Multilingual Grammars for the Semantic Web

Aarne Ranta

Dagstuhl Seminar on Multilingual Semantic Web, 6 September 2012

MOLTO Multilingual Online Translation

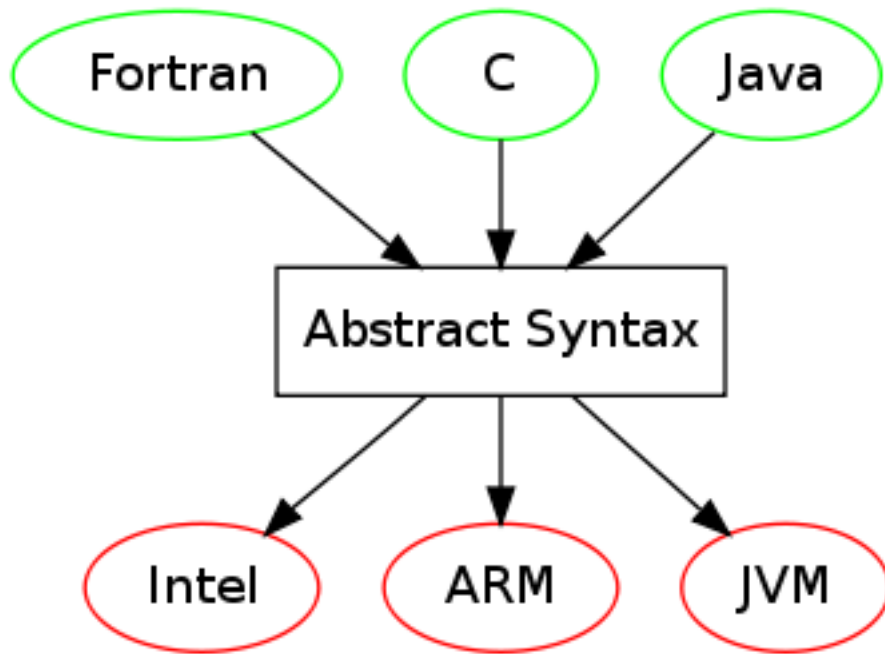Non multa, sed multum not quantity but quality

ABOUT | NEWS | EVENTS

MOLTO's mission is to develop a set of tools for translating texts between *multiple languages* in *real time* with *high quality*. MOLTO will use multilingual grammars based on semantic interlinguas.

