

# Grammatical Framework

Aarne Ranta

Department of Computer Science and Engineering,  
Chalmers University of Technology and University of Gothenburg

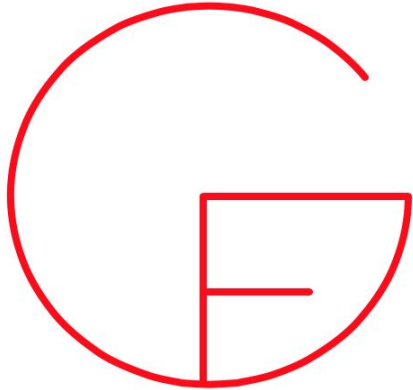
Digital Grammars AB

FOSS North, Gothenburg, 26 May 2016



REMU

digital  grammars  
Language technology to rely on.



*The mission of GF is to formalize the grammars of the world and to make them available for computer applications.*

<http://www.grammaticalframework.org>

Since 1998

# Available?

Back in 1998, grammar software was mostly proprietary

- Xerox XFST and XLE
- SRI Core Language Engines

But Xerox was kind enough to let GF join the bazaar.

# Computer applications

Translation

Localization

Query systems

Voice communication

Language learning

Text analysis

# Why grammars?

Let us start with a translation demo:

The Wide-Coverage Translator on the GF Cloud

<http://cloud.grammaticalframework.org/wc.html>

(Also available on as a mobile app: the GF Offline Translator for Android and iOS)

# Wide Coverage Translation Demo

English

I don't speak Swedish.  
But my computer can translate for me!  
The translation uses grammars for 15 languages.  
The program uses colours to indicate the expected quality.  
Green translations are best.  
Yellow translations are not quite so good.  
Translations which is red is usually very bad but can serve as back-up.  
Google translate is probably better on the average.  
But our translation supports optimization for specific domains.

Enter text to translate above

Italian

Non parlo svedese.  
Ma il mio computer sa tradurre per me!  
La traduzione usa grammatiche per 15 lingue.  
Il programma usa televisore a colori per [indicate\_V2] [indicate\_V2] [indicate\_V2] la qualità [indicate\_V2]  
Traduzioni verdi sono [best\_Adv].  
Traduzioni gialle non sono [quite\_AdA] così buone.  
Traduzioni quali è il rosso [ConjVPS].  
[google] traduci [copula\_Chunk] probabilmente il [better\_N] carponi lui la media.  
Ma la nostra traduzione supporta dell'ottimizzazione per domini specifici.

Swedish

Jag talar inte svenska.  
Men min dator kan översätta för mig!  
Översättningen använder grammatiker för 15 språk.  
Programmet använder färger för att indikera den förväntade kvaliteten.  
Gröna översättningar är bäst.  
Gula översättningar är inte ganska så goda.  
Översättningar vilka är röda är vanligen mycket dålig men kan tjäna som reserv.  
[google] översatt är troligt bättre hälft på det medelta.  
Men vår översättning stödjer optimering för specifika domäner.

Chinese

我不说瑞典语。  
但是我的计算机在为了我会翻译!  
翻译在为了15种语言以语法。  
程序在[InOrderToVP]以色盲。  
绿翻译在最好。  
黄翻译不相当如此好。  
翻译红是哪通常非常糟糕但是在与备份一样能服务。  
[google]翻译是可能地长辈上个平均。  
但是我们的翻译在为了个别的域承受优化。

# Main stream machine translation

Google, Bing, Baidu

Statistics and machine learning

Automatically constructed from data

- Parallel text: millions of words
- Single language text: billions of words

**So, why grammars?**



# Advantage 1: grammaticality

English Swedish Hebrew Detect language ▾



Finnish Swedish English ▾

Translate

My car is red.  
My house is red.  
My old house is not red.



Min bil är röd.  
Mitt hus är röd.  
Min gamla hus är inte röd.

# Advantage 2: rule-based coverage

English Swedish Finnish Detect language ▾



Finnish Swedish English ▾


Translate



olut, olutta, oluen, oluena, olueksi,  
oluessa,oluesta, olueeseen, oluella, oluelta,...  
Sinä haiset oluelta!  
Oluesta siirryttiin viineihin.




beer, beer, beer, beers, beer, beer, beer, beer,  
beer, oluelta, ...  
You stink oluelta!  
Beer moved wines.

# Advantage 3: reliability




English Swedish Finnish Detect language 


 Dutch Chinese (Simplified) English  [Translate](#)

Min mor är svensk.  
Min mor är inte svensk. 




我的母亲是瑞典的。  
我的母亲是瑞典的。


# Advantage 3: reliability

English Swedish Finnish Detect language   Dutch Chinese (Simplified) English  [Translate](#)


Min mor är svensk.   
Min mor är inte svensk.

我的母亲是瑞典的。  
我的母亲是瑞典的。

English Swedish Finnish Detect language   Dutch Arabic English  [Translate](#)

Min mor är svensk.   
Min mor är inte svensk.

My mother is Swedish.  
My mother is Swedish.

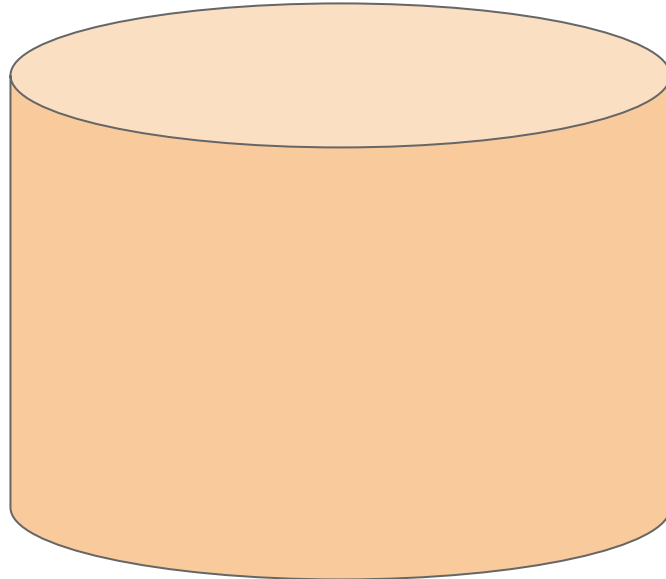
Min far är svensk.   
Min far är inte svensk.

My father is Swedish.  
My father is not Swedish.

# Advantage 4: compactness

GF Offline Translator: 15 modules, 30 MB in total

Google translate offline: 182 modules, 150 MB each

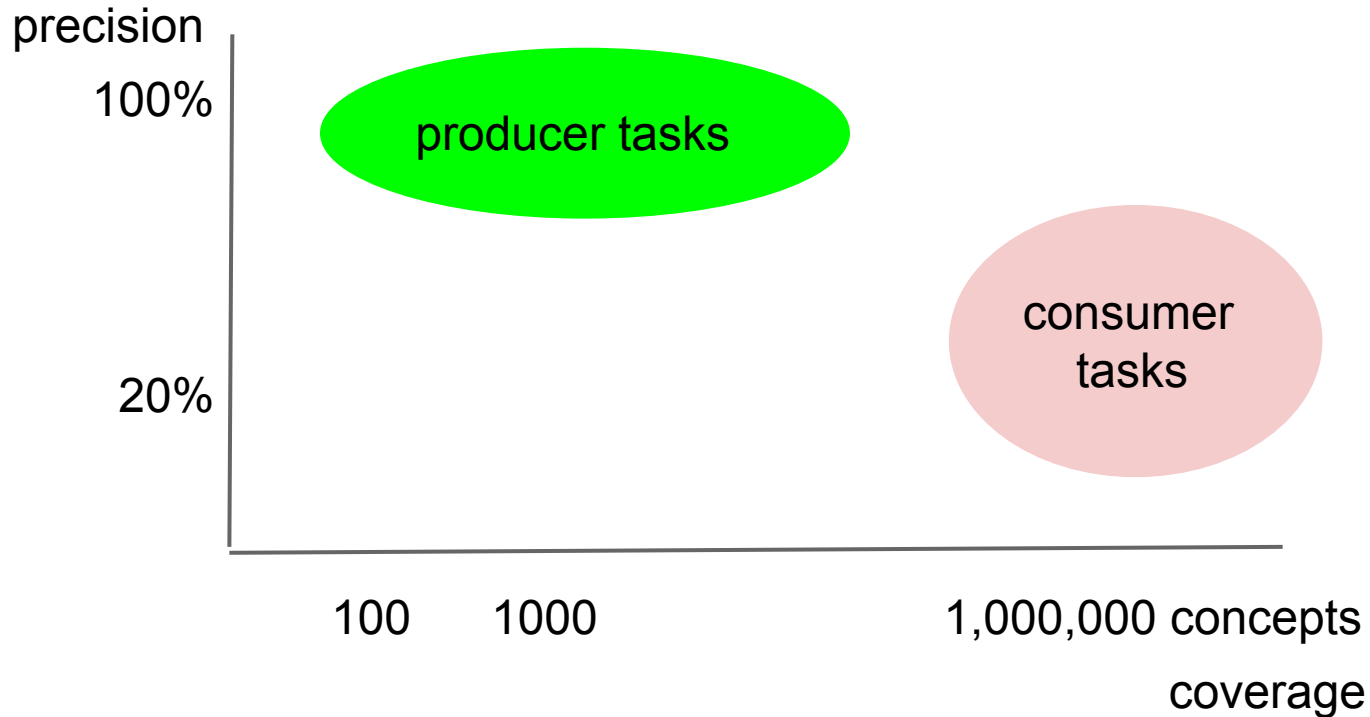


# Advantage 5: programmability

Grammars are programs that can be

- inspected
- understood
- debugged
- adapted

# Advantage 6: domain adaptability



**Any disadvantages?**



# **Disadvantage 1: hard to build**

“A PhD in linguistics and years of work”

Cf. statistics: a BSc in computer science and some weeks waiting for training to complete.

