



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

Multilingual CNL in GF and at Digital Grammars

Aarne Ranta

Department of Computer Science and Engineering,
Chalmers University of Technology & University of Gothenburg
and
Digital Grammars AB



World Forum on Controlled Languages
Florence, 16-17 March 2018



CNL = Controlled Natural Language

= language defined by a formal grammar (as the term is used in this talk)

Multilingual CNL = set of languages with a common abstract syntax

GF = Grammatical Framework - a programming language for multilingual CNLs

- started 1998 at Xerox, now open source, 250+ developers, 30+ languages

Digital Grammars Gothenburg AB - a start-up company using GF

- one of the 15+ companies using GF, started by GF core developers

Example: GDPR

GDPR

General Data Protection Regulation, Official Journal of the EU

24 official EU languages

60-80k words in each language

2500-3000 unique lemmas in each language

To enter in force 25 May 2018

Commercial project at Digital Grammars AB with Signatu AS (Oslo)

With

Georg Philip Krog *international law*

Christina Unger *abstract syntax, German*

Jordi Saludes *Spanish*

Sara Negri *Italian*

Daniel von Plato *Italian*

Grégoire Détérez *French*

Markus Forsberg *corpus analysis*

Koen Lindström Claessen *word alignment*

Thomas Hallgren *visual effects*

John Camilleri *all aspects of lexicon and supporting software*

Our task

Construct a CNL that permits

- creating documents in multiple languages automatically
- searching for concepts in many languages simultaneously
- reasoning about content independently of language

→ identify translation-equivalent concepts in the corpus

Concept and its translations

German
rechtmäßig

French
légal

English
legal

Abstract
legal

Italian
legale

Finnish
laillinen

Sense distinctions

German
juristisch

English
legal

French
juridique

Abstract
legal_2

Italian
giuridico

Finnish
juridinen

Grammatical information

German

juristisch,
juristischer,
juristischen,
juristisches,

English
legal

French

juridique,
juridiques

Abstract

legal_2 : Adj

Italian

giuridico,
giuridica,
giuridici,
giuridiche

Finnish

juridinen,
juridisen,
juridista,
juridisena,
juridisessa,

Multiword expression

German

Rechtswirkung,
Rechtswirkungen
(feminine)

English

legal effect,
legal effects

French

effet juridique,
effets juridiques
(masculine)

Abstract
legal_effect :
Noun

Italian

effetto giuridico,
effetti giuridici
(masculine)

Finnish

oikeusvaikutus,
oikeusvaikutuksen,
oikeusvaikutusta,
oikeisvaikutuksena,
oikeusvaikutukseksi,
oikeusvaikutuksessa,

Expression with arguments

German
unbeschadet
X(genitive)

English
without
prejudice to X

French
sans
préjudice de X

Abstract
without_prejudice
: NP -> Adv

Italian
fatto salvo X,
fatta salva X,
fatti salvi X,
fatte salve X

Finnish
sanotun
kuitenkaan
rajoittamatta
X(partitive)

Syntax rule

German

ap(gender(cn),n,c)

cn(n,c)

English

ap cn(n)

French

cn(n)

ap(gender(cn),n)

Abstract

AdjCN :