http://www.molto-project.eu

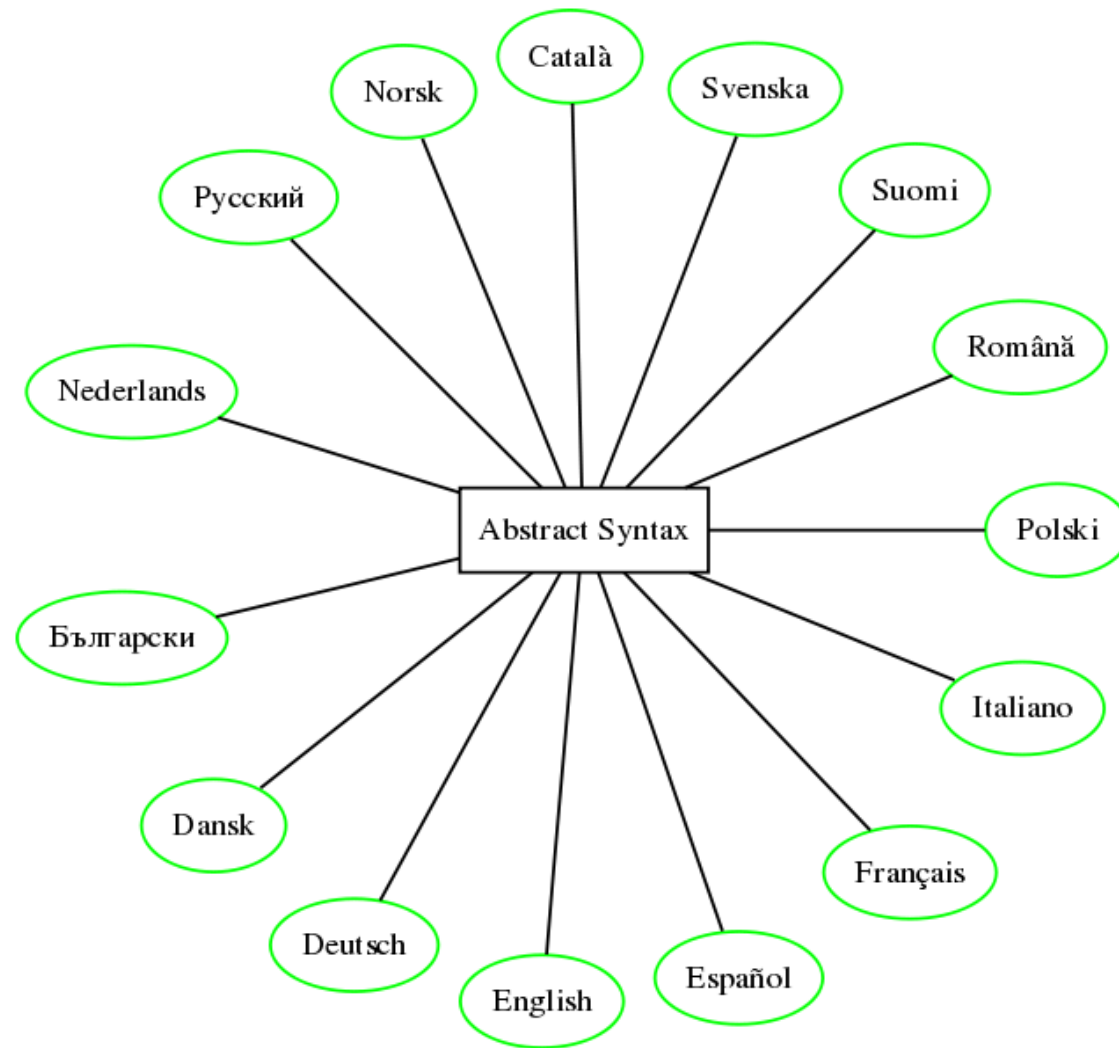# Where interlingua rules today: compilers


```
                                                        bipush 6
6 + x  -->  (EAdd TInt (EInt 6) (EVar TInt "x")) -->    iload 8
                                                        iadd
```

MOLTO languages

Two things learned from the past

No universal interlingua:

- instead, domain-specific interlinguas $\sim$ ontologies
- *The Rosetta stone is not a monolith, but a boulder field.*

Yes universal concrete syntax:

- no hand-crafted *ad hoc* grammars
- but a general-purpose **Resource Grammar Library**

Grammar example: the predicate "x likes y"

Abstract syntax:

```
fun Like : Person -> Item -> Fact
```

Concrete syntax (first approximation):

```
lin Like x y = x ++ "likes" ++ y            -- Eng
lin Like x y = x ++ "tycker om" ++ y        -- Swe
lin Like x y = x ++ "pitää" ++ y ++ "stä"   -- Fin
lin Like x y = y ++ "piace a" ++ x          -- Ita
```

Complexity of concrete syntax

Italian: agreement, rection, clitics (*il vino piace a Maria* vs. *il vino mi piace* ; *tu mi piaci*)

```
lin Like x y = y.s ! nominative ++ case x.isPron of {
  True  => x.s ! dative ++ piacere_V ! y.agr ;
  False => piacere_V ! y.agr ++ "a" ++ x.s ! accusative
  }
oper piacere_V = verbForms "piaccio" "piaci" "piace" ...
```

Moreover: contractions (*tu piaci ai bambini*), tenses, mood, ...