# Disadvantage 2: limited coverage

“Syntax error”

“All grammars leak”

Cf. statistics: billions of words to recombine.

**Making grammars easier to build:**

**grammar engineering tools**

# The GF programming language

High level source: functional programming

Resource Grammar Libraries

Compiler and IDE tools

Binary runtime: PGF (Portable Grammar Format)

## ParadigmsSwe.mkV “glömmer”

	<b>active</b>	<b>passive</b>
<b>infitiv</b>	glömma	glömmas
<b>presens</b>	glömmer	glöms
<b>preteritum</b>	glömde	glömdes
<b>supinum</b>	glömt	glömts
<b>imperativ</b>	glöm	glöms

### Particip Presens

		<b>obest</b>	<b>best</b>
<b>nom</b>	<b>sg</b>	glömmande	glömmandet
	<b>pl</b>	glömmanden	glömmandena
<b>gen</b>	<b>sg</b>	glömmandes	glömmandets
	<b>pl</b>	glömmandens	glömmandenas

### Particip Perfekt

		<b>nom</b>	<b>gen</b>
<b>obest</b>	<b>utr</b>	glömd	glömds
	<b>neut</b>	glömt	glömts
	<b>pl</b>	glömda	glömdas
<b>best</b>	<b>sg</b>	glömda	glömdas
	<b>pl</b>	glömda	glömdas

ParadigmsFin.mkN “olut” “oluita”

	<b>yksikkö</b>	<b>monikko</b>
<b>nominatiivi</b>	olut	oluet
<b>genetiivi</b>	oluen	oluiden
<b>partitiivi</b>	olutta	oluita
<b>translatiivi</b>	olueksi	oluiksi
<b>essiivi</b>	oluena	oluina
<b>inessiivi</b>	oluessa	oluissa
<b>elatiivi</b>	oluesta	oluita
<b>illatiivi</b>	olueen	oluisiin
<b>adessiivi</b>	oluella	olulla
<b>ablatiivi</b>	oluelta	oluilta
<b>allatiivi</b>	oluelle	oluille
<b>abessiivi</b>	oluetta	oluitta
<b>komitatiivi</b>		oluline
<b>instruktiivi</b>		oluin

Function	Type	Example
genericCl	VP -> CI	one sleeps
mkCl	NP -> V -> CI	she sleeps
mkCl	NP -> V2 -> NP -> CI	she loves him
mkCl	NP -> V3 -> NP -> NP -> CI	she sends it to him
mkCl	NP -> VV -> VP -> CI	she wants to sleep
mkCl	NP -> VS -> S -> CI	she says that I sleep
mkCl	NP -> VQ -> QS -> CI	she wonders who sleeps
mkCl	NP -> VA -> A -> CI	she bec
mkCl	NP -> VA -> AP -> CI	she bec
mkCl	NP -> V2A -> NP -> A -> CI	she pai
mkCl	NP -> V2A -> NP -> AP -> CI	she pai
mkCl	NP -> V2S -> NP -> S -> CI	she ans
mkCl	NP -> V2Q -> NP -> QS -> CI	she ask
mkCl	NP -> V2V -> NP -> VP -> CI	she beg
mkCl	NP -> VPSlash -> NP -> CI	she beg
mkCl	NP -> A -> CI	she is o
mkCl	NP -> A -> NP -> CI	she is o
mkCl	NP -> A2 -> NP -> CI	she is n
mkCl	NP -> AP -> CI	she is v
mkCl	NP -> NP -> CI	she is th
mkCl	NP -> N -> CI	she is a
mkCl	NP -> CN -> CI	she is a
mkCl	NP -> Adv -> CI	she is h
mkCl	NP -> VP -> CI	she alw
mkCl	N -> CI	there is
mkCl	CN -> CI	there is
mkCl	NP -> CI	there at
mkCl	NP -> RS -> CI	it is she

- API: mkUtt (mkCl she\_NP wonder\_VQ (mkQS (mkQC1 who\_IP sleep\_V)) )
- Afr: sy vra hom af wie slaap
- Bul: ma ce y'cydea koiu cnu
- Cat: ella es pregunta qui dorm
- Chi: 她好奇谁睡觉
- Dan: hun undres hvem som sover
- Dut: ze vraagt zich af wie slaapt
- Eng: she wonders who sleeps
- Est: tema arutleb kes magab
- Fin: hän ihmettelee , kuka nukkuu
- Fre: elle se demande qui dort
- Ger: sie wundert sich , wer schläft
- Gre: αυτή αναρωτιέται ποιος κοιμάται
- Hin: वह हैरान होती है कि कौन सोता है
- Ita: lei si domanda chi dorme
- Jpn: 彼女は誰が寝るか質問する
- Lav: viņa brīnās , kurš guļ
- Mlt: hi tikkuntempla min jorqod
- Mon: мүүний зайхдаг нь
- Nep: उनी अचम्म हुन्छिन्की को सुत्छन्
- Nor: hun undrer hvem som sover
- Pes: constant not found: wonder\_VQ
- Pnb: او حیران ہوتی ہے کہ کون سووتا ہے
- Pol: ona zastanawia się , kto śpi
- Ron: ea se întrebă cine doarme
- Rus: она интересуется кто спит
- Snd: هو حیران ٿی ٿی ت کیر سمھی ٿو
- Spa: ella se pregunta quién duerme
- Swc: hon undrar vem som sover
- Tha: หล่อนประหลาดใจว่าใครนอนหลับ
- Urd: وہ حیران ہوتی ہے کہ کون سووتا ہے

# GF online editor for simple multilingual grammars

## Management

View:

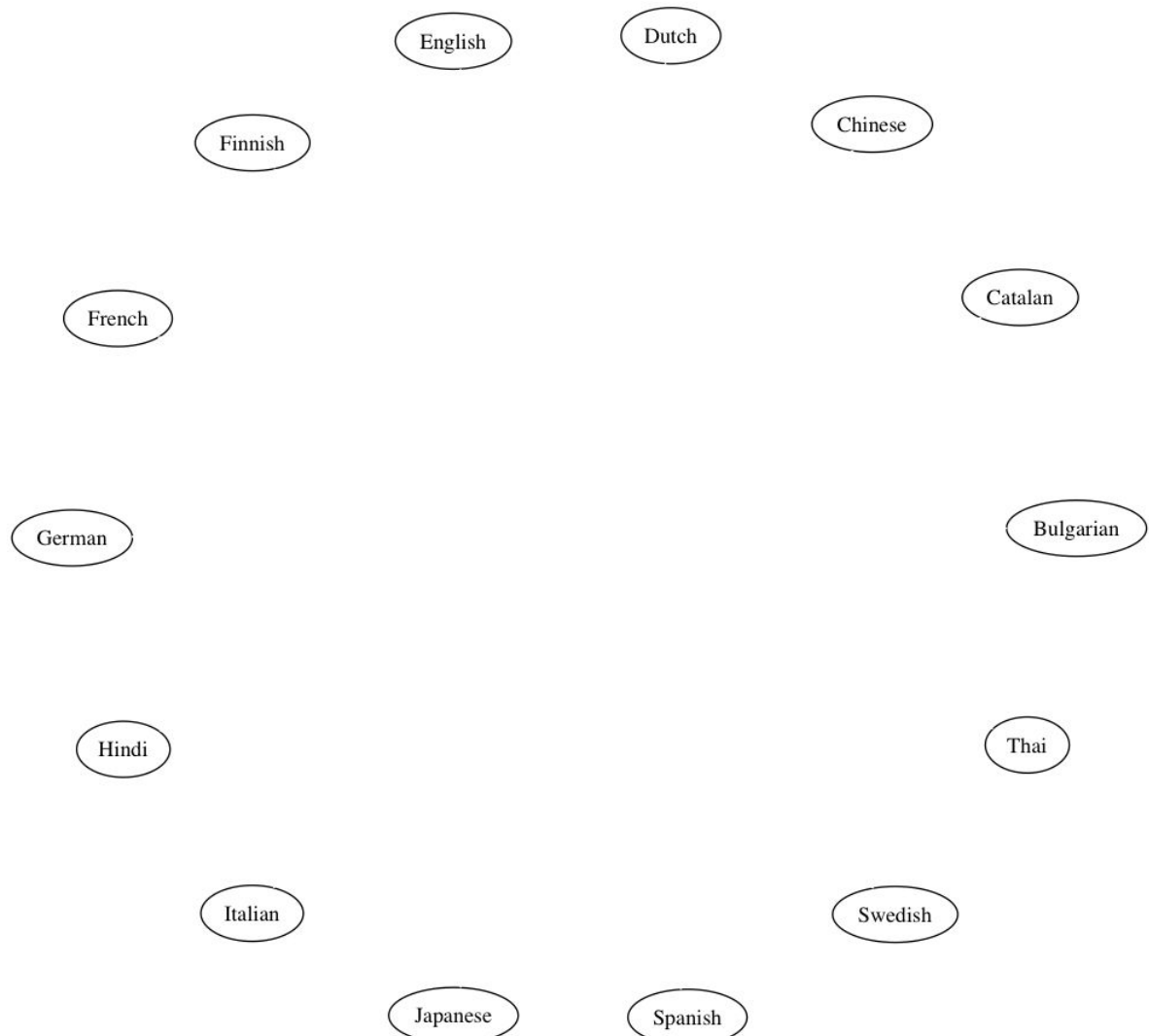
Abstract	English	Finnish	Swedish
<b>cat</b> Action	CI	CI	CI
<b>cat</b> Category	CN	CN	CN
<b>cat</b> Position	NP	NP	NP
<b>cat</b> Report	S	S	S
Appoint : Position → Position → Action	\ position1 position2 → mkCI position1 (mkV2 (mkV "appoint")) position2	\ position1 position2 → mkCI position1 (mkV2 (mkV "nimittää")) position2	\ position1 position2 → mkCI position1 (mkV2 (mkV "utse")) position2
Dismiss : Position → Position → Action	\ position1 position2 → mkCI position1 (mkV2 (mkV "dismiss")) position2	\ position1 position2 → mkCI position1 (mkV2 (mkV "erottaa")) position2	\ position1 position2 → mkCI position1 (mkV2 (mkV "avskeda")) position2
PIndef : Category → Position	\ category → mkNP a_Det category	\ category → mkNP a_Det category	\ category → mkNP a_Det category
PTwo : Category → Position	\ category → mkNP (mkNumeral "2") category	\ category → mkNP (mkNumeral "2") category	\ category → mkNP (mkNumeral "2") category
Board : Category	mkCN (mkN "board")	mkCN (mkN "hallitus")	mkCN (mkN "styrelse")
DivisionHead : Category	mkCN (mkN "division head")	mkCN (mkN "osasto" (mkN "päällikkö"))	mkCN (mkN "avdelningchef")
VicePresident : Category	mkCN (mkN "vice president")	mkCN (mkN "varapuheenjohtaja")	mkCN (mkN "vice ordförande")
PAll : Category → Position	\ category → mkNP all_Predet (mkNP aPl_Det category)	\ category → mkNP all_Predet (mkNP thePl_Det category)	\ category → mkNP all_Predet (mkNP aPl_Det category)
RPerfect : Action → Report	\ action → mkS anteriorAnt action	\ action → mkS anteriorAnt action	\ action → mkS anteriorAnt action
RPresent : Action → Report	\ action → mkS action	\ action → mkS action	\ action → mkS action
RFuture : Action → Report	\ action → mkS futureTense action	\ action → mkS action	\ action → mkS futureTense action
PDef : Category → Position	\ category → mkNP the_Det category	\ category → mkNP the_Det category	\ category → mkNP the_Det category
RNegFuture : Action → Report	\ action → mkS futureTense negativePol action	\ action → mkS futureTense negativePol action	\ action → mkS futureTense negativePol action

