

Grammatical Framework and Machine Translation

Aarne Ranta

University of Gothenburg
and
Digital Grammars AB

FinMT Helsinki 12 September 2016



REMU

digital Grammars
Language technology to rely on.

After NMT?

After NMT?

GSM

After NMT?

GSM

= Grammar-based Semantic Methods

The GF approach

Interlingual translation based on grammars and
(some) statistics.

The GF approach

Interlingual translation based on grammars and (some) statistics.

and also:

Grammars and software available for other approaches.

Since 1998, 30+ languages, 200+ community



Nieminen, Camilleri 2013-

Wide coverage:

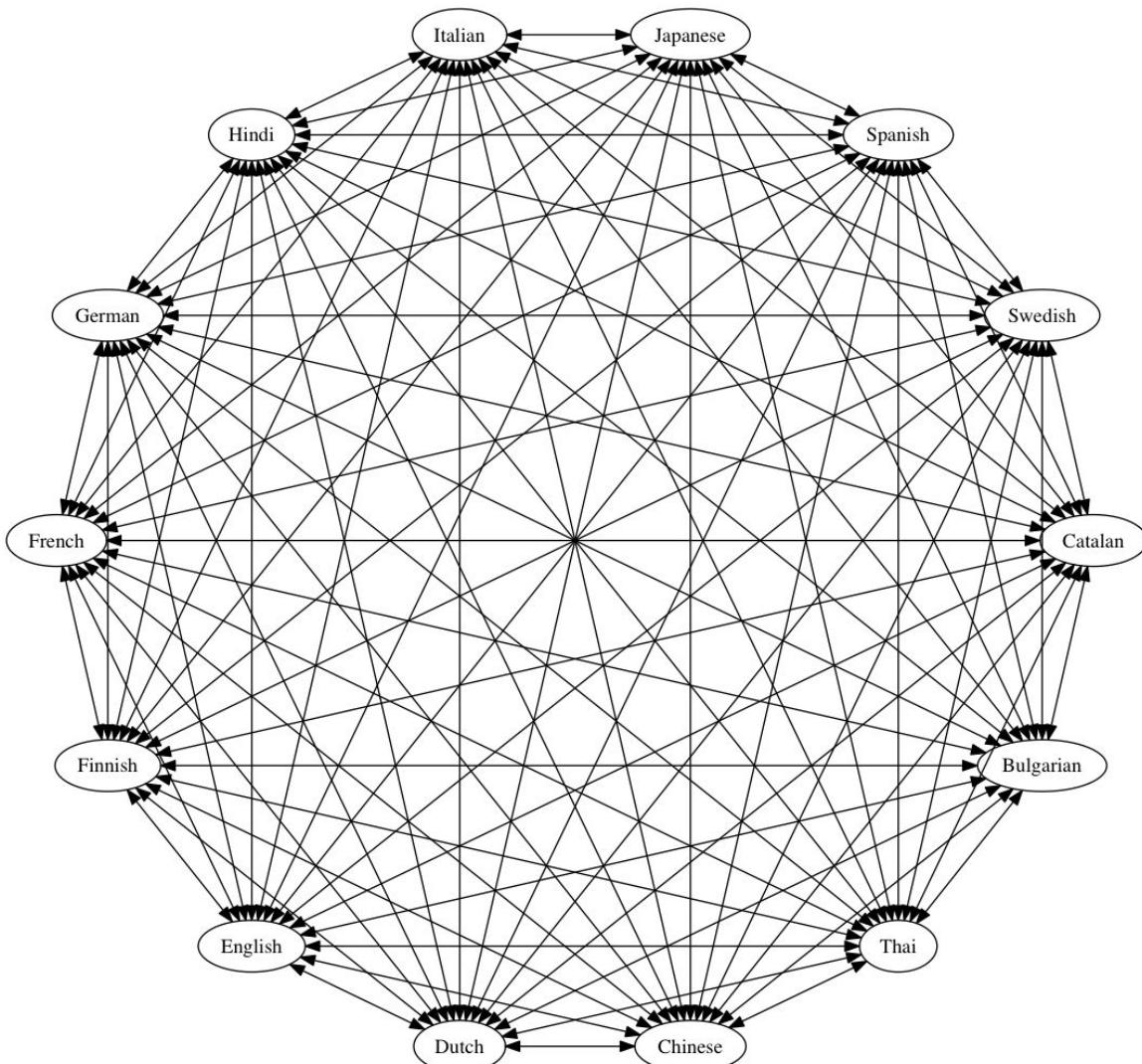
English Swedish German Dutch French Italian Spanish
Catalan Bulgarian Russian Finnish Estonian Japanese Thai
Chinese Hindi

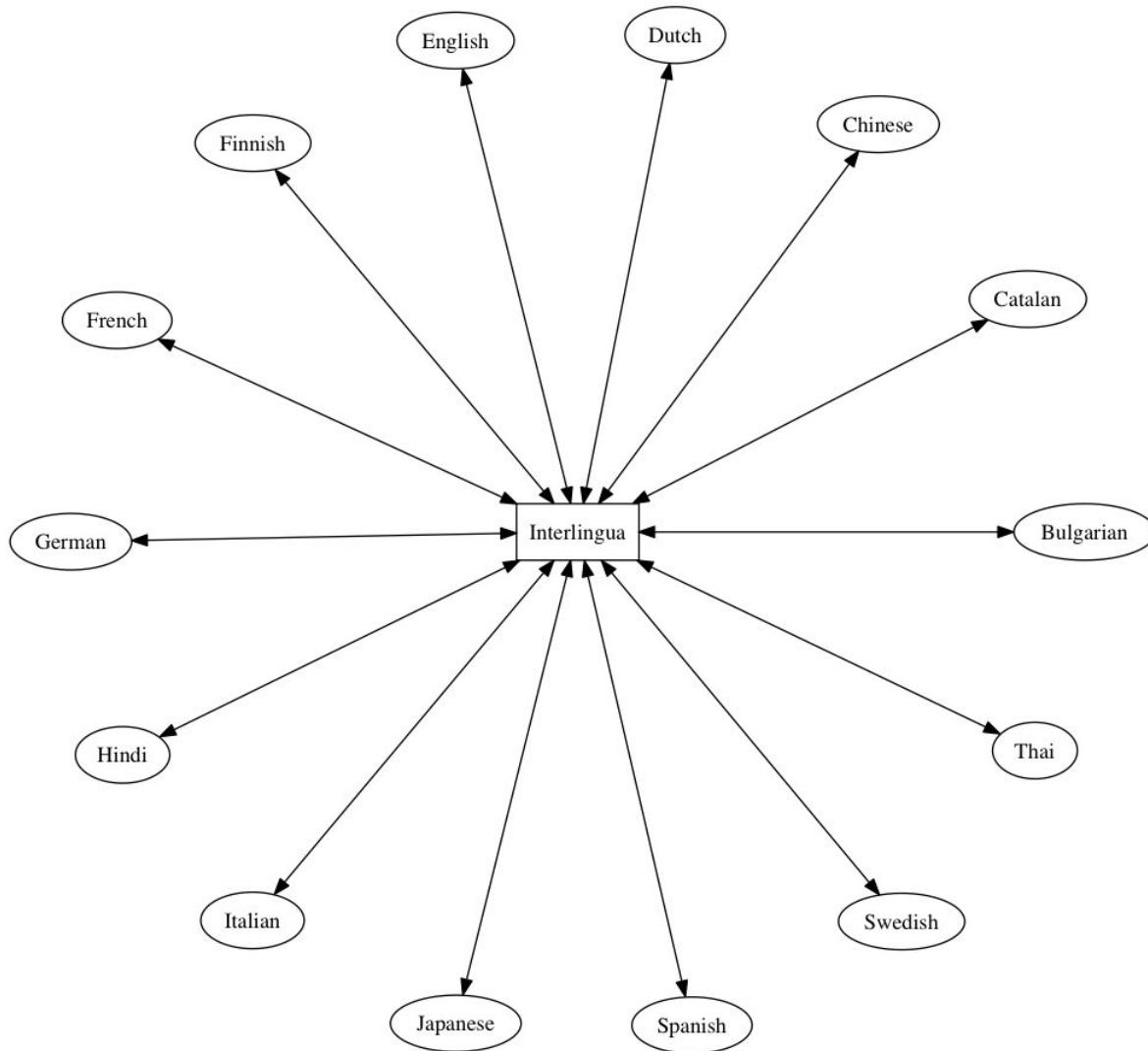
Complete Resource Grammar Library:

Norwegian Danish Afrikaans Romanian Polish Russian
Latvian Mongolian Urdu Punjabi Sindhi Greek Maltese
Nepali Persian