AP -> CN -> CN

Italian

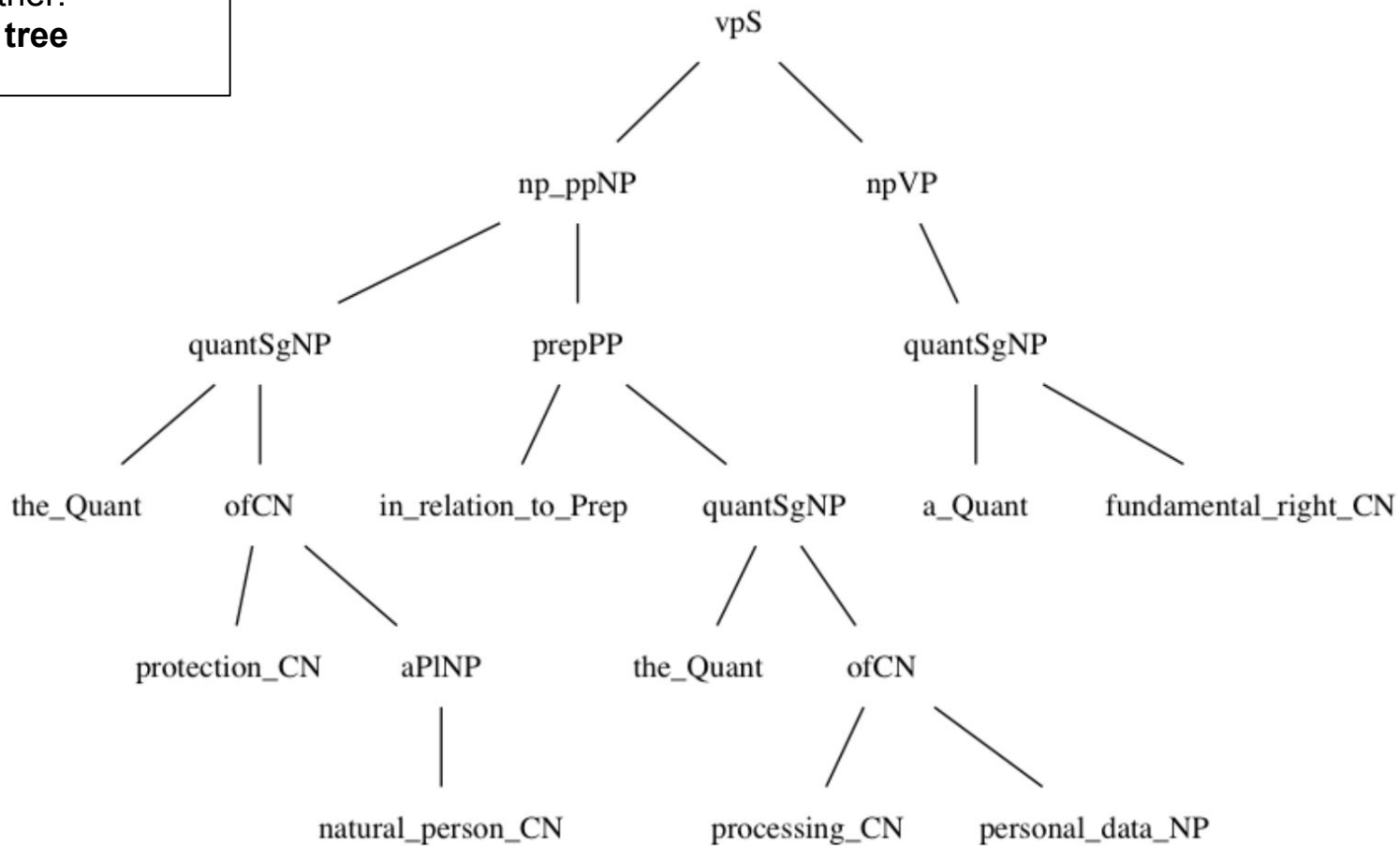
cn(n)

ap(gender(cn),n)