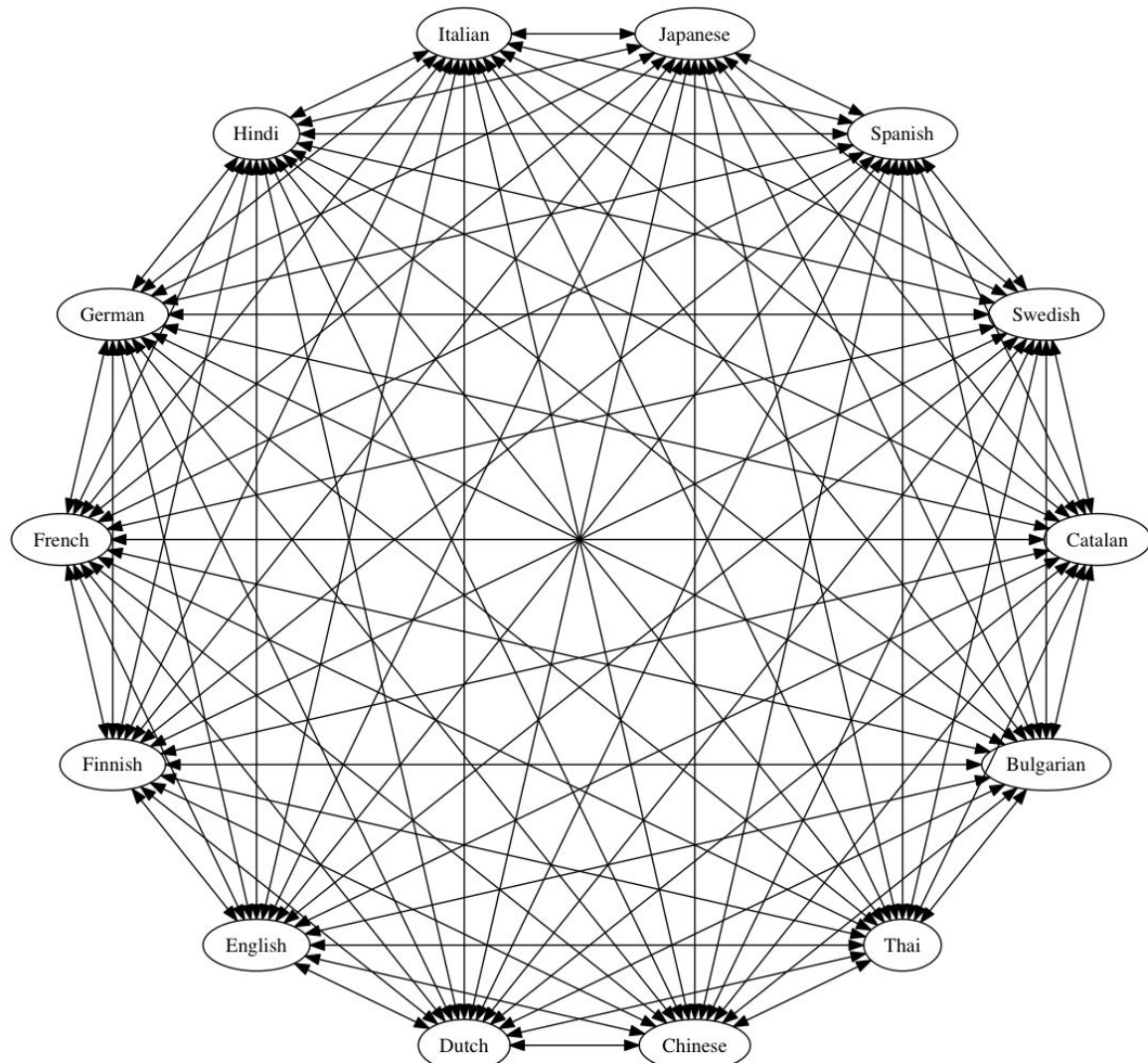
+

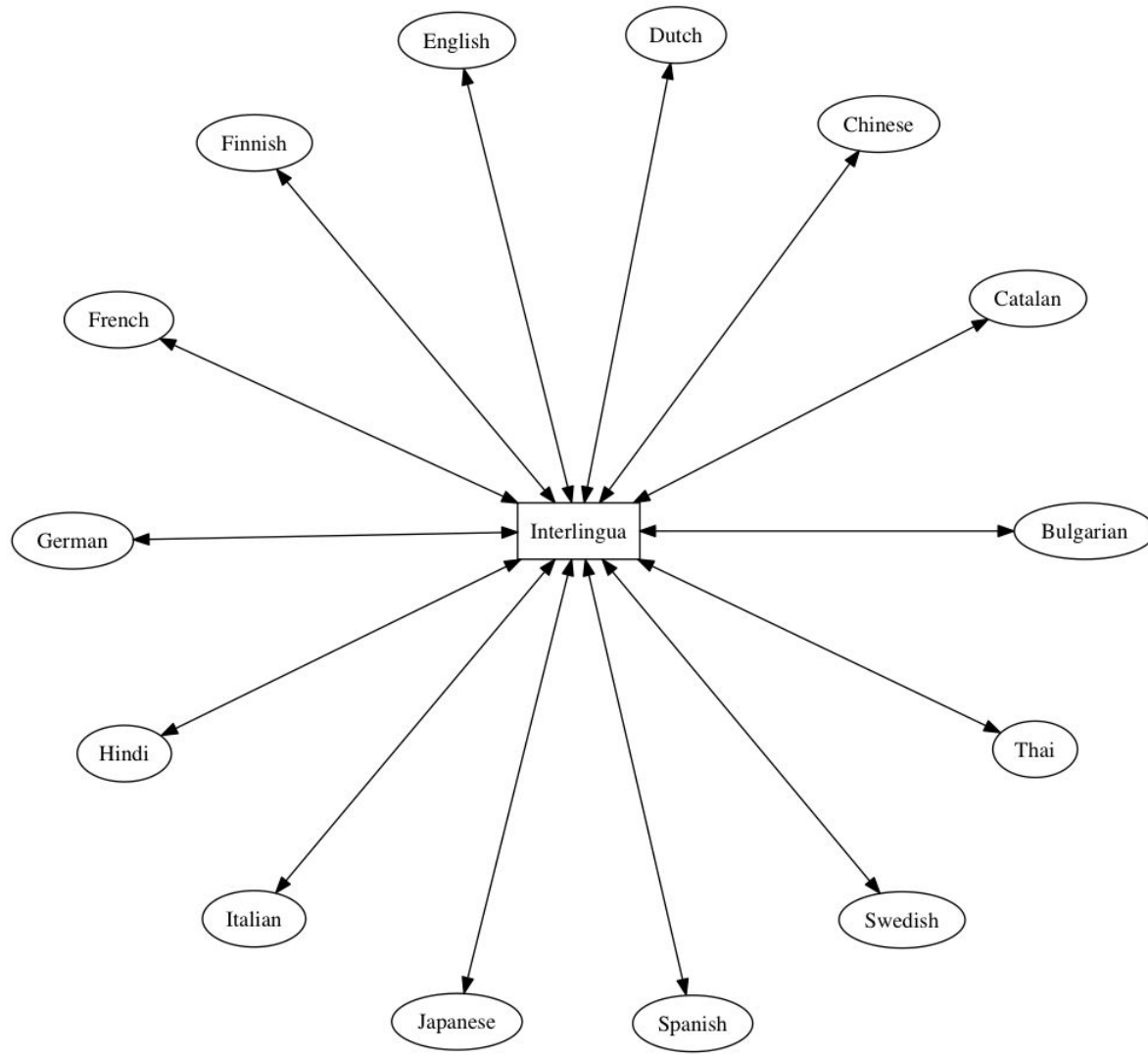


**Making grammars easier to build:**

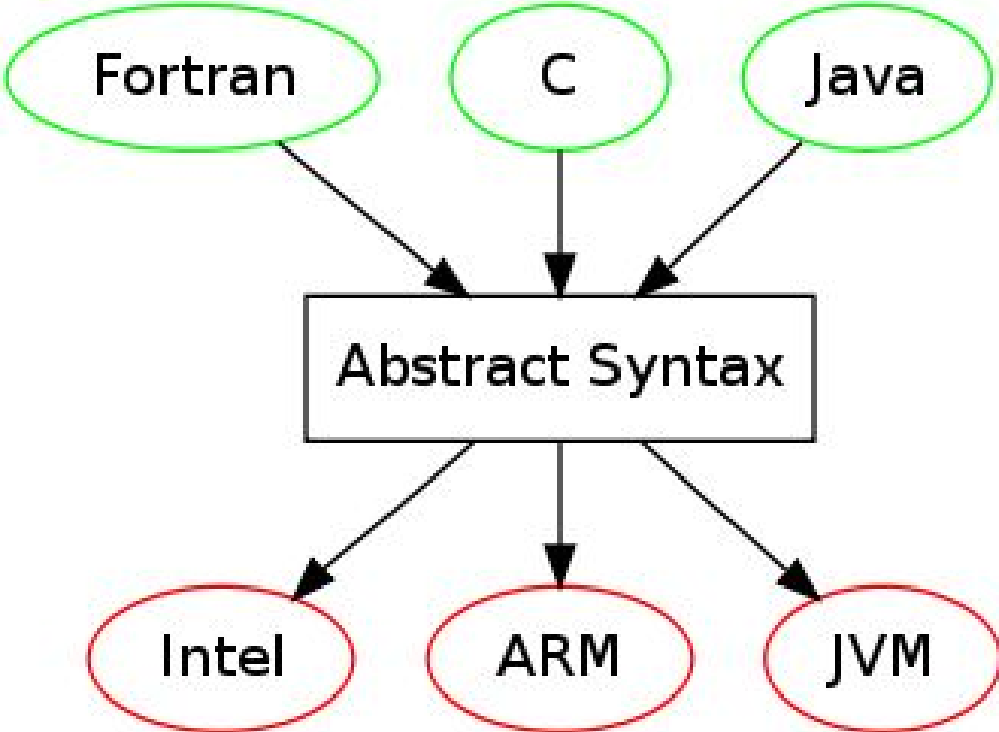
**interlinguality**







Inspiration: multi-source multi-target compilers



**Making grammars easier to build:**

**community**

# GF languages and contributors



Nieminen, Camilleri 2013-

	Norwegian	Danish	Afrikaans
	English	Swedish	German Dutch
Romanian	French	Italian	Spanish Catalan
Polish	Bulgarian	Finnish	Estonian
Russian	Japanese	Thai	Chinese Hindi
Latvian	Mongolian	Urdu	Punjabi Sindhi
Greek		Maltese	Nepali Persian
Latin	Turkish	Hebrew	Arabic Amharic Swahili



**Making grammars easier to build:**

**documentation, tutorials, training**



# Grammatical Framework

A programming language for multilingual grammar applications

## Use GF

- [GF Cloud](#) 
- [Android app](#)
- [Other Demos](#)
- [Download GF](#)
- [GF Eclipse Plugin](#)
- [GF Editor Modes](#)
- [User Group](#)
- [Bug Reports](#)
- [Blog](#)

## Learn GF

- [QuickStart](#)
- [QuickRefCard](#)
- [GF Shell Reference](#)
- [GF Summer School](#)
- [The GF Book](#)
- [GF Tutorial](#)
- [Reference Manual](#)
- [Best Practices](#) [PDF]
- [Library Synopsis](#)
- [Library Tutorial](#) [PDF]
- [Coverage Map](#)

## Develop GF

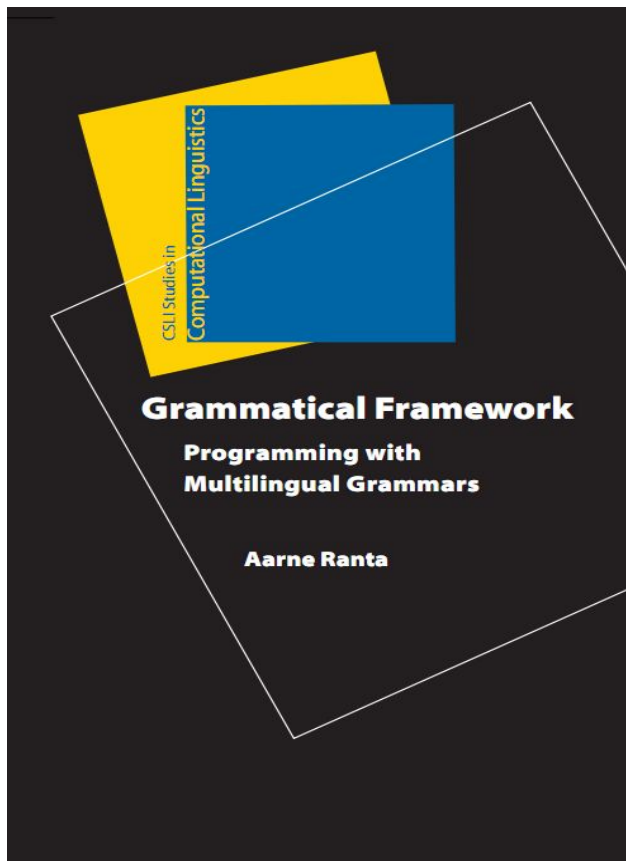
- [build](#) [passing](#)
- [GF Developers Guide](#)
- [GitHub mirror](#)
- [Wiki](#)
- [Browse Source Code](#)
- [Authors](#)

## Develop Applications

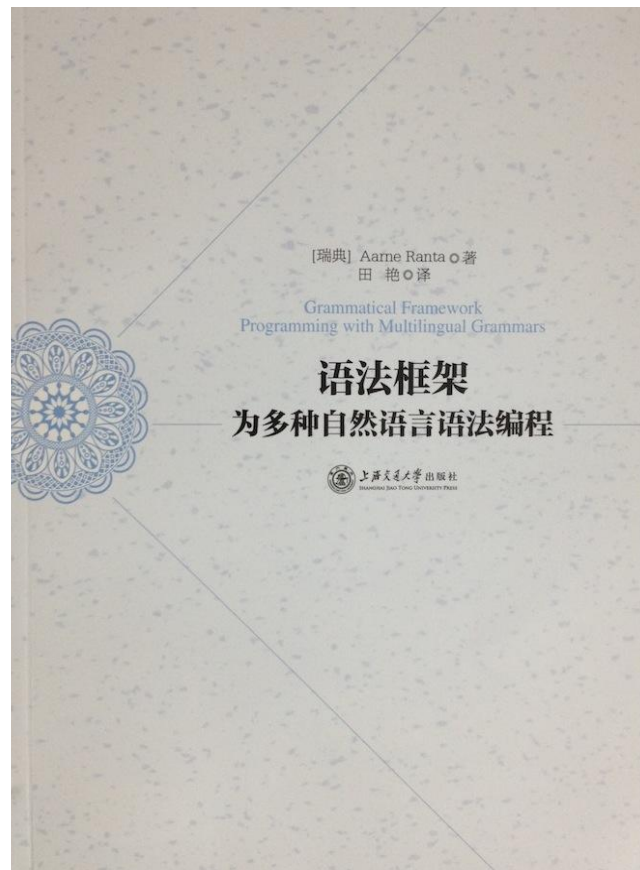
- [PGF library API \(Haskell\)](#)
- [PGF library API \(Python\)](#)
- [GF on Android \(new\)](#)
- [GF on Android \(old\)](#)

## Related to GF

- [Publications](#)
- [GF Summer Schools](#)
- [The REMU Project](#)
- [The MOLTO Project](#)
- [GF on Wikipedia](#)
- [Digital Grammars AB](#)



CSLI, Stanford, 2011



Shanghai Jiao Tong University press, 2014

# GF Summer School 2011

GF Resou



## Frontiers of Multilingual Te

August 15-26, 2011

Barcelona, Spain.