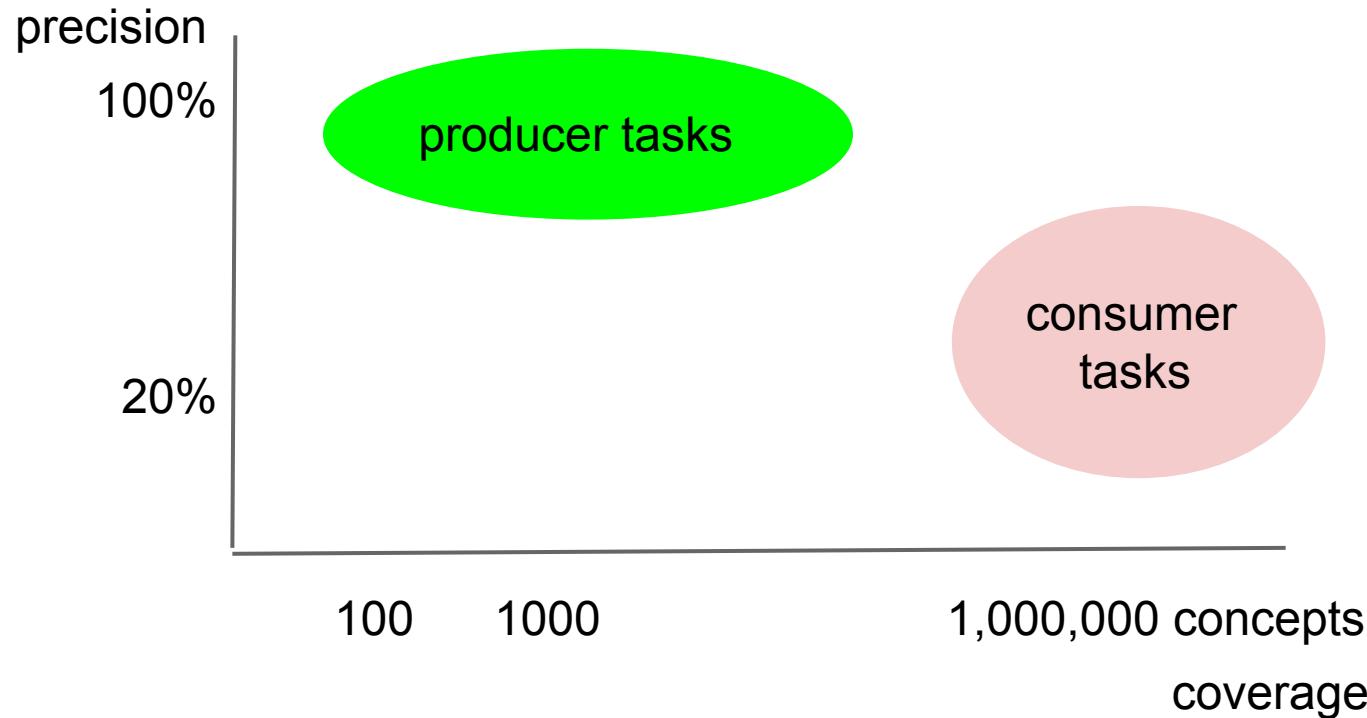
In progress:

Latin Slovene Turkish Basque Icelandic Hebrew Arabic
Amharic Swahili ...

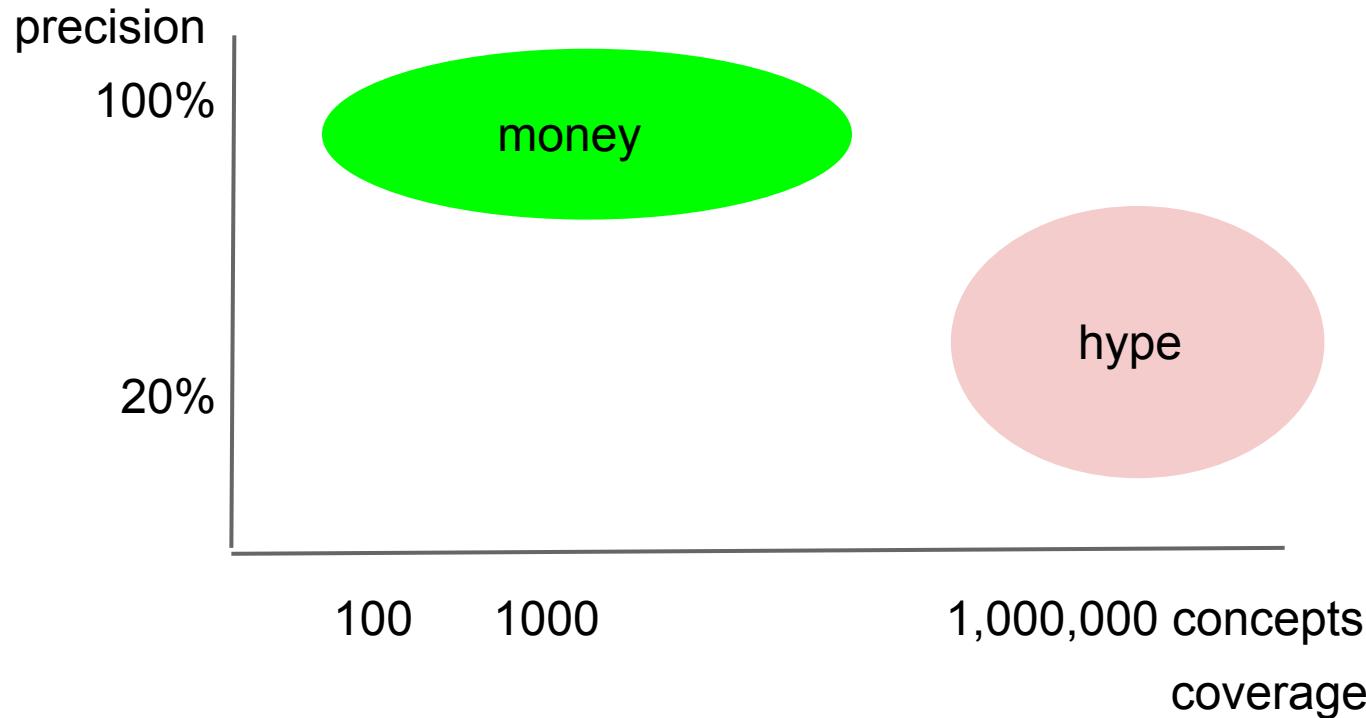




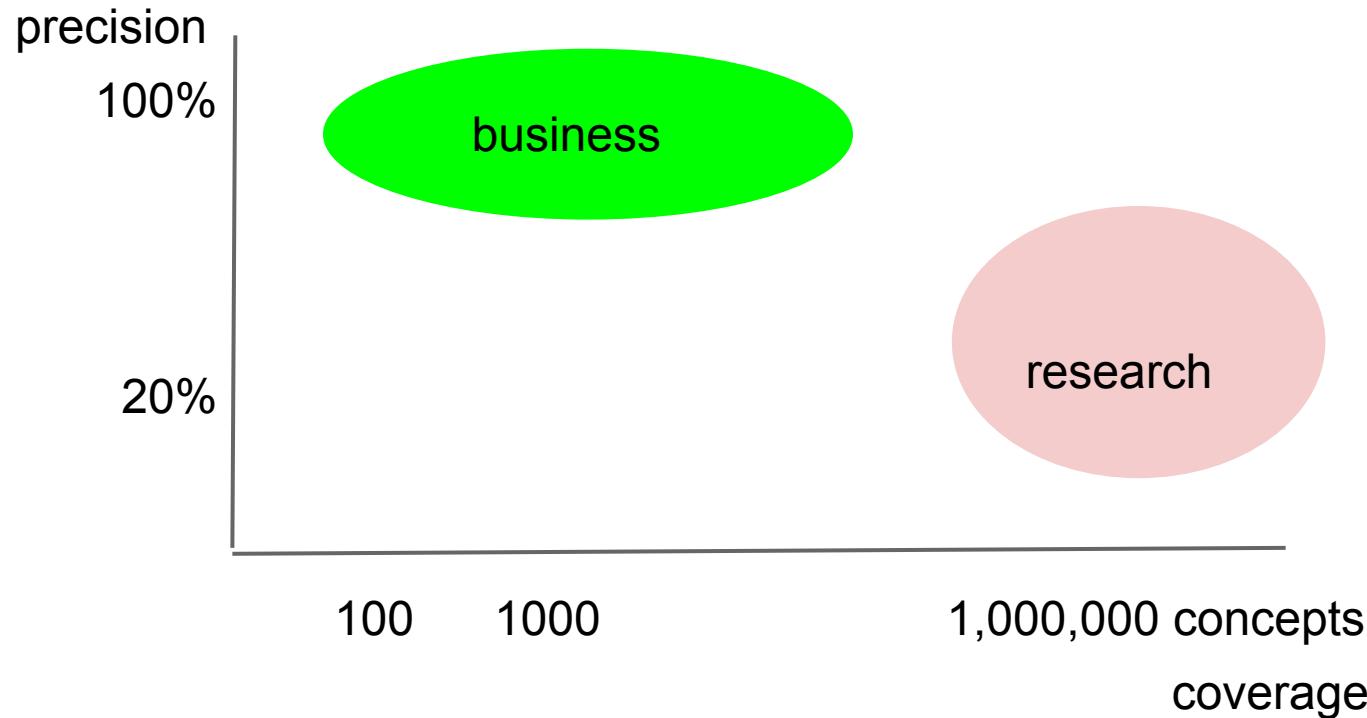
MOLTO 2010-2013



Kay (RBMT 2012)