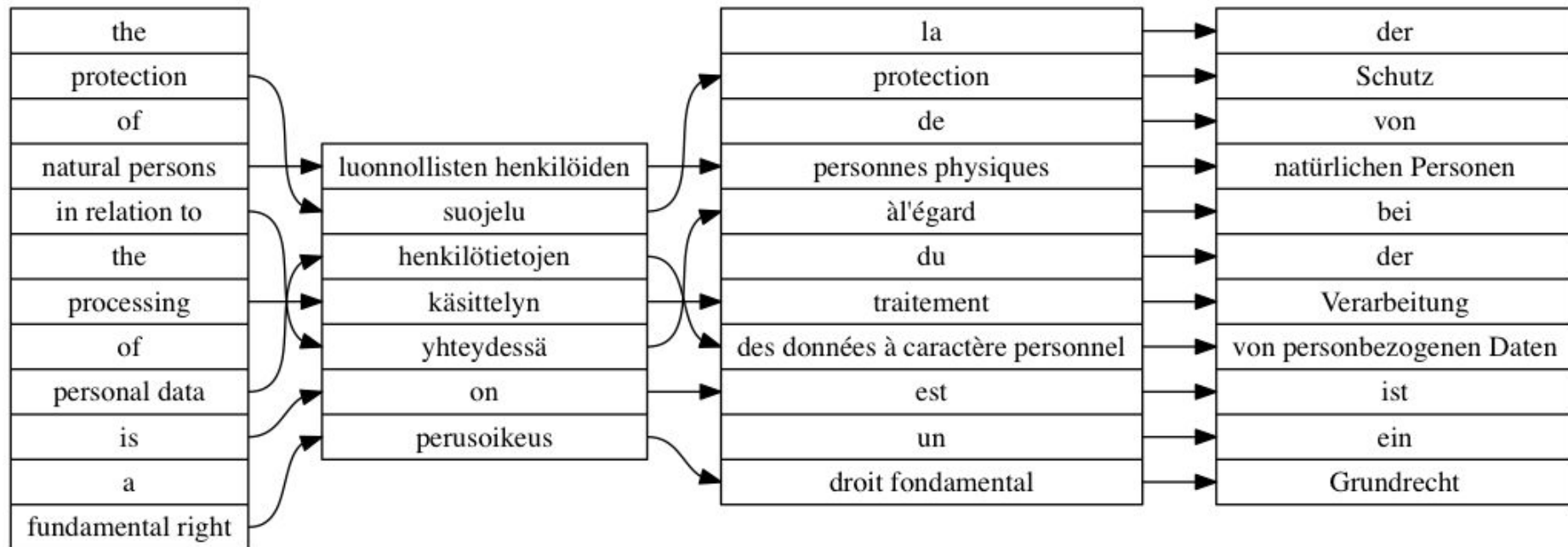
Finnish

ap(n,c) cn(n,c)

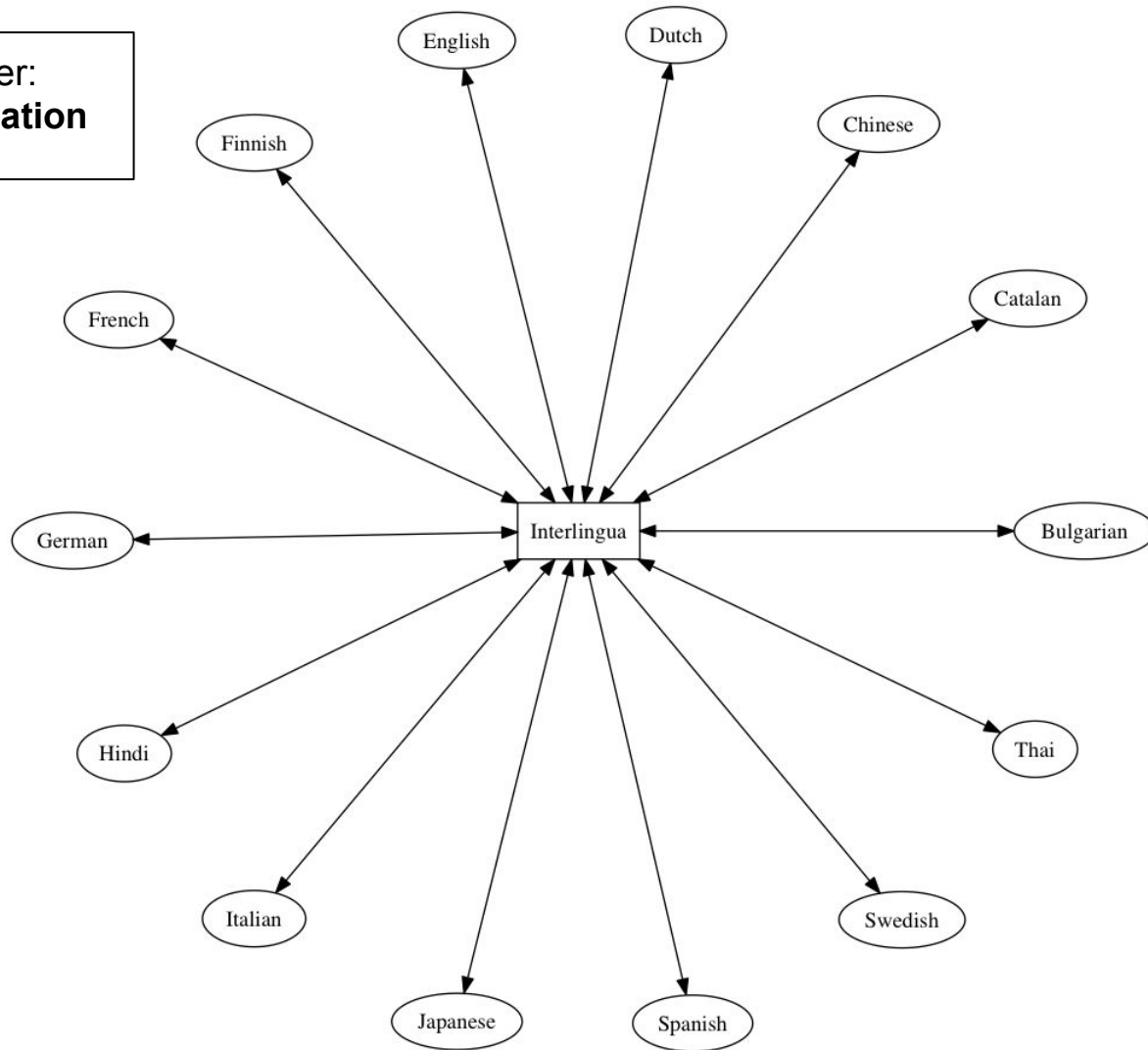
Putting it all together:
abstract syntax tree



