatava osa / “helppo toteuttaa”

Käsin kirjoitettava osuus

MOLTO-project.eu

MOLTO Multilingual Online Translation
Non multa, sed multum nicht vielerlei treiben, sondern eine Sache intensiv und genau

HOME EVENTS INTERNAL WORKPLAN DELIVERABLES PUBLICATIONS FAQ

RECENT PUBLICATIONS

[Ontology-based lexicon management in a multilingual translation system - a survey of use cases](#)
[Periodic Management Report T30](#)
[WP4 Current Status - M30](#)
[Comparing human perceptions of post-editing effort with post-editing operations](#)
[More...](#)

RECENT NEWS

[Your Friendly Evaluation Coordinator here...](#)
[Call for participation: CLARA Thematic training course on evaluation of Human Language Technologies](#)
[GF mirrored on GitHub](#)
[CfP: User Interfaces for Theorem Provers UITP'12 Post-Proceedings](#)

DEMOS

[MOLTO Application Grammars](#)
[MOLTO KRI](#)
[Query Patents](#)
[GF Cloud Services](#)



AVAILABLE SIGNUPS

Congratulations, you have replied to all of the content on the site that is accepting signups.

[VIEW](#) [WHAT LINKS HERE](#)

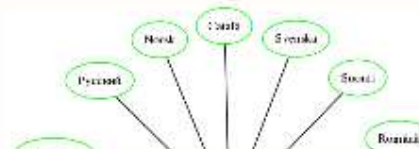
Welcome to the MOLTO website!

Sites

- UGOT: University of Gothenburg, Sweden (coordinator)
- UHEL: University of Helsinki, Finland
- UPC: Universitat Politècnica de Catalunya, Barcelona, Spain
- Ontotext: Ontotext AD, Sofia, Bulgaria
- BI: Be Informed, Apeldoorn, The Netherlands
- UZH: University of Zurich, Zürich, Switzerland

MOLTO Languages

MOLTO translation is based on GF and its resource grammar library, which currently covers the languages shown in the diagram besides here. MOLTO tools will enable the specialization of both the Abstract Syntax and the language fragments to idiomatic usages in new domains of application.



EVENTS

[Call for Book Chapter: "Towards the Multilingual Semantic Web"](#)
(to be published by Springer, autumn 2013)

(5 days)

[CfP: User Interfaces for Theorem Provers UITP'12 Post-Proceedings](#)

(18 days)

MOLTO

- EU-rahoitteinen tutkimusprojekti, 2010-2013
- Kattaa 12 Euroopan virallista kieltä
 - lisäksi 3 lähialueen kieltä (Cat, Rus, Nor)
- 4 yliopistoa:
 - Göteborg, Helsinki, UPC (Barcelona), UZH (Zürich)
- Yhteistyökumppanit:
 - Ontotext AD (Sofia, Bulgaria)
 - Be Informed (Apeldoorn, Alankomaat)

MOLTO

- Useita konekäännösmaalleja
 - Matematiikka (ongelmien ja vastausten sanallistaminen)
 - Patentit (lääketiede, kemia)
 - Kulttuuriperintö (museokielioppi)
 - Tiedon sanallistaminen (ontologiahaku, semanttinen wiki)
 - Matkaopas (Q&A, paikkoja, aikoja, lukumääriä, nimiä, laatuja)

MOLTO

- Työvälineitä konekäännökseen
 - Käännössovellukset (mobiili, web, javascript)
 - Kääntäjän työvälineet
 - Kielimallin kehittäjän apuvälineet
 - Hybridimallien kehittäminen
 - Robusti jäsenitys (avoimempi kielimalli)

MOLTO: Demoja ja sovelluksia

- Turistin monikieliopas
- Kääntäjän työkalut
- Tiedonhaku (Knowledge Retrieval)
- Semantic Wiki
- Leksikon tallennus ontologiamallissa

- Robusti jäsennin
- Hybridi käänno-smalli ("Soft" & "Hard")

Kiitos

Seppo . Nyrkko (ät) helsinki .fi

www.grammaticalframework.org

www.molto-project.eu