Digital Grammars AB / REMU

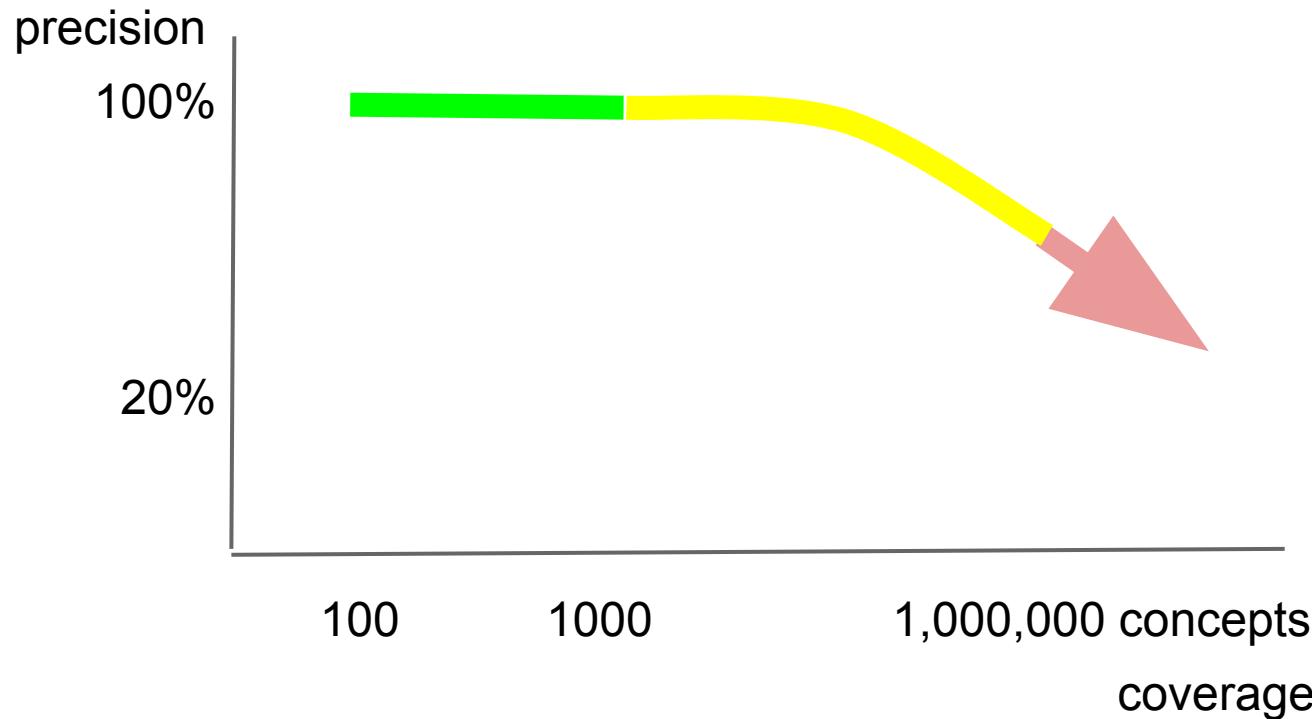




- *Svängrumsytan utanför dörren lutar 1% i sidled.*
- *The turning space outside the gate tilts 1% sideways.*
- *Kääntymätila oven ulkopuolella kallistuu 1% sivusuunnassa.*
- *Der Schwenkbereich außerhalb der Tür neigt sich um 1% seitlich.*
- *La zona de giro fuera de la puerta se inclina 1% de lado.*

<http://www.t-d.se/sv/TD2/>

Wide-coverage translation



GF Offline Translator

Finnish

Hindi

你爱我们吗

est-ce que tu nous aimes

my hovercraft is full of eels

min svävare är full av ålar

questo programma traduce

тази програма превежда

kaupassa on olutta

<https://play.google.com/store/apps/details?id=org.grammaticalframework.ui.android>

<https://itunes.apple.com/us/app/gf-offline-translator/id1023328422?mt=8>

K. Angelov, B. Bringert & A. Ranta,
Speech-enabled hybrid multilingual
translation for mobile devices,
EACL 2014.

Carrier

9:57 AM

Finnish Thai

Info

I don't speak Swedish

jag talar inte svenska

but my phone can translate for me

aber mein Fernsprecher kann für mich
übersetzen

the best translations are green

le migliori traduzioni sono verdi

red translation maybe not grammatical

la traducción roja quizás no
gramatical

you can translate from all the fourteen
languages

あなたは全部十四個の言葉の壁から訳
すことができます

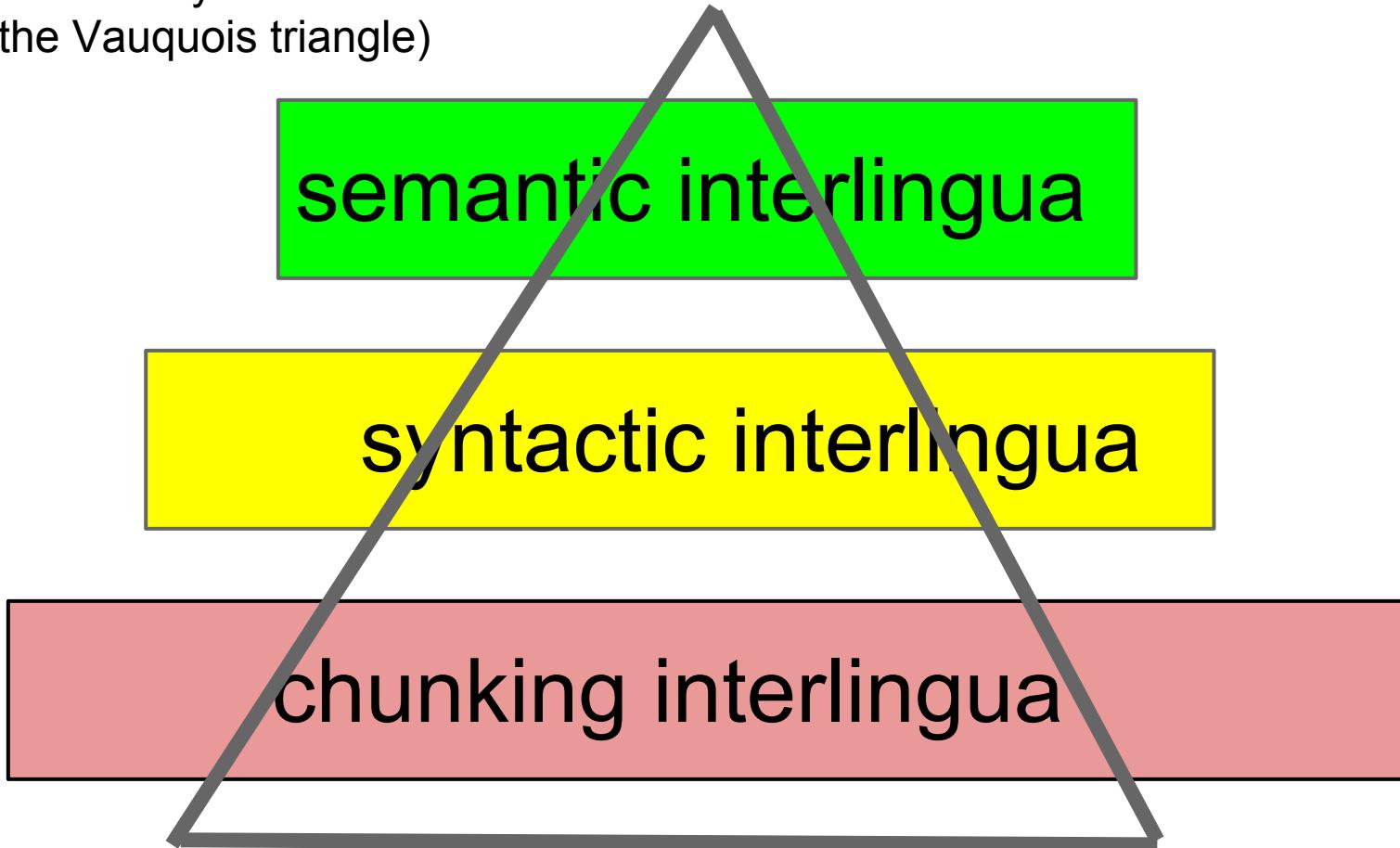
je t'aime

我爱你

the best translations are green

Translate

Translation layers
(cf. the Vauquois triangle)





Wide Coverage Translation Demo

English

Clear



Finnish



Colors

Translate

Grammars...

My daughter is hungry.

The vice dean kicked the bucket.

We copes from grammar errors by this way.

Tyttärelläni on nälkä..

Pahedekaani potkaisi sankoa.

Me kestää tämän tien kautta kielioppivirheistä.

Try Google Translate



Wide Coverage Translation Demo

English

Clear



Finnish

 Colors

Translate

Grammars...

My daughter is hungry.

The vice dean kicked the bucket.

We copes from grammar errors by this way.

Tyttärelläni on nälkä..

Pahedekaani potkaisi sankoa.

Me kestää tämän tien kautta kielioppivirheistä.

Try Google Translate

My daughter is hungry.

The vice dean kicked the bucket.

We copes from grammar errors by this way.