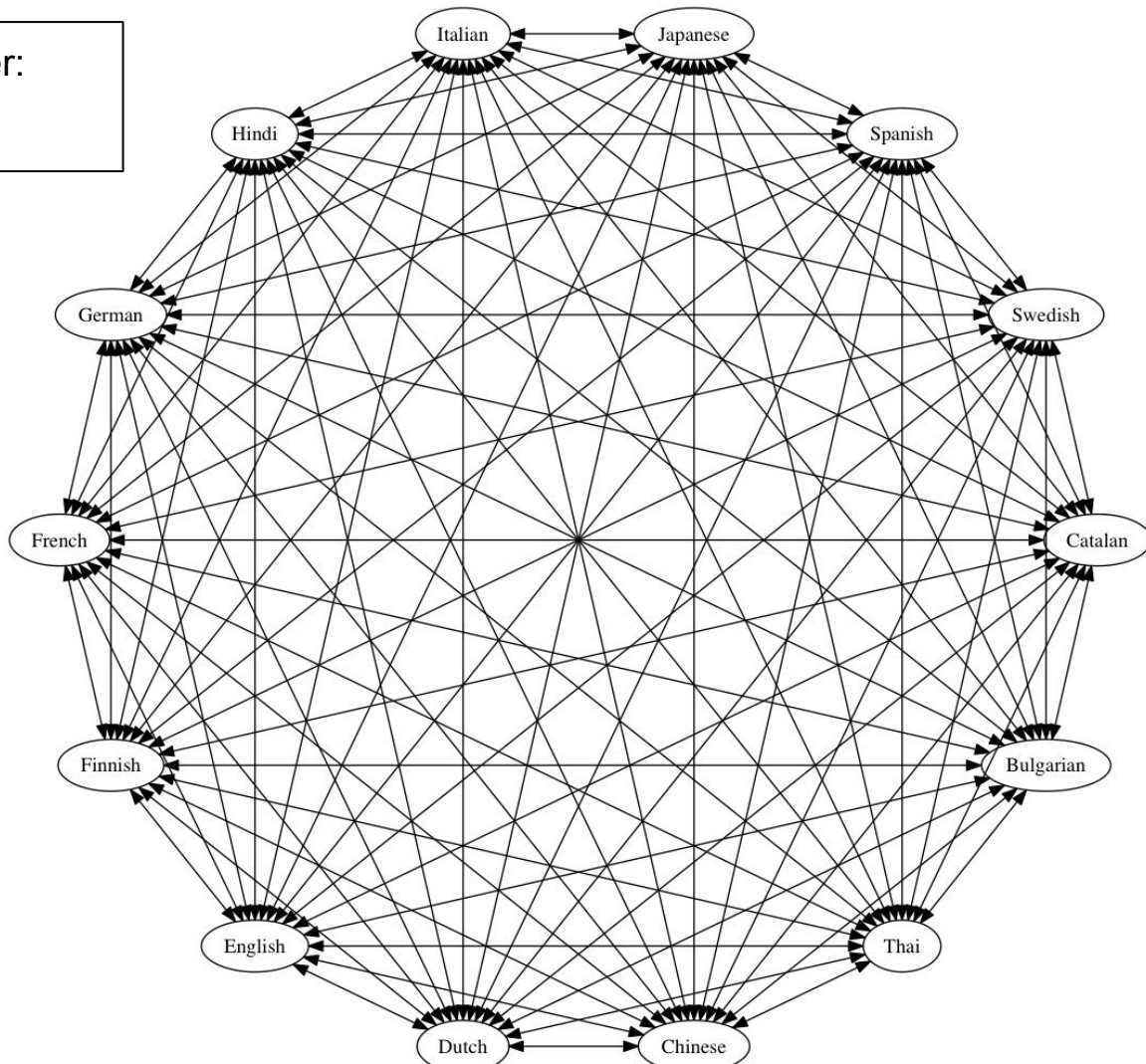
Putting it all together:
concept alignment