The GF Resource Grammar Library

Currently for 25 languages; 3-6 months for a new language.

40+ contributors 2001-

Complete morphology, comprehensive syntax, lexicon of irregular words.

Afrikaans	Bulgarian	Catalan	Danish	Dutch
English	Finnish	French	German	Hindi
Italian	Japanese	Latvian	Nepali	Norwegian
Persian	Punjabi	Polish	Romanian	Russian
Sindhi	Spanish	Swedish	Thai	Urdu

Smart paradigms

```
mkV "like"
mkV "cry"
mkV "go" "went" "gone"

mkV "pitää"
mkV "talo"
mkV "poika" "pojan"
```

Concrete syntax for predications

Common syntax API:

```
lin Like x y = mkCl x (mkV2 (mkV "like")) y        -- Eng
lin Like x y = mkCl x (mkV2 (mkV "tycker") "om") y  -- Swe
lin Like x y = mkCl x (mkV2 (mkV "pitää") elative) y -- Fin
lin Like x y = mkCl y (mkV2 piacere_V dative) x      -- Ita
```

From predications to complex phrases (MOLTO Phrase-book)

Bul:

Аз знам , че нейният съпруг харесва това италианско вино.

Cat:

Sé que aquest vi italià agrada a el seu home.

Dan:

Jeg ved at hendes mand holder af denne italienske vin.

Dut:

Ik weet dat haar man van deze Italiaanse wijn houdt.

Eng:

I know that her husband likes this Italian wine.

Fin:

Tiedän että hänen miehensa pitää tästä italialaisesta viinistä.

Fre:

Je sais que ce vin italien plaît à son mari.

Ger:

Ich weiß , dass ihr Mann diesen italienischen Wein mag.

मैं जानता हूँ कि उस का पति इस इतालवी शराब को पसंद करता है.

-- I (male) know that her husband likes this Italian wine.

मैं जानती हूँ कि उस का पति इस इतालवी शराब को पसंद करता है.

-- I (female) know that her husband likes this Italian wine.

Ita:

So che questo vino italiano piace a suo marito.

Lav:

Es zinu , ka šis itāļu vīns garšo viņas vīram.

Nor:

Jeg vet at mannen hennes liker denne italienske vinen.

Pes:

”Û‏ من می دانم که او این شراب ایتالیایی را دوست دارد شوهر

Pol:

Wiem , że jej mąż lubi to włoskie wino.

Ron:

Eu ştiu cã soţului sãu îi place acest vin italian.

Rus:

Я знаю , что её муж нравдит это итальянское вино.

Spa:

Sé que este vino italiano gusta a su marido.

Swe:

Jag vet att hennes man tycker om det här italienska vinet.

Tha:

ผม รู้ ว่า สามี ของ หล่อน ชอบ เหล้าองุ่น อิตาเลียน ขวด นี้

-- I (male) know that her husband likes this Italian wine.

ฉัน รู้ ว่า สามี ของ หล่อน ชอบ เหล้าองุ่น อิตาเลียน ขวด นี้

-- I (female) know that her husband likes this Italian wine.

From sentence to text (MOLTO Museum Grammar)

```
MkGenText GSM980019Obj AnnaLindskog OilPainting (MkColour Black) (MkSize (SIntInt 435 365))
  (MkMaterial Canvas) (MkYear (YInt 1885)) (MkMuseum GoteborgsCityMuseum)
```

- PaintingEng: The girl was painted on canvas by Anna Lindskog in 1885. It is of size 435 by 365 and it is painted in black. This oil painting is displayed at the City Museum of Gothenburg.

- PaintingFin: Maalauksen Flickan on maalannut Anna Lindskog kankaalle vuonna 1885. Se on kokoa 435 kertaa 365 ja se on maalattu mustalla. Tämä öljymaalaus on esillä Göteborgin kaupunginmuseossa.