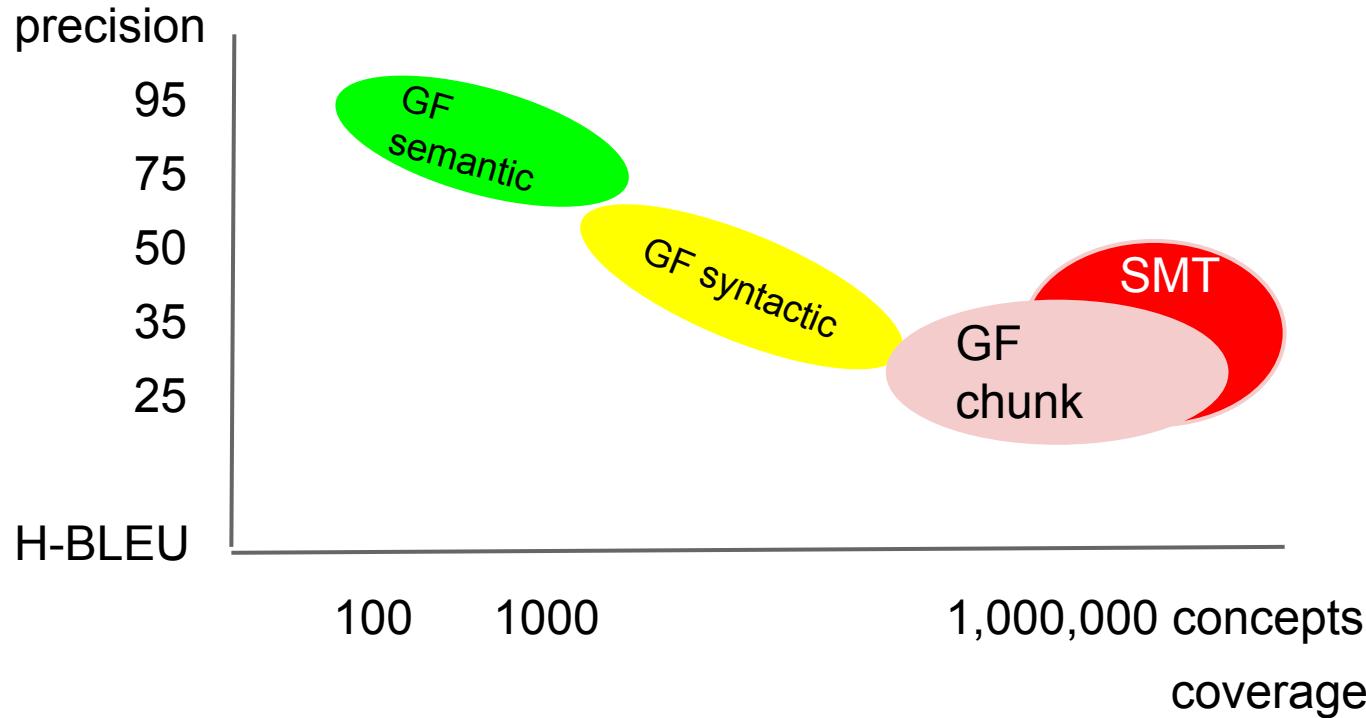


Tyttäreni on nälkäinen.

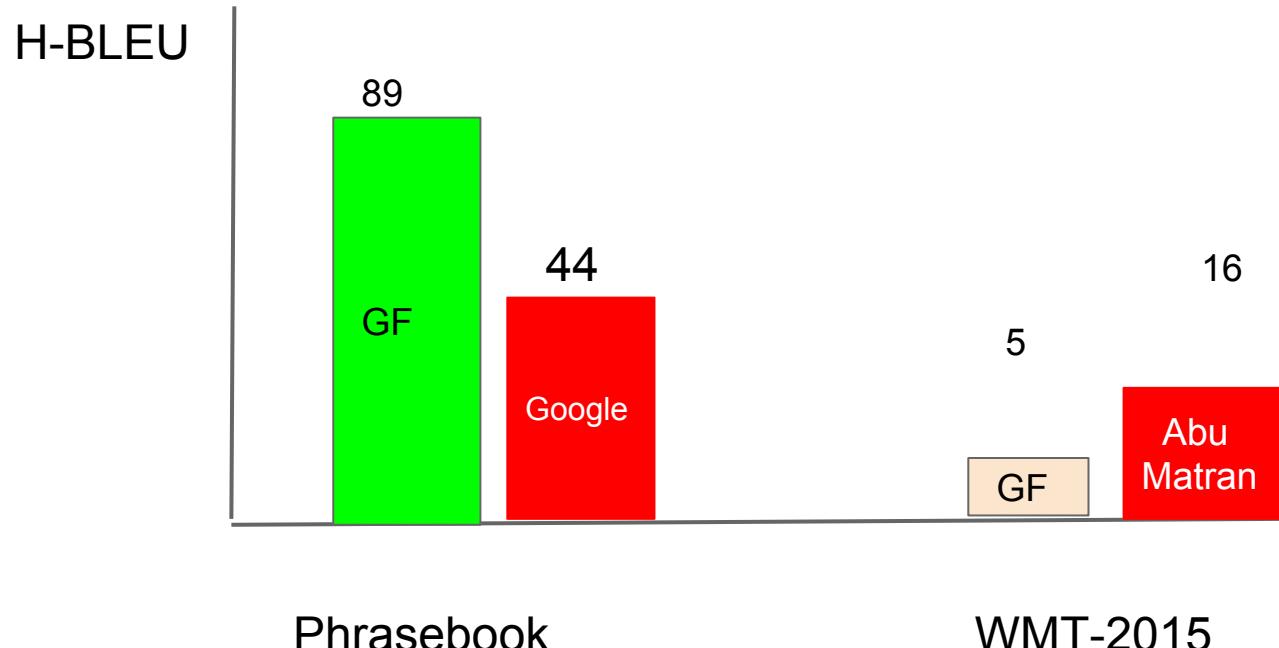
Varapuheenjohtaja dekaani potkaisi ämpäri.

Me selviää mistä kielioppivirheitä tällä tavalla.

Where we are now



Some scores (English-Finnish)



Phrasebook

Rautio and Koponen, MOLTO Deiverable 9.2, 2013.

WMT-2015

Kolachina and Ranta, WMT 2015.

Swe TD to Fin, Ger, Spa

	GF, BLEU	Google, BLEU
Fin, CNL	77	31
Fin, robust	31	20
Fin, all	73	28
Ger, CNL	75	37
Ger, robust	33	34
Ger, all	73	37
Spa, CNL	76	28
Spa, robust	39	25
Spa, all	74	38

language	GE	PE	PES	Ws/h	LoC
Finnish	14	6	1200	370	1292
German	18	6	1200	300	1290
Spanish	38	8	900	160	1291
Swedish	12	-	-	-	1303
English	12	-	-	-	1284
abstract	12	-	-	-	1286
prepare	12	-	-	-	400
total	118	20	1100	160	8056

Table 2: The grammar writing and translation effort. GE = grammar engineering (hours), PE = postediting (hours), PES Ws (post-editing speed words per hour), Ws/h (words translated per hour of work), LoC = lines of GF/Haskell code. The “total” for PES and Ws/h means the average of all languages.

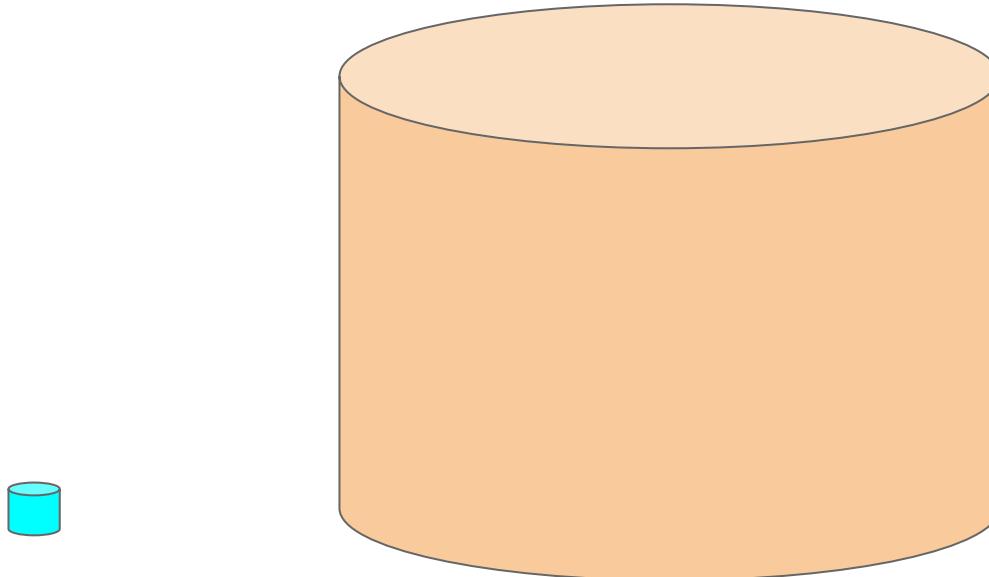
GF vs. Statistical Translation

- + grammatical correctness
- + feedback: trees, colours
- + less dependent on language data
- + compact size of multilingual systems

Compactness

GF Offline Translator: 16 modules, 35 MB in total

Google/Baidu translate offline: 210 modules, 30-150 MB each



GF vs. Statistical Translation

- + grammatical correctness
- + feedback: trees, colours
- + less dependent on language data
- + compact size of multilingual systems
- non-compositional idioms
- contextual disambiguation
 - cases for hybrid methods

Grammars...

PhD in linguistics

Grammars...

PhD in linguistics

5 years of work

Grammars...

PhD in linguistics

5 years of work

brittle

Grammars...