Putting it all together:
Interlingual translation



Putting it all together:
linear scale-up



Putting it all together:
shared reasoning

*Who controls
the controller?*

*Wer kontrolliert
den
Verantwortlichen?*

*Qui contrôle le
responsable
du traitement?*

*Chi controlla il
titolare del
trattamento?*

*Kuka valvoo
rekisterinpitäjää?*

**(Question
Who
(Control
(Definite
Data_controller
)
)
)**

**(Answer
(Definite
data_
protection_
supervisory_
authority
)
)
)**

*The data
protection
supervisory
authority.*

*Die
Datenschutzau-
sichtsbehörde.*

*L'autorité de
contrôle de la
protection des
données.*

*L'autorità di
controllo della
protezione dei
dati.*

*Tietosuojavalvon-
taviranomainen.*

Some statistics about the GDPR lexicon

	abstract
functions	3525
atomic	3272
syntactic	139
construction	114

	English	German	French	Italian	Spanish
word tokens	55186	54903	62198	55296	57383
word types	2625	4153	3206	3520	3498
lemma types	2555	3053	2467	2689	2478
multiword functions	590	227	574	559	594
multiwords in corpus	8-13%	3-10%	10-21%	10-20%	11-21%

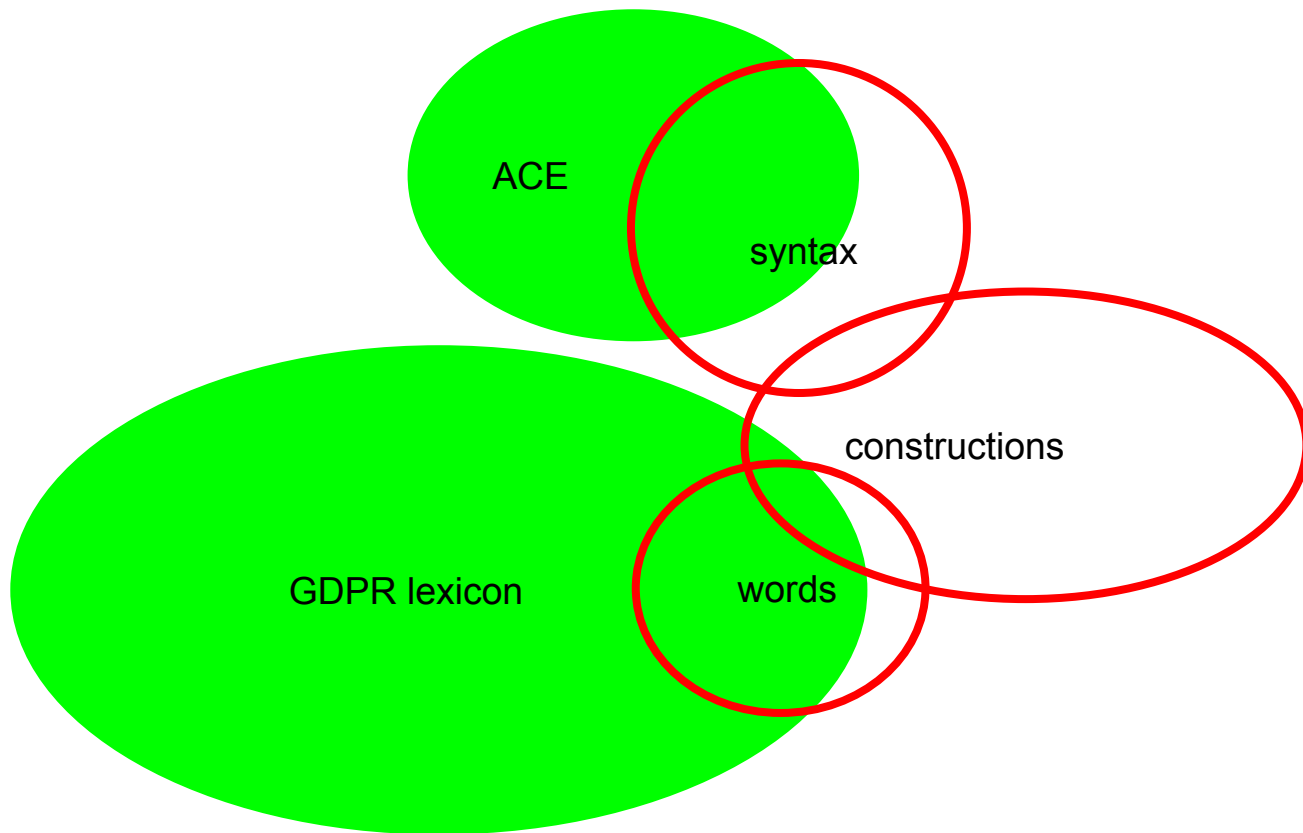
Is this the largest multilingual corpus ever analysed at this level of precision?