- PaintingFre: Le tableau Flickan a été peint sur toile par Anna Lindskog en 1885. Il est de taille 435 sur 365 et il est peint en noir. Cette peinture à l' huile est exposée dans le musée municipal de Göteborg.

- PaintingIta: Il quadro Flickan è stato dipinto su tela da Anna Lindskog nel 1885. Misura 435 per 365 ed è dipinto in nero. Questo dipinto ad olio è esposto nel museo municipale di Goteburgo.

- PaintingSwe: Flickan målades på duk av Anna Lindskog år 1885. Den är av storlek 435 gånger 365 och den är målad i svart. Den här oljemålningen är utställd på Göteborgs stadsmuseum.

Using grammars for querying (MOLTO Patent Grammar)

MOLTO is funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement FP7-ICT-247914.

Natural Language Queries

zeigen sie da| Search

zeigen sie das ablaufdatum
zeigen sie das medikament
zeigen sie das zulassungsdatum

zeigen sie alle informationen über alle verabreichungsarten von FAMOTIDINE
zeigen sie alle informationen über alle darreichungsformen von GANCICLOVIR
zeigen sie das zulassungsdatum vom patent für MIACALCIN
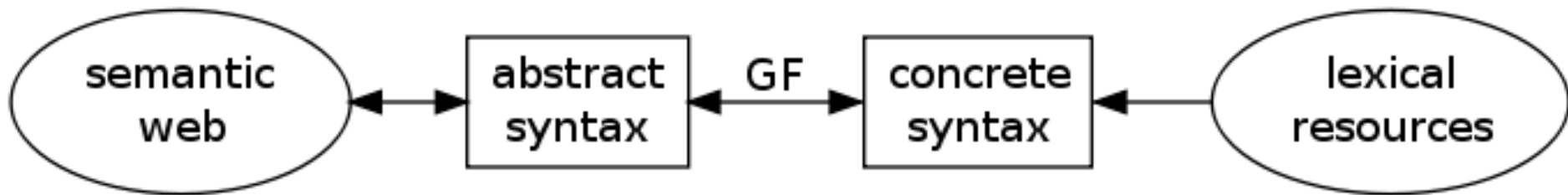
Workflow summary