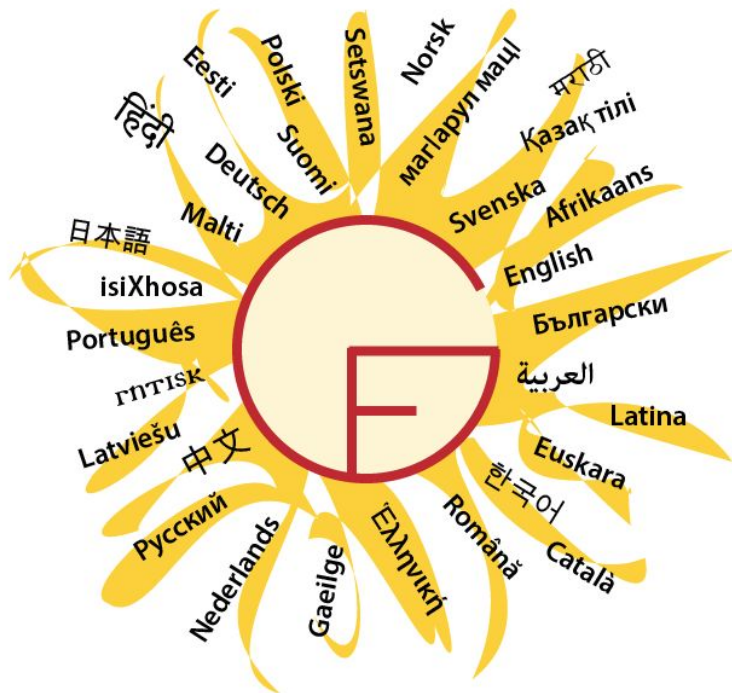
# Third GF Sum Frontiers of Multilingual

Special theme of 2013: Sci



Lake Chiemsee, Bavaria  
18<sup>th</sup>–30<sup>th</sup> August, 2013

# Fourth GF Summer School 2015

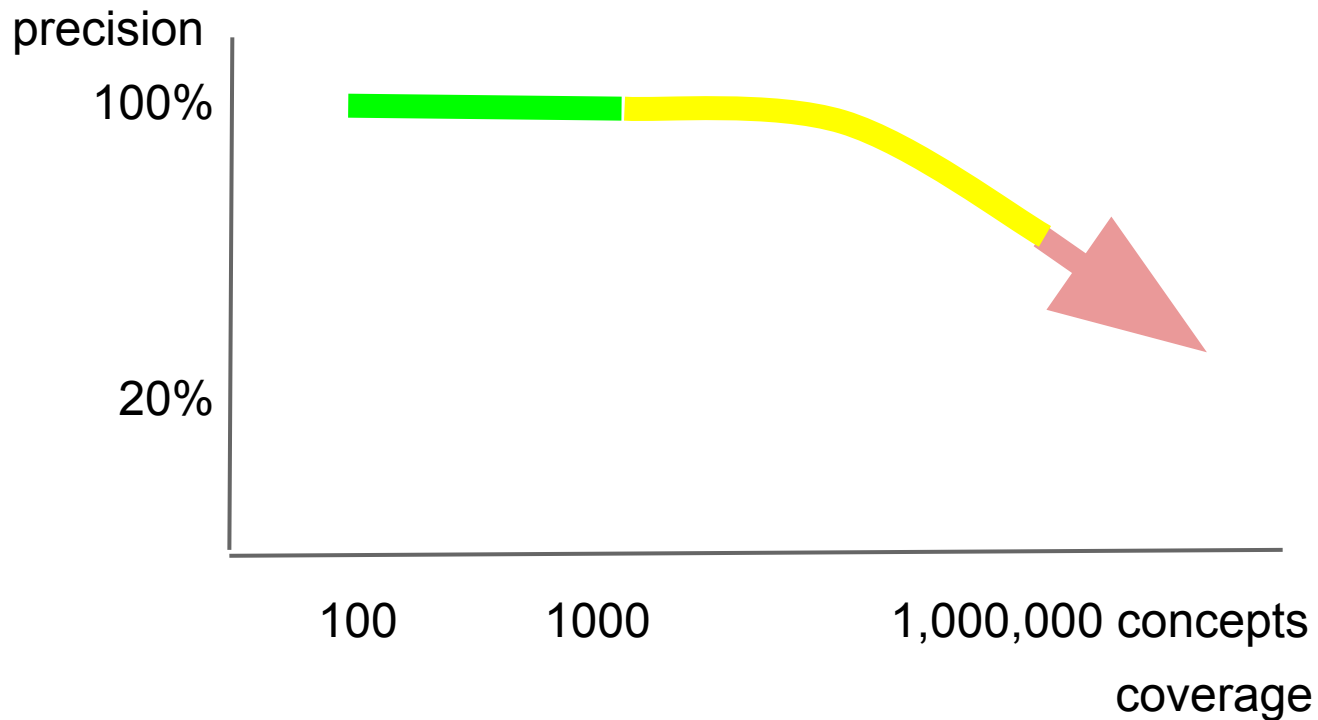


Gozo, Malta  
13<sup>th</sup>–24<sup>th</sup> July, 2015

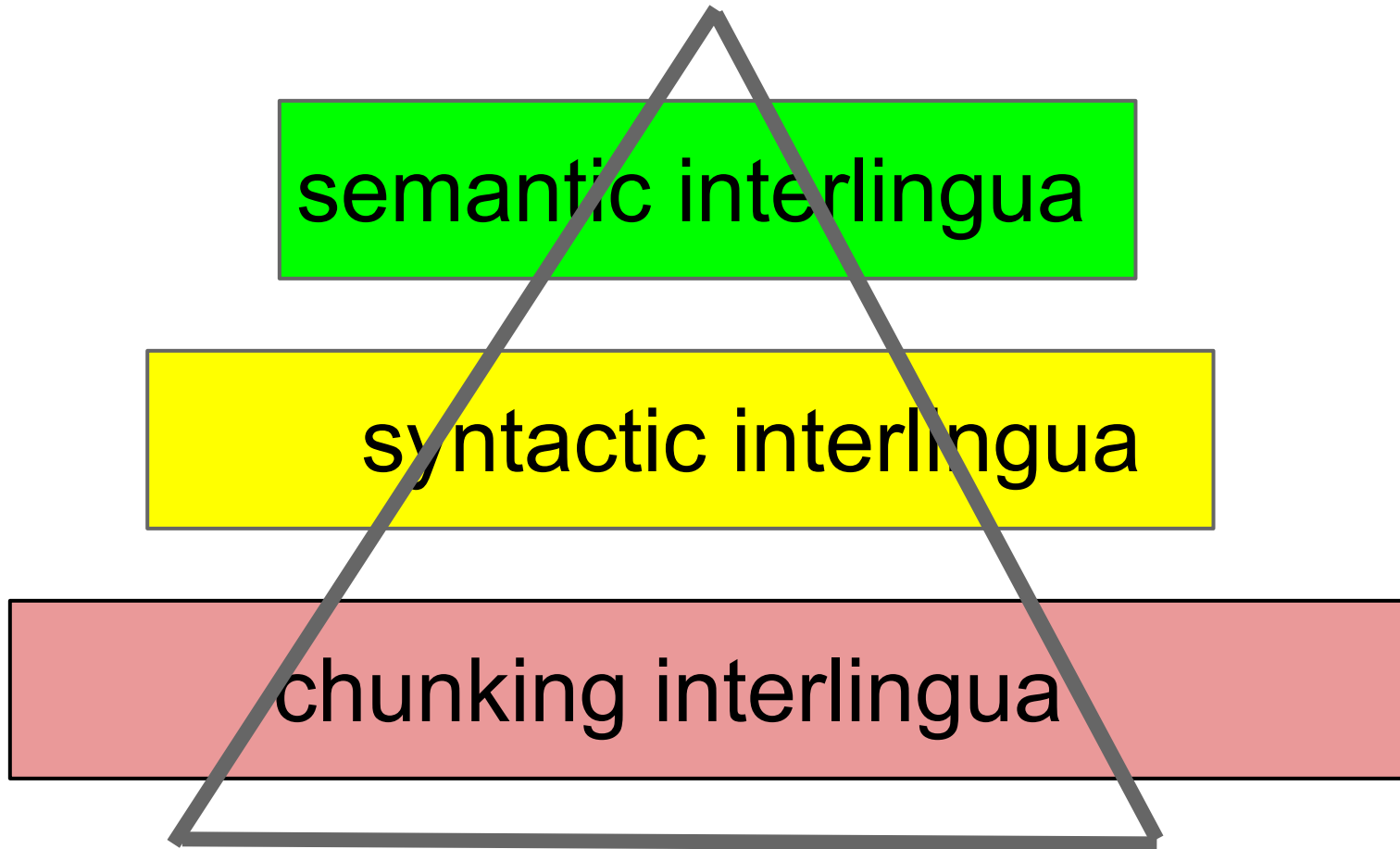
GF Summer School  
Gothenburg, Sweden  
August 17-28, 2009