PhD in linguistics

5 years of work

brittle

obsolete

Grammars...

statistics

PhD in linguistics

BSc in computer science

5 years of work

5 weeks of waiting

brittle

robust

obsolete

state of the art

Grammar formalisms

Tradition (1970-1980's)

Linguistic theory

Definition of grammaticality

Monolith

Experts

Monolingual

GF (1998-)

Compiler construction

Transduction

Framework

Distributed work

Multilingual

Grammar building effort

First grammar on cloud IDE

minutes

Grammar building effort

First grammar on cloud IDE minutes

Commercial application grammar days

Grammar building effort

First grammar on cloud IDE minutes

Commercial application grammar days

Resource grammar months

Grammar building effort

First grammar on cloud IDE minutes

Commercial application grammar days

Resource grammar months

Wide coverage grammar

Grammar building effort

First grammar on cloud IDE minutes

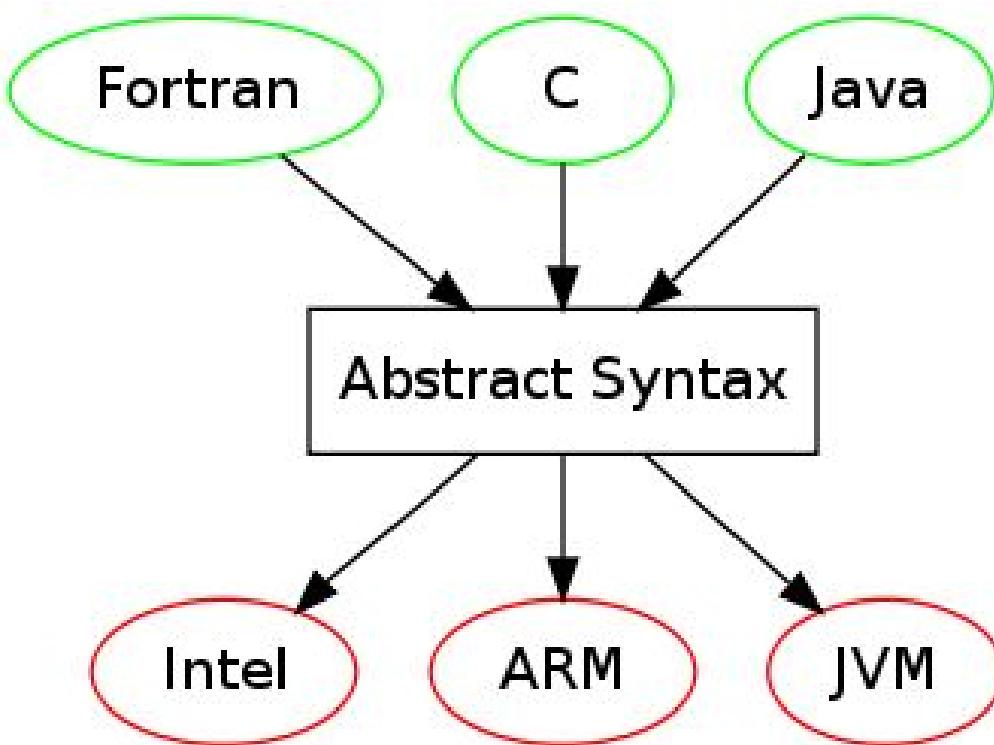
Commercial application grammar days

Resource grammar months

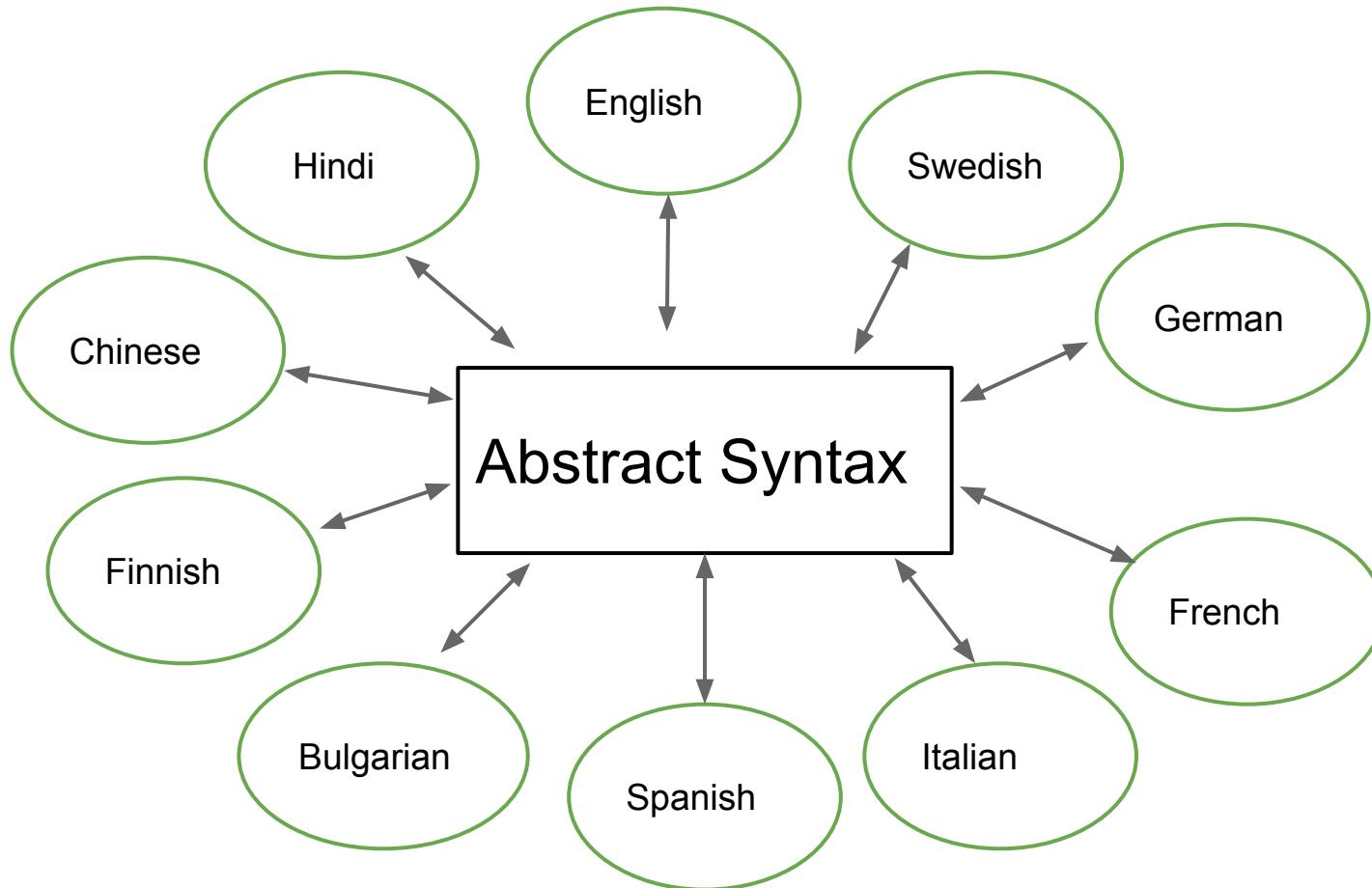
Wide coverage grammar days

How it works

Model: multi-source multi-target compiler



“Compiling natural language”



Example: subject-verb-object predication

abstract syntax functions

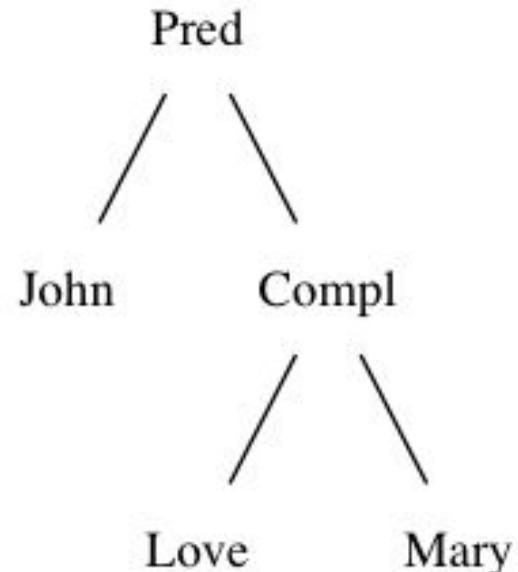
```
fun Pred : NP -> VP -> S
```

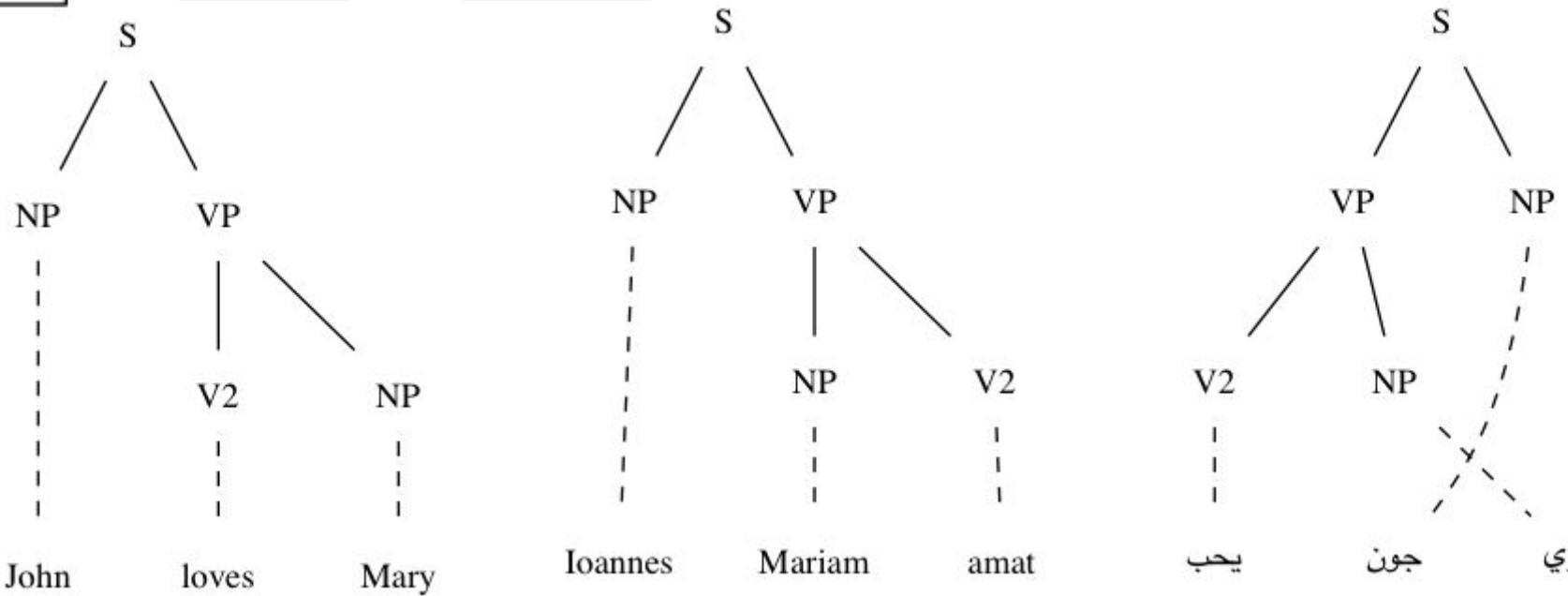
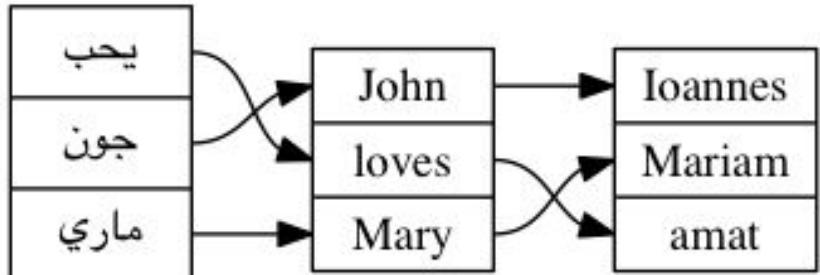
```
fun Compl : V2 -> NP -> VP
```

```
fun John : NP
```

```
fun Mary : NP
```

```
fun Love : V2
```