A screen dump from the lexicon demo page: <https://gdprlexicon.com>

9 results for "subject"

English	German	French	Italian	Spanish
be subject to X	X unterliegen	être soumise à X	essere sottoposta a X	ser sujeta a X
be subject to X	X unterstehen	être placée sous les ordres de X	sottostare X	ser sujeta a X
be subject to X	X unterworfen sein	faire l' objet de X	essere soggiogata a X	ser sujeta a X
person subject to law	Rechtsunterworfene	justiciable	persona soggetta alla legge	persona sujeta al ley
subject	betroffen	concerné	soggetto	sujeto
subject	Gegenstand	sujet	soggetto	objeto
subject-matter	Gegenstand	objet	oggetto	objeto
subject to X	vorbehaltlich X	sous reserve de X	soggetto a X	siempre que se den X
subject to X	fallend unter X	soumis à X	soggetto a X	interesados a X

From lexicon to CNL



ACE = Attempto
Controlled English
(Fuchs & al.)

Example: English original

7 Do we perform automated decision-making and profiling?

7.1 We use your personal data to make automatic decisions about you. The automatic decisions are made by a computer and are made without a human influence ('automatic decision-making').

7.2 We use your personal data to make automatic assessments about your personal characteristics and behaviour. The automatic assessments may include analysis of your characteristics or predictions of your behaviour. The automatic assessments of your characteristics and behaviour are made by a computer and are without a human influence ('automatic profiling').

7.3 Our 'automatic decisions' or 'automatic profiling' can have a significant actual impact on your circumstances, behaviour or choices or on your rights or legal status.

7.4 The legal ground for our 'automatic decisions' or 'automatic profiling' about you is

7.4.1 your explicit consent to the 'automatic decisions' or 'automatic profiling' for specified purposes.

7.4.2 the need of the 'automatic decisions' or 'automatic profiling' for the entry into a contract or the performance of a contract between you and us.

7.4.3 the authorization of the 'automatic decisions' or 'automatic profiling' by Union or Member State law to which we are subject.

7.4.4 the need of the 'automatic decisions' or 'automatic profiling' for the reasons of a public interest that is substantial, on the basis of Union or Member State law.

7.4.5 the purposes of our legitimate interests or of a third party.

7.4.6 the following legal ground: XXX.

7.5 Your personal data on which our computer bases the 'automatic decisions' or 'automatic profiling'

7.5.1 are sensitive personal data.

7.5.2 are not sensitive personal data.