**Increasing the coverage of grammars**

# Graceful degradation



Translation layers (the Vauquois triangle)



what is your wife's name

vad heter din fru

the vice president kicked the bucket

skruvstädspresidenten  
sparkade hinken

long time no see

lång tid nej ser

translation by **meaning**

- correct
- idiomatic

translation by **syntax**

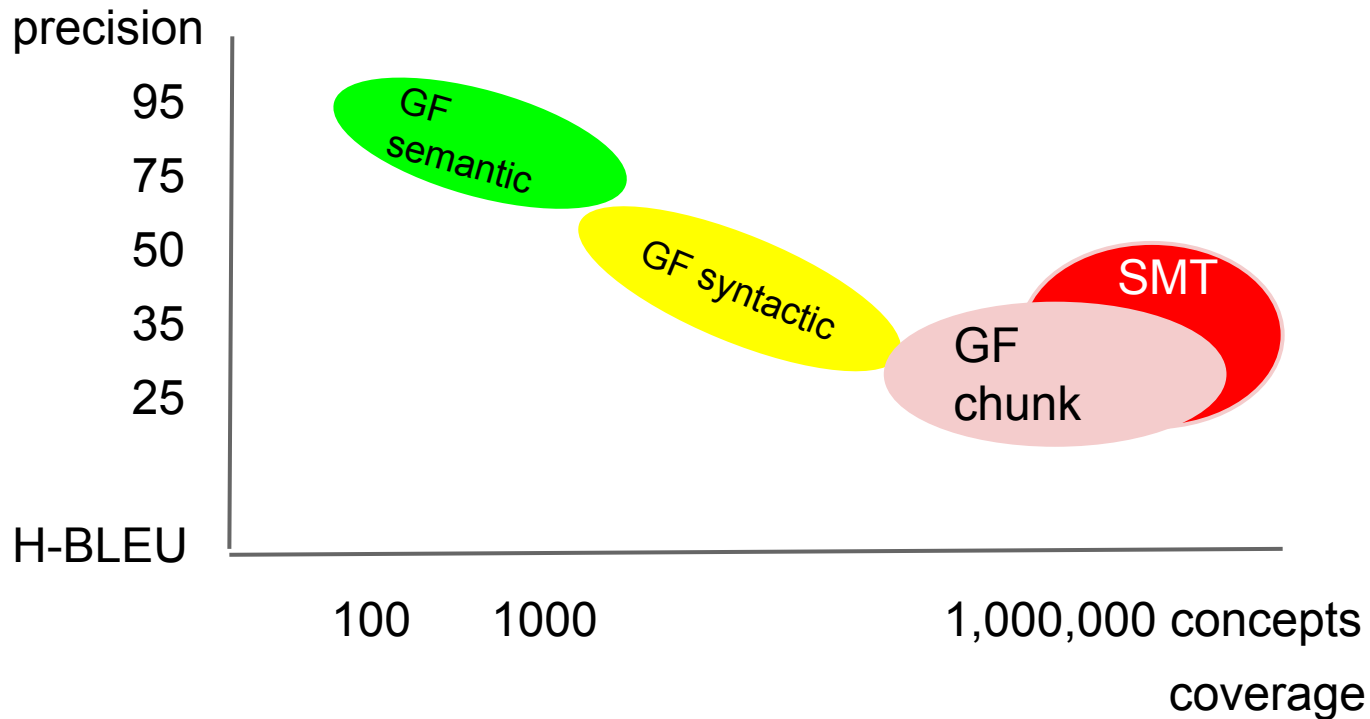
- grammatical
- often strange
- often wrong

translation by **chunks**

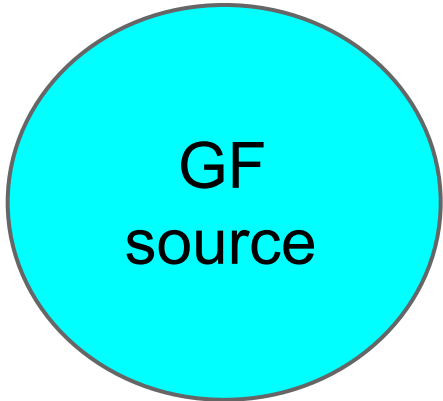
- probably ungrammatical
- probably wrong