Concrete syntax linearizations

SV0: lin Pred s vo = s ++ vo

lin Compl v o = v ++ o

SOV: lin Pred s vo = s ++ vo

lin Compl v o = o ++ v

VSO: lin Pred s vo = vo.verb ++ s ++ vo.obj

lin Compl v o = {verb = v ; obj = o}

Inflection tables

```
lin Love = table {  
    Ind Pres Sg P1 => “rakastan” ;  
    Ind Pres Sg P2 => “rakastat” ;  
    Ind Pres Sg P3 => “rakastaa” ;  
    ... -- up to 240 forms  
}
```

```
lin Love = “愛” ;
```

Multilingual abstraction

Abstract syntax:

- tree structure: constituency, word senses

Concrete syntax:

- words, word order
- inflection, agreement

Expressive power

Synchronous CFG (context-free grammars)

- operating on strings



Synchronous MCFG (*Multiple* CFG)

- operating on tuples (tables, records)

The UD connection

GF

Common abstract syntax

30+ languages

Parsing and generation

Mechanic mapping to UD

Work on grammar writing

Universal Dependencies

Shared POS tags and labels

30+ languages

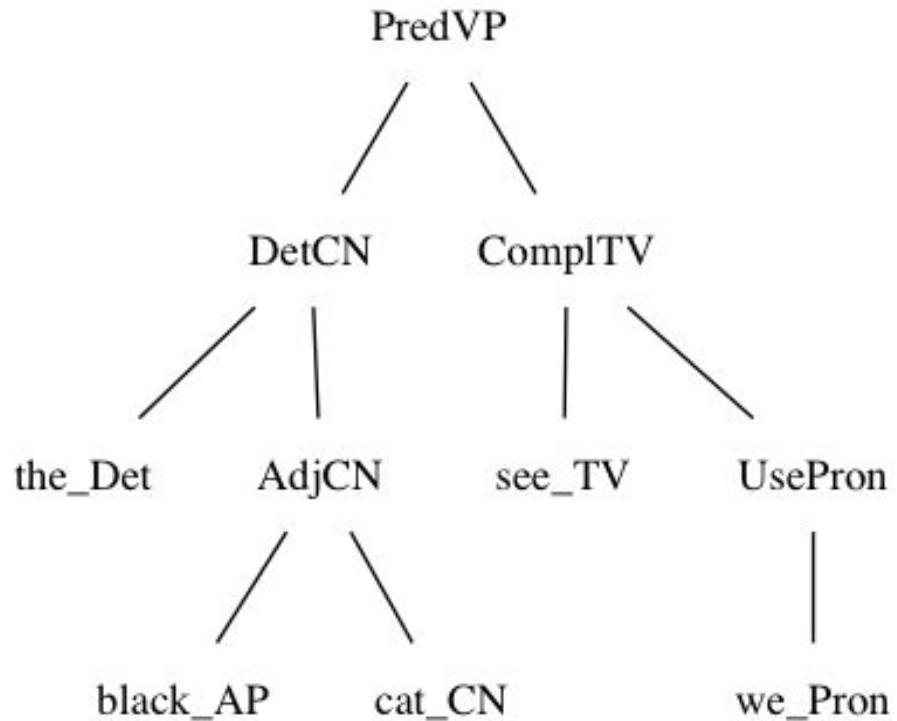
Parsing only

Partial retrieval of GF

Work on tree annotation

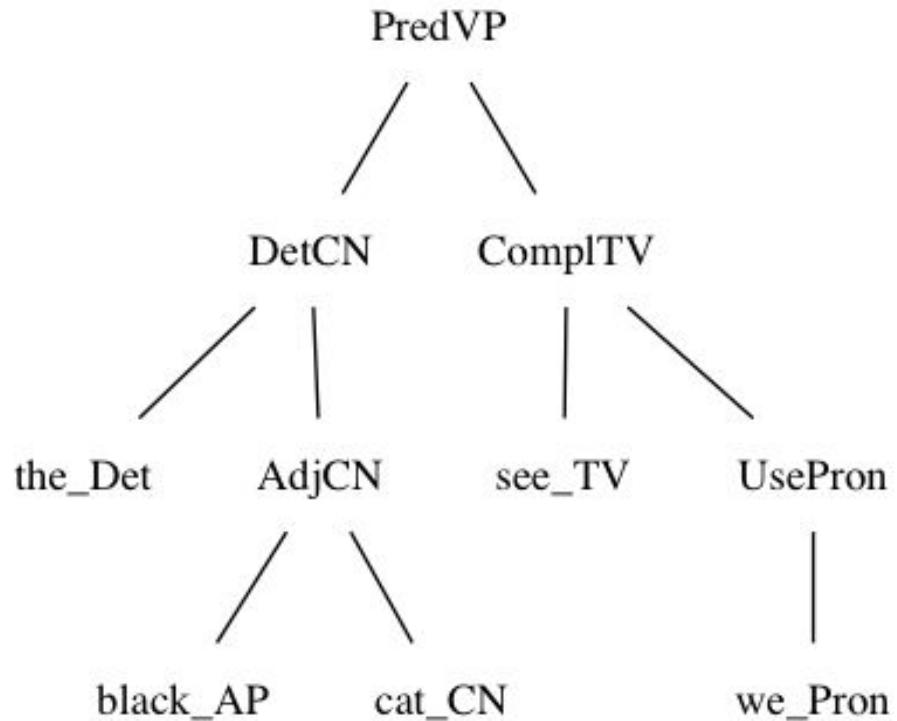
From AST to dependency tree

- put labels on subtrees and hide functions

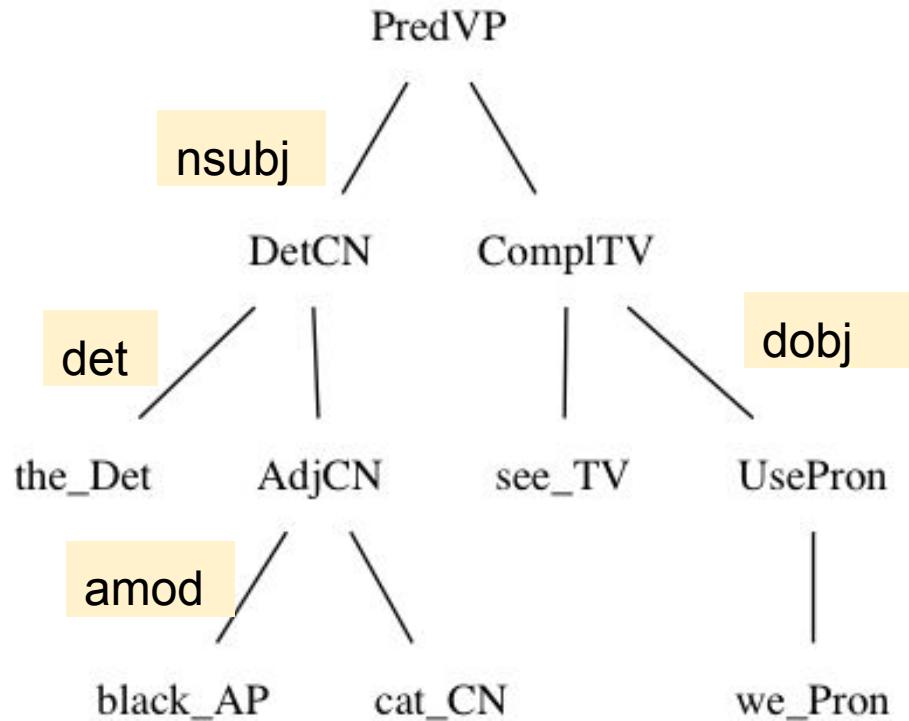


Dependency configuration

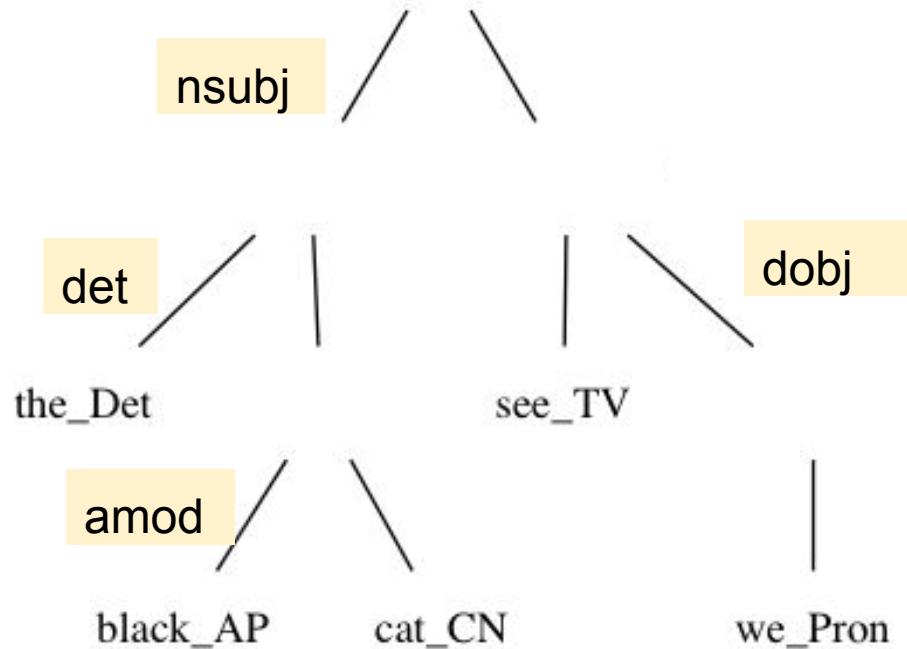
PredVP	nsubj	head
ComplTV	head	dobj
DetCN	det	head
AdjCN	amod	head

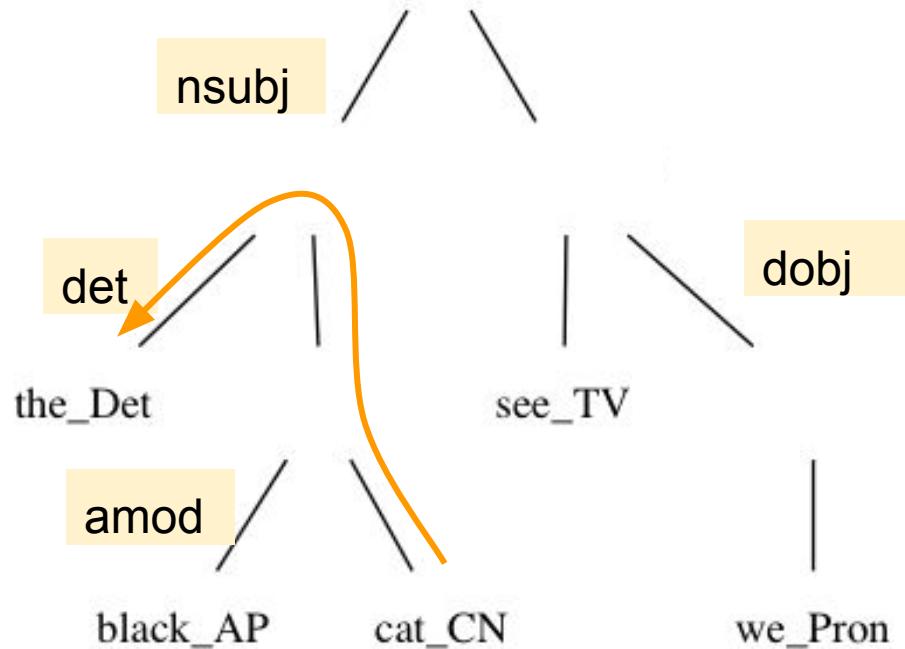


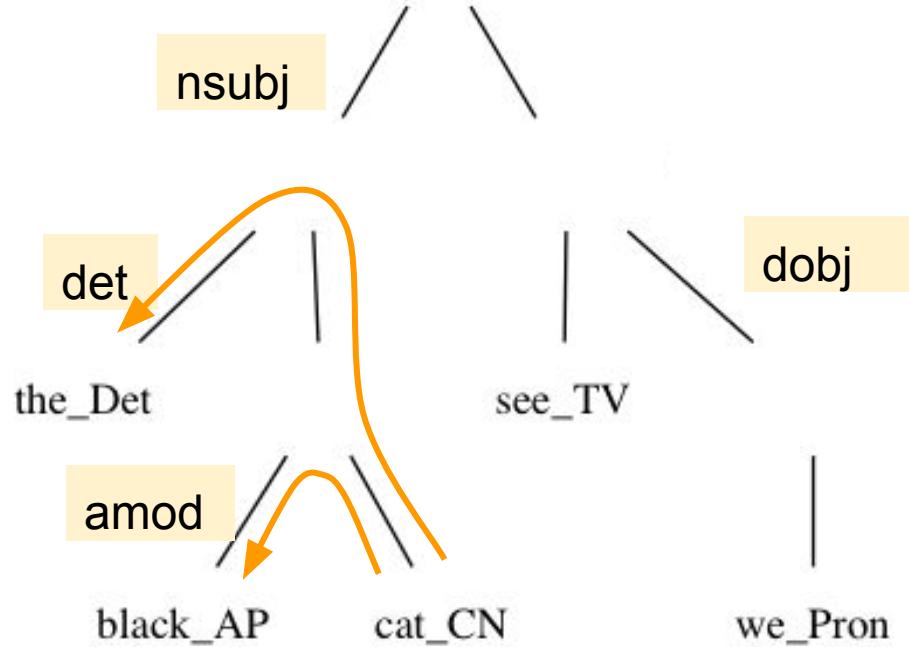
Dependency configuration

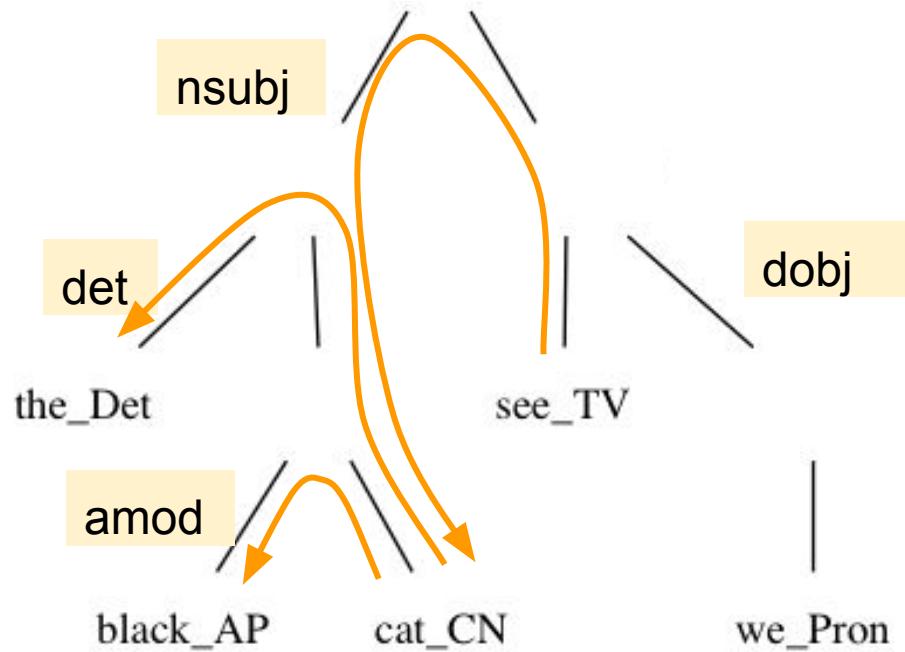


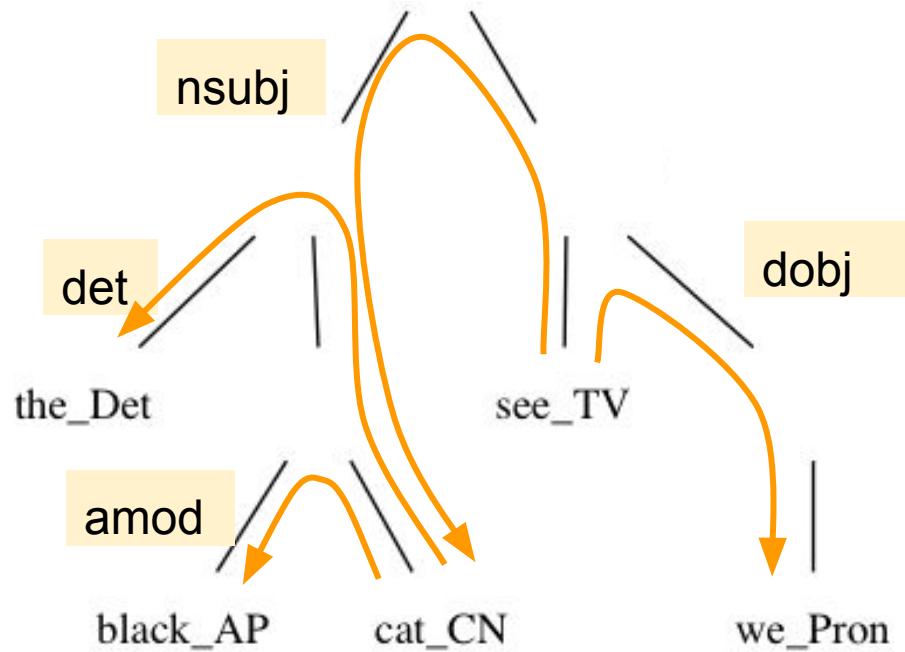
PredVP	nsubj	head
ComplTV	head	dobj
DetCN	det	head
AdjCN	amod	head