7.6 We base the 'automatic decisions' or 'automatic profiling' about you on the following reasoning:

7.6.1 XXX.

7.7 Our 'automatic decisions' or 'automatic profiling' may have consequences for you: XXX.

7.8 We have suitable measures that safeguard your rights, freedoms and legitimate interests against the 'automatic decisions' or 'automatic profiling' that produce a legal effect for you or that affect your circumstances, behaviour or choices significantly. We make it possible for you:

7.8.1 to activate a human intervention on our side of the 'automatic decisions' or 'automatic profiling'.

7.8.2 to express your point of view about our 'automatic decisions' or 'automatic profiling'.

7.8.3 to obtain an explanation of our 'automatic decisions' or 'automatic profiling'.

7.8.4 to challenge our 'automatic decisions' or 'automatic profiling'.

7.9 We make it possible for you to express your concerns about the safeguards that are related to 'automatic decisions' or 'automatic profiling' about you through our

7.9.1 postal address: XXX.

7.9.2 email address: XXX.

German translation (work in progress)

7 Führen wir automatisierte Entscheidung und Profilierung aus?

7.1 Wir verwenden Ihre personenbezogenen Daten um automatische Entscheidungen über Sie zu machen. Die automatischen Entscheidungen werden durch einen Computer gemacht und sie werden ohne menschlichen Einfluss gemacht („automatische Entscheidung“).

7.2 Wir verwenden Ihre personenbezogenen Daten um automatische Bewertungen über Ihre personenbezogenen Eigenschaften und Verhalten zu machen. Die automatischen Bewertungen dürfen Analyse von Ihren Eigenschaften oder von Prognosen Ihres Verhaltens beinhalten.

Die automatischen Bewertungen von Ihren Eigenschaften und Ihr Verhalten werden durch einen Computer und ohne menschlichen Einfluss gemacht („automatische Profilierung“).

7.3 Unsere „automatischen Entscheidungen“ oder „automatische Profilierung“ können einen bedeutenden aktuellen Einfluss auf Ihre Umstände, Ihr Verhalten, Ihre Wahlen, Ihre Rechte oder Ihren rechtlichen Status haben.

7.4 Der rechtliche Grund für unsere „automatischen Entscheidungen“ oder „automatische Profilierung“ über Sie ist

7.4.1 Ihre ausdrückliche Einwilligung zu den „automatischen Entscheidungen“ oder „automatische Profilierung“ für angegebene Zwecke.

7.4.2 Der Bedarf an die „automatischen Entscheidungen“ oder „automatische Profilierung“ für das Inkrafttreten in einen Vertrag oder die Leistung von einem Vertrag zwischen Ihnen und uns.

7.4.3 Die Genehmigung von den „automatischen Entscheidungen“ oder „automatische Profilierung“ durch Gesetz der Union oder eines Mitgliedstaats, zu dem wir betroffen sind.

7.4.4 Der Bedarf an die „automatischen Entscheidungen“ oder „automatische Profilierung“ für die Anlässe von einem hoheitlichen Interesse, das deutlich ist, auf der Basis von Gesetz der Union oder eines Mitgliedstaats.

7.4.5 Die Zwecke von unseren berechtigten Interessen oder von einer dritten Partei.

7.4.6 Der nachstehende rechtliche Grund: XXX.

7.5 Ihre personenbezogenen Daten, auf denen unser Computer die „automatischen Entscheidungen“ stützt, oder „automatische Profilierung“

7.5.1 Sind sensible personenbezogene Daten.

7.5.2 Sind nicht sensible personenbezogene Daten.

7.6 Wir stützen die „automatischen Entscheidungen“ oder „automatische Profilierung“ über Sie auf der nachstehenden Schlussfolge:

7.6.1 XXX.

7.7 Unsere „automatischen Entscheidungen“ oder „automatische Profilierung“ dürfen Auswirkungen für Sie haben: XXX.

7.8 Wir haben geeignete Maßnahmen, die Ihre Rechte, Freiheiten und berechtigte Interessen gegen die „automatischen Entscheidungen“ oder „automatische Profilierung“ aufrechterhalten, die eine rechtliche Auswirkung für Sie entfalten oder die Ihre Umstände beeinträchtigen, Verhalten oder Wahlen eindrucksvoll. Wir ermöglichen es für Sie:

7.8.1 Ein menschliches Eingreifen auf unserer Seite von den „automatischen Entscheidungen“ oder „automatischer Profilierung“ zu aktivieren.