Given: ontology

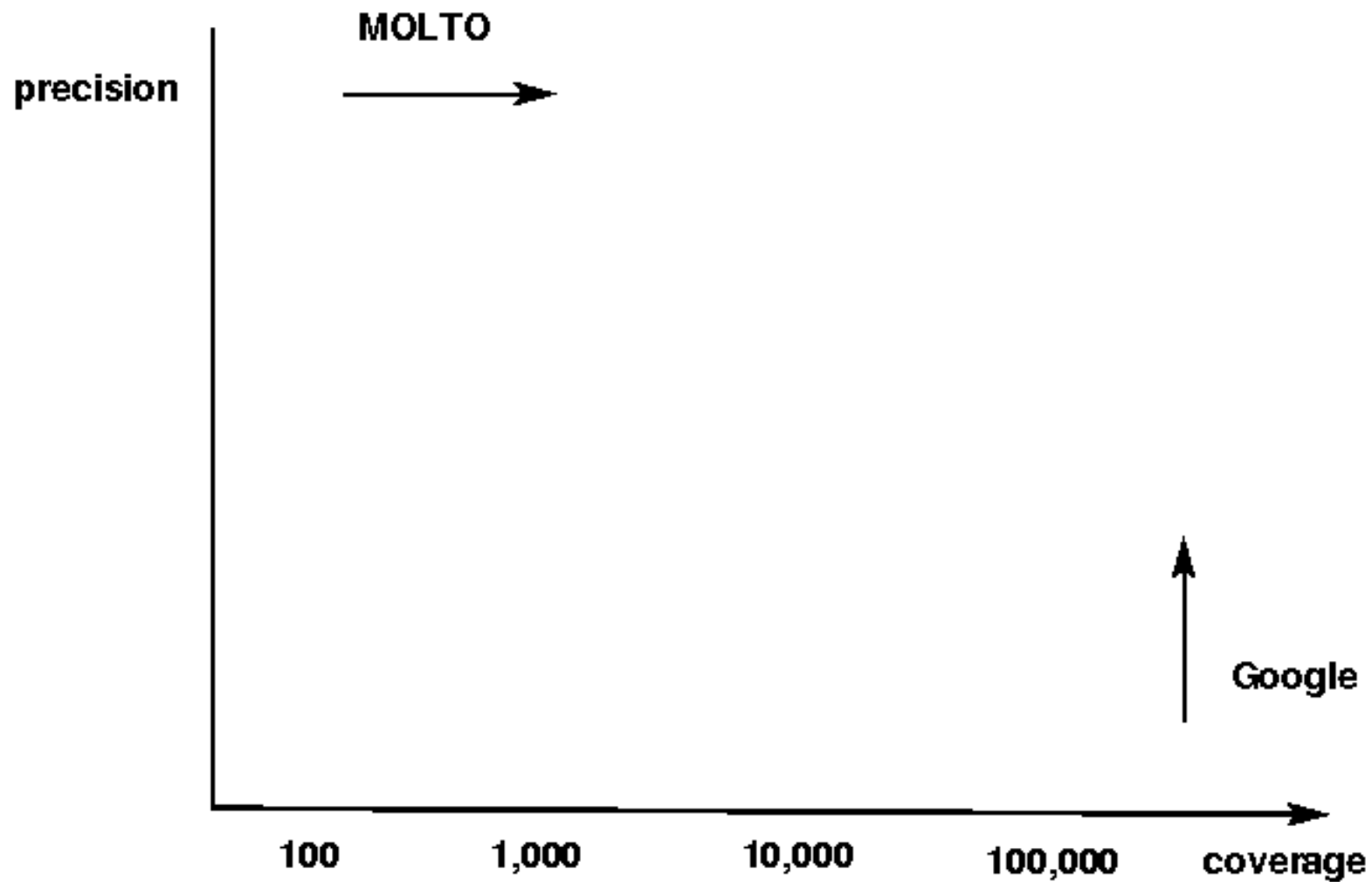**Obtain**: abstract syntax

Provide: concrete syntax

Obtain: verbalization, querying

Pages equipped with ontologies... may soon be extended to translation systems

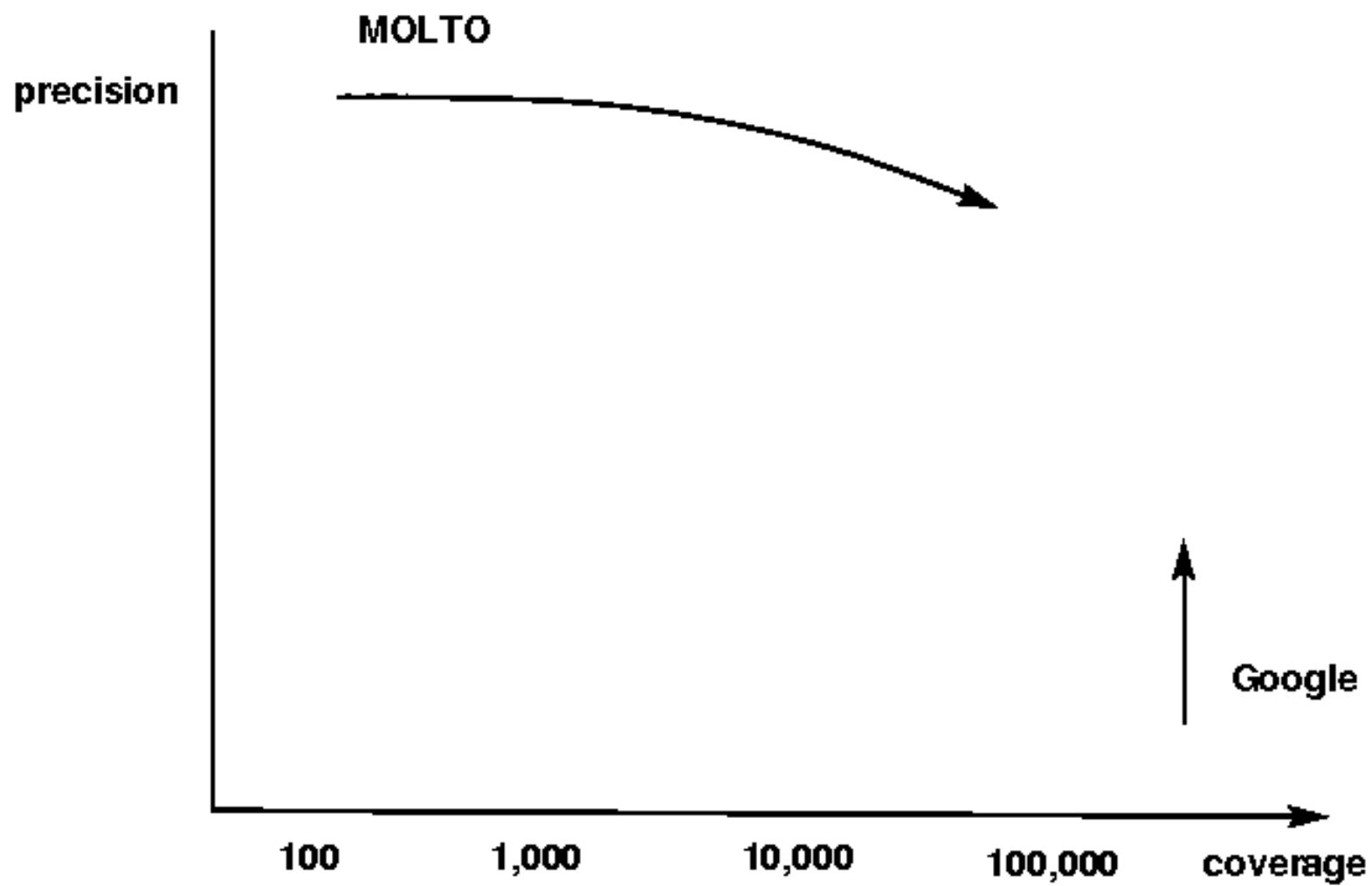# GF, Semantic Web, lexical resources

Scaling up

Learning GF grammars from examples

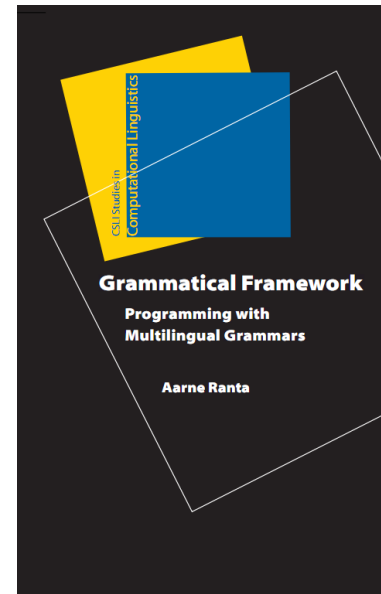| Abstract syntax | `Like She He` | first grammarian |
English example	*she likes him*	first grammarian
German translation	*er gefällt ihr*	translator/SMT syste
resource tree	`mkCl he_NP gefallen_V2 she_NP`	GF parser
concrete syntax rule	`Like x y = mkCl y gefallen_V2 x`	variables renamed

Rationale: SMT is *good* for sentences that are *short* and *frequent*

A possible scenario: controlled trade-off precision/quality

GF

Open source (GPL/LGPL/BSD)

http://www.grammaticalframework.org

Demo this afternoon; http://www.grammaticalframework.org/demos/phrasebo

My Question

Use cases for precision-oriented multilingual semantic web.