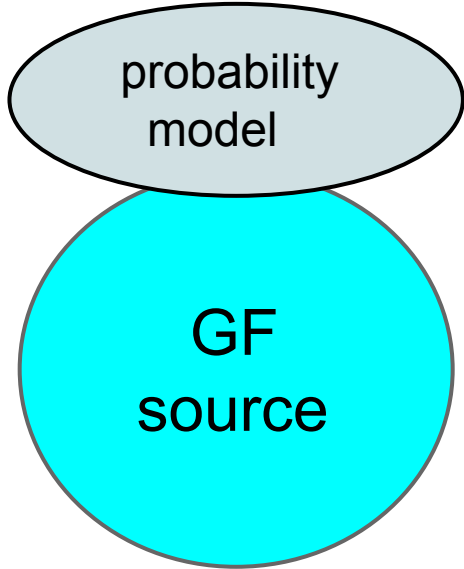


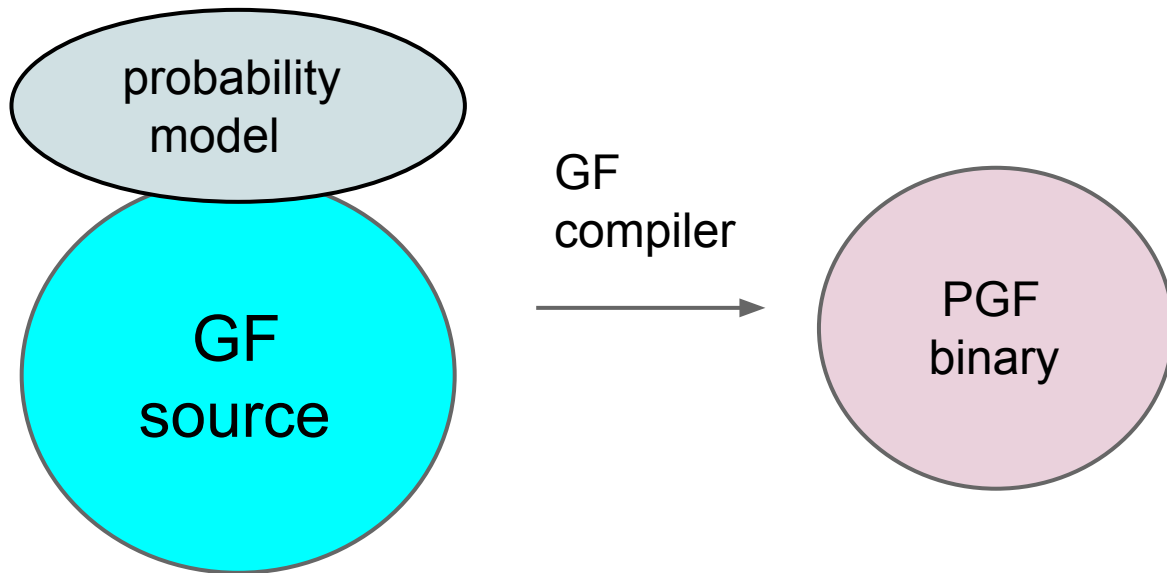
# Where we are now

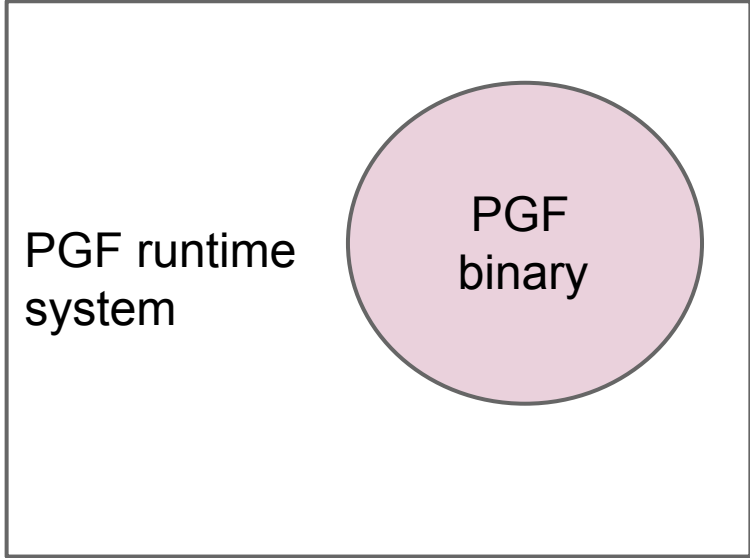


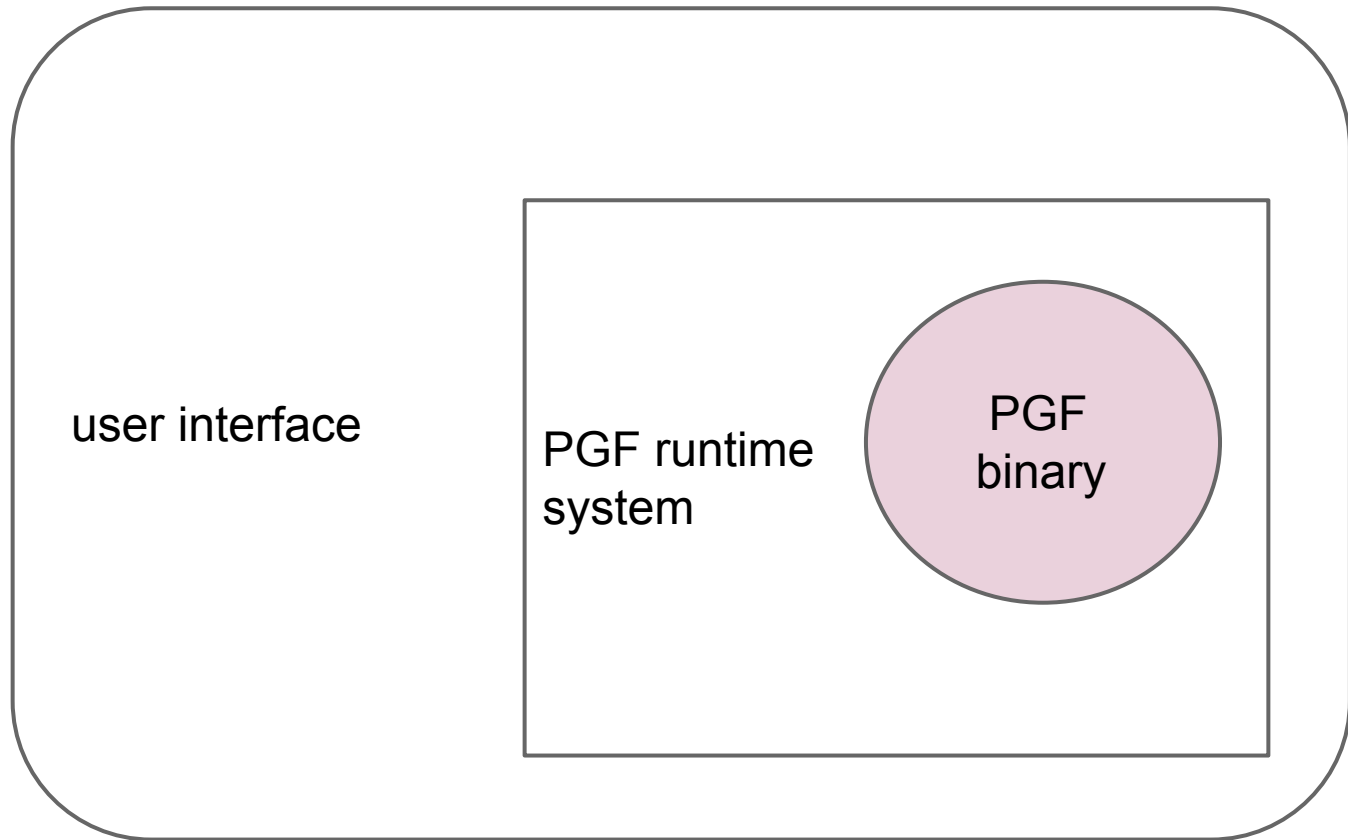
# **Building GF applications**







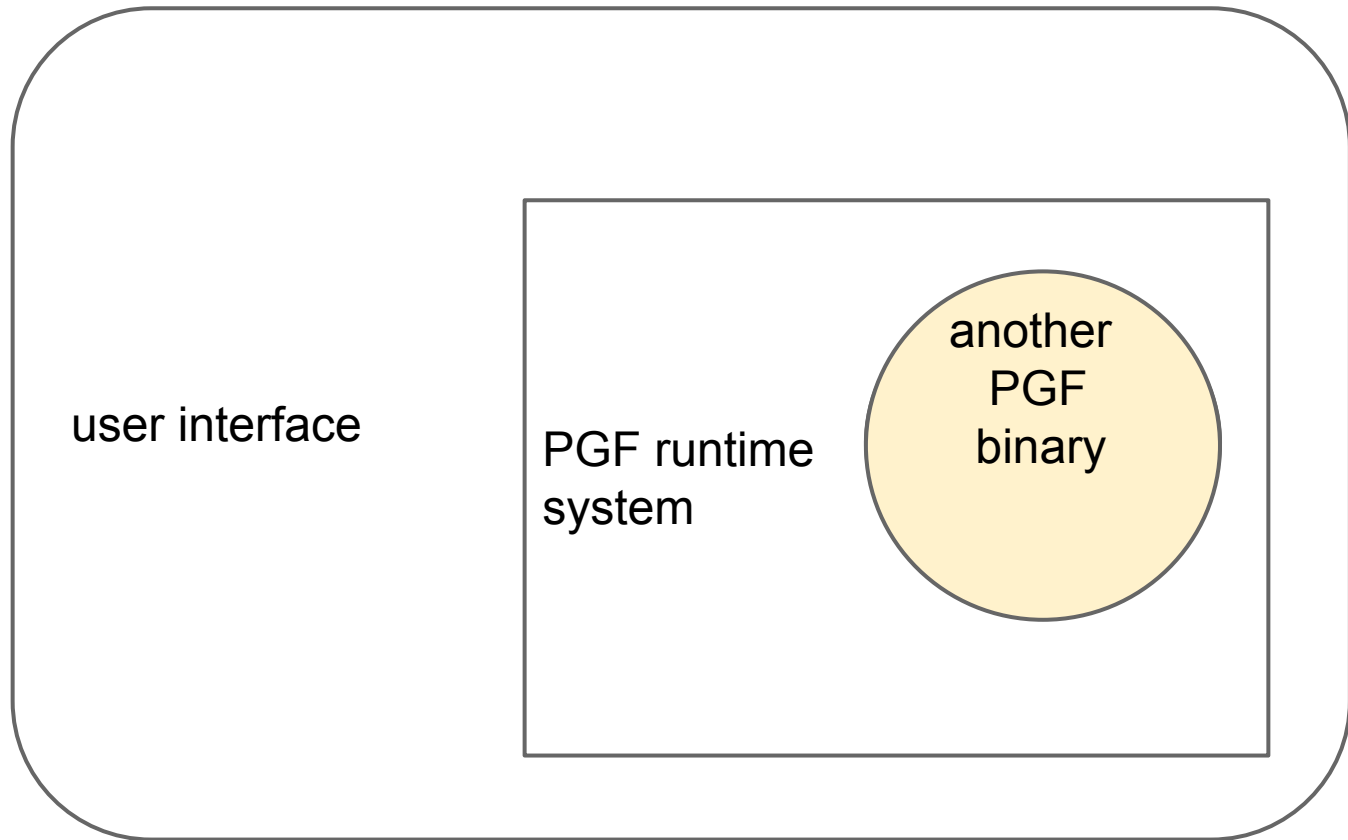




user interface

PGF runtime  
system

PGF  
binary



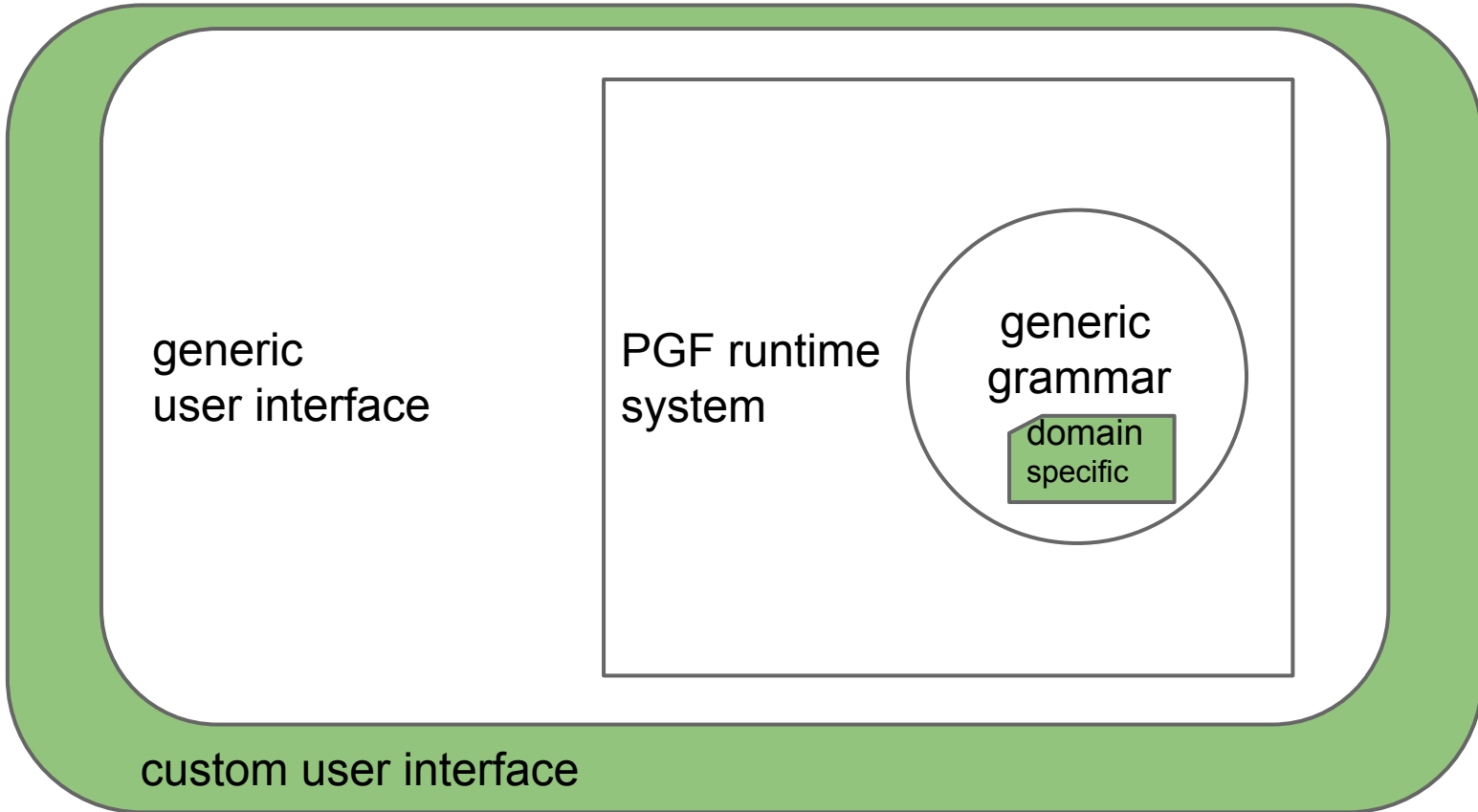
user interface

PGF runtime  
system

another  
PGF  
binary

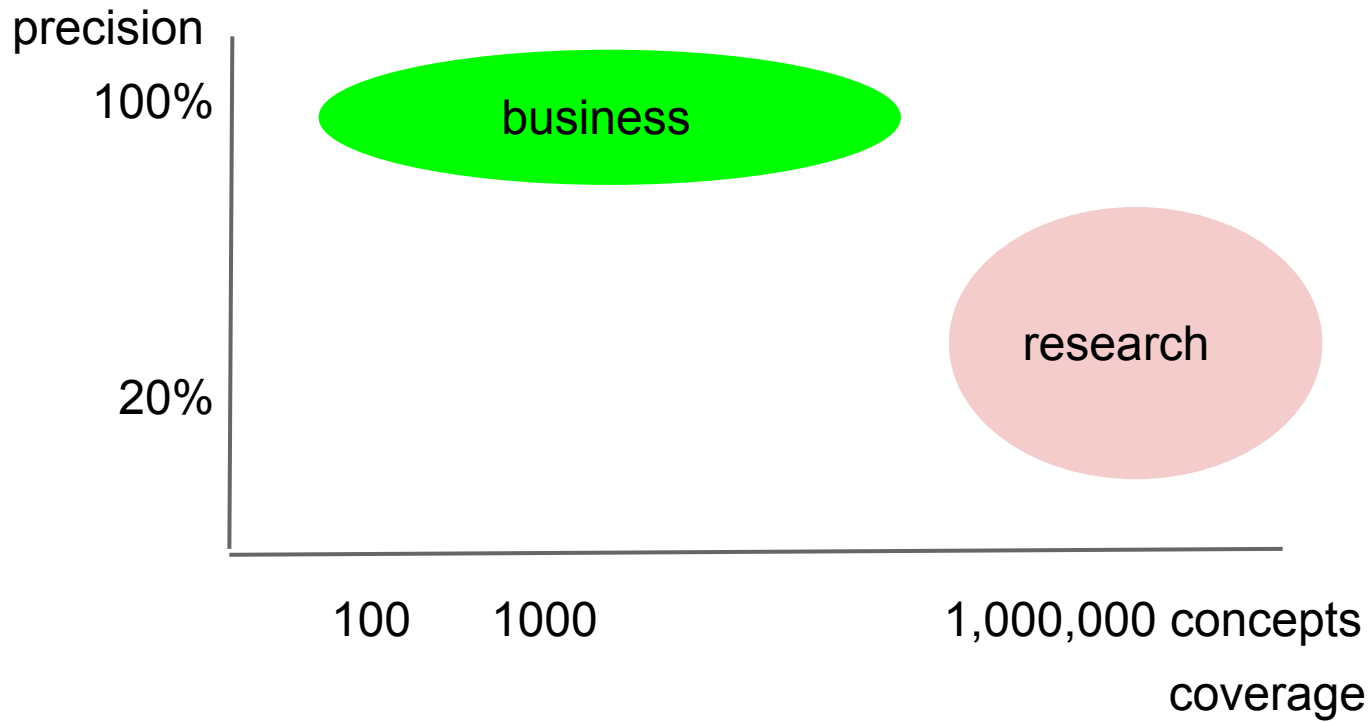


**White:** free, open-source (LGPL/BSD). **Green:** a business idea



**Making business with GF:**

**Digital Grammars AB**



*We build solutions for customers who*

- need to translate to many languages*
- need to update the content often*
- find Google translate to have too low quality*
- find human translation to be too expensive  
or too slow*

# Candidates

Health care

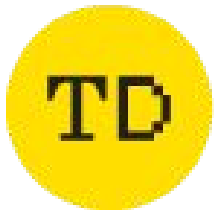
E-commerce

Legal documentation, contracts

Technical documentation, manuals

Robot journalism

Software localization



- 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

UttSTD (PredUttTD (AdvNPTD (DetCNTD (DetQuant DefArt NumSg) (UseNTD svängrumsyta\_NTD)) (PrepNPTD utanför\_Prep (DetCNTD (DetQuant DefArt NumSg) (UseNTD dörr\_NTD)))) (AdvVPTD (luta\_VPTD (procentMeasure 1)) i\_sidled\_AdvTD))

**Talkamatic**  
FREE DIALOGUE



digital **G**rammars  
Language technology to rely on.

`next_membership_level_sys_answer silver (next_membership_points_sys_answer integer0_99_50)`

`test_mockup_travelChi: 您有五十个常旅客点符合会员条件, 您现在是在伦敦.`

`test_mockup_travelDut: je hebt vijftig punten nodig om het zilveren niveau te bereiken`

`test_mockup_travelEng: you need fifty points to reach silver level`

`test_mockup_travelFin: sinä tarvitset viisikymmentä pistettä päästäksesi hopeatasolle`

`test_mockup_travelFre: tu as besoin de cinquante points pour atteindre le niveau argent`