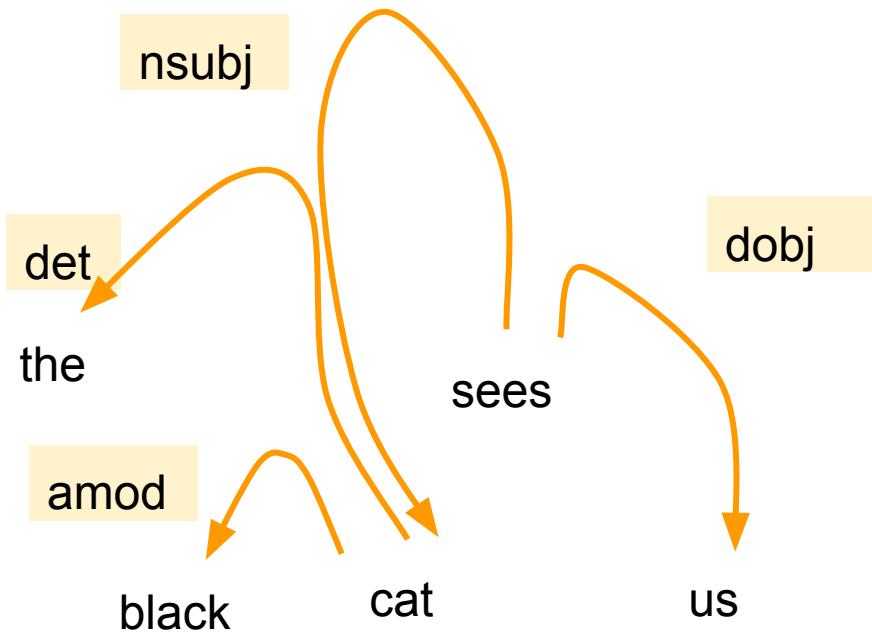






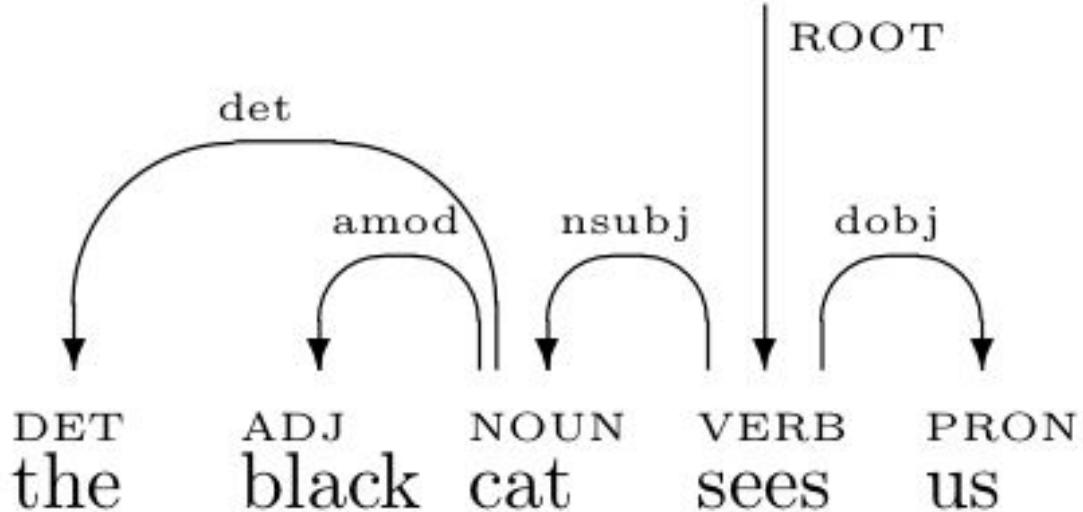


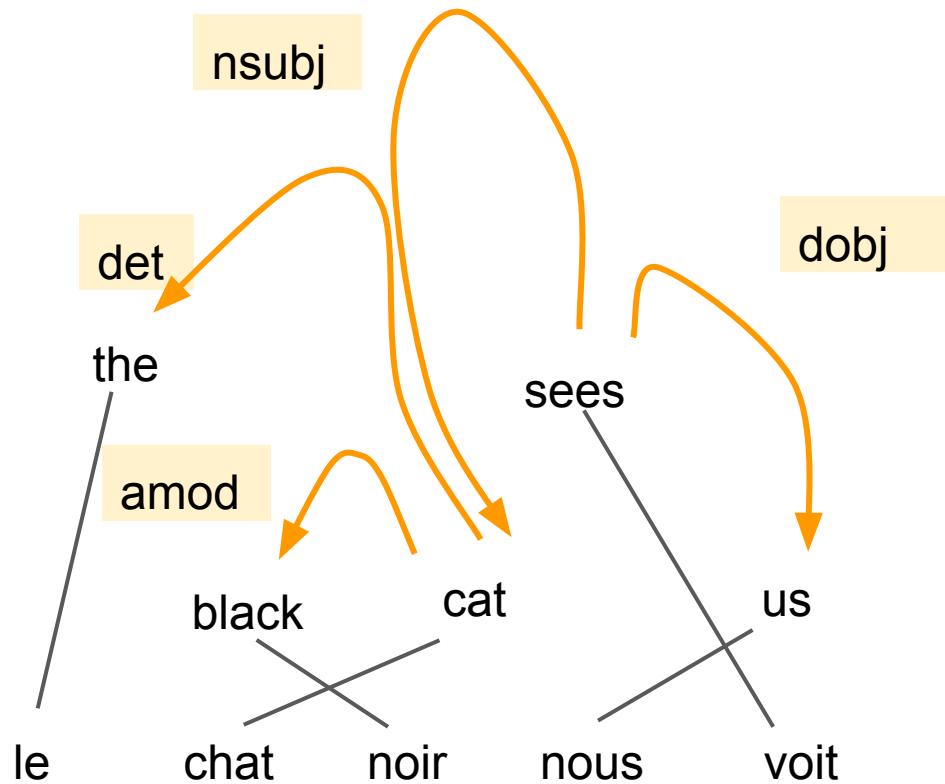


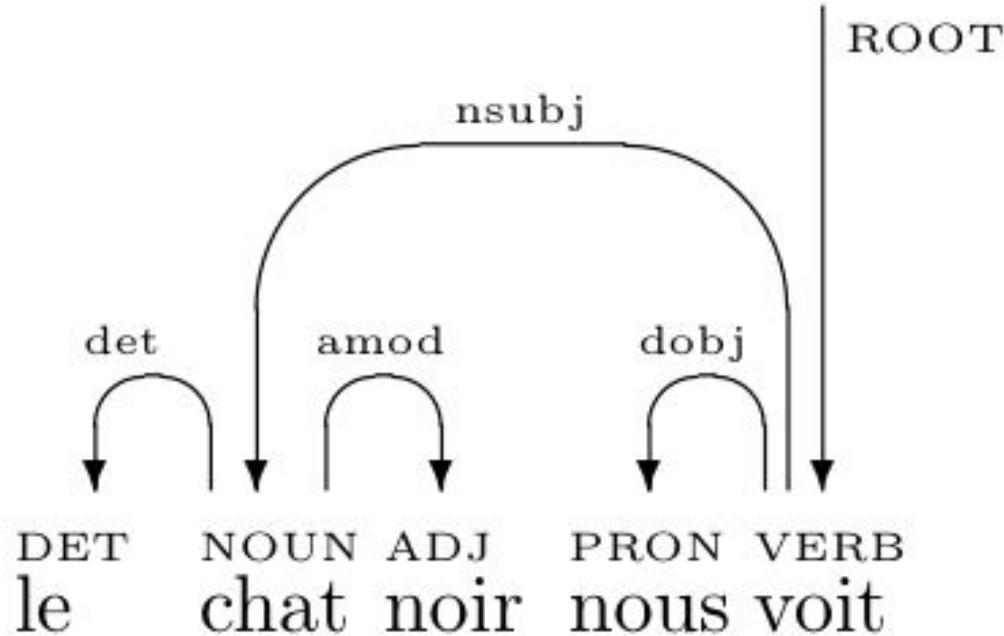


POS configuration

Det	DET
AP	ADJ
CN	NOUN
TV	VERB
Pron	PRON







GF to UD:

- Translation of treebanks
- Rule-based dependency parsing
- Principled approach to annotations

UD to GF:

- Improved statistical models
- Fast robust parsing

Take home points

GF

- Grammar engineering without tears
- Interlingual translation
- Domain adaptation
- Abstract Syntax ~ Universal Dependencies
- Libraries for over 30 languages, growing

<http://www.grammaticalframework.org>



Wide Coverage Translation Demo

English

Clear



Finnish

Colors

Translate

Grammars...

Around midnight, the father of the boy informed Ilta-Sanomat that the boy has been found.

Stealing food is a common crime in student halls.

Keskiyön ympärillä, pojан isä informoi Ilta-Sanomat:ia että poika on löydetty.

Ruoan varastaminen on tavallinen rikos opiskelijalämpiöissä.

Enter text to translate above

Try Google Translate

1 2 3 4 5 6 7 8 9 10

en-US Stealing food is a common crime in student halls → Ruoan varastaminen on tavallinen rikos opiskelijalämpiöissä fi-FI

80.74987

PhrUtt NoPConj (UttS (UseCl (TTAnt TPRes ASimul) PPos (PredVP (GerundNP (ComplV2 **steal_V2** (MassNP (UseN **food_N**)))) (AdvVP (UseComp (CompNP (DetCN (DetQuant IndefArt NumSg) (AdjCN (PositA **common_1_A**) (UseN **crime_N**)))))) (PrepNP **in_Prep** (DetCN (DetQuant IndefArt NumPl) (UseN (CompoundN **studentFem_N hall_N**))))))) NoVoc

KITOS