`test_mockup_travelGer: Sie brauchen fünfzig Punkte um das Silberniveau zu erreichen`

`test_mockup_travelIta: avete bisogno di cinquanta punti per raggiungere il livello argento`

`test_mockup_travelSpa: necesitas cincuenta puntos para llegar al nivel plata`

Home Visual Scenes

Brian - Visual Scenes

<http://brian.com/visual-scenes>

Välj vad du vill prata om

Huvudmeny

Skrivet språk

Talat språk

Logo mat

Trädgården

Fika

Se på TV

I don't know where my wife is	non riesco a trovare mia moglie	jag vet inte var min fru är	ich weiss nicht, wo meine Frau ist	ne vem, kje je moja žena
I don't know where my husband is	non riesco a trovare mio marito	jag vet inte var min man är	ich weiss nicht, wo mein Mann ist	Ne vem, kje je moj muž
I'm hurt	mi sono ferito	jag har skadat mig	ich bin verletzt	Poškodovan sem
I need a doctor	ho bisogno di un medico	jag behöver en läkare	ich brauche einen Arzt	rabim zdravnika
I'm allergic to penicillin	Sono allergico alla penicillina	jag är allergisk mot penicillin	ich bin allergisch gegen Penicilin	sem alergičen na penicilin
I'm dizzy	mi gira la testa	jag är yr	mir ist schwindelig	Sem omotična
I have nausea	ho la nausea	jag mår illa	ich habe Übelkeit	Čutim slabost, počutim se slabo
my arm hurts	mi fa male il braccio	jag har ont i armen	mein Arm tut weh	me roka boli



# **Other companies that have used GF**

Galois, USA

Lingsoft, Finland

Ontotext, Bulgaria

Be Informed, The Netherlands

Talkamatic, Sweden

**Take home points**

# **GF is an open source software project with**

- 200 active members from 20+ countries
- grammar libraries for over 30 language
- support for mobile and web applications
- scales up to wide coverage (but low quality)
- mature for commercial precision applications
  - translation
  - natural language interaction