7.8.2 Ihren Standpunkt über unsere „automatischen Entscheidungen“ oder „automatische Profilierung“ auszudrücken.

7.8.3 Eine Erläuterung von unseren „automatischen Entscheidungen“ oder von „automatischer Profilierung“ zu erhalten.

7.8.4 Unsere „automatischen Entscheidungen“ oder „automatische Profilierung“ anzufechten.

7.9 Wir ermöglichen es für Sie um Ihre Sorgen über die Garantien zu „automatischen Entscheidungen“, die dazugehörig sind, oder „automatische Profilierung“ über Sie auszudrücken

7.9.1 Postanschrift: XXX

7.9.2 E-Mail-Adresse: XXX

Italian translation (work in progress)

7 Facciamo processo decisionale automatizzato e profilazione?

7.1 Impieghiamo i Suoi dati personali per fare decisioni automatiche su Lei. Le decisioni automatiche si fanno da un computer e loro si fanno senza influenza umana (il «processo decisionale automatico»).

7.2 Impieghiamo i Suoi dati personali per fare valutazioni automatiche sulle Sue caratteristiche personali e comportamento.

Le valutazioni automatiche possono comprendere analisi delle Sue caratteristiche o previsioni del Suo comportamento.

Le valutazioni automatiche delle Sue caratteristiche ed il Suo comportamento si fanno da un computer e senza influenza umana (la «profilazione automatica»).

7.3 Le nostre «decisioni automatiche» o la «profilazione automatica» possono avere un'influenza sulle Sue circostanze attuale significativa, il Suo comportamento, le Sue scelte, i Suoi diritti o il Suo stato legale.

7.4 Il motivo legale per le nostre «decisioni automatiche» o la «profilazione automatica» su Lei è

7.4.1 Il Suo accettare esplicito alle «decisioni automatiche» o la «profilazione automatica» per fini specificati.

7.4.2 La necessità delle «decisioni automatiche» o la «profilazione automatica» per l'entrata in un contratto o l'esecuzione di un contratto fra Lei e noi.

7.4.3 L'autorizzazione delle «decisioni automatiche» o la «profilazione automatica» da legge dell'Unione o uno stato membro a cui siamo soggetti.

7.4.4 La necessità delle «decisioni automatiche» o la «profilazione automatica» per i motivi di un interesse che è considerevole pubblico, sul base di legge dell'Unione o uno stato membro.

7.4.5 I fini dei nostri interessi legittimi o di una parte terza.

7.4.6 Il motivo legale seguente: XXX.

7.5 I Suoi dati su cui il nostro computer basa le «decisioni automatiche» personali o la «profilazione automatica»

7.5.1 Sono dati personali sensibili.

7.5.2 Non sono dati personali sensibili.

7.6 Basiamo le «decisioni automatiche» o «profilazione automatica» su Lei sul ragionamento seguente:

7.6.1 XXX.

7.7 Le nostre «decisioni automatiche» o la «profilazione automatica» possono avere effetti per Lei: XXX.

7.8 Abbiamo misure che tutelano i Suoi diritti, libertà ed interessi legittimi contro le «decisioni automatiche» o «profilazione automatica» che producono un effetto legale per Lei o che riguardano le Sue circostanze adeguate, comportamento o scelte in modo significativo. Le consentiamo:

7.8.1 Attivare un intervento umano sul nostro lato delle «decisioni automatiche» o «profilazione automatica».

7.8.2 Esprimere la Sua opinione sulle nostre «decisioni automatiche» o «profilazione automatica».

7.8.3 Ottenere una spiegazione delle nostre «decisioni automatiche» o «profilazione automatica».

7.8.4 Impugnare le nostre «decisioni automatiche» o «profilazione automatica».

7.9 Le consentiamo per esprimere i Suoi problemi sulle garanzie che sono appartenenti a «decisioni automatiche» o «profilazione automatica» su Lei

7.9.1 indirizzo postale: XXX

7.9.2 indirizzo di posta elettronica: XXX

Rationale: reliable translation

The state of the art in machine translation

English Swedish French Detect language ▾ ↔ Translate

Min mor är inte svensk. × 我的母亲是瑞典的。
Min mor är svensk. 我的母亲是瑞典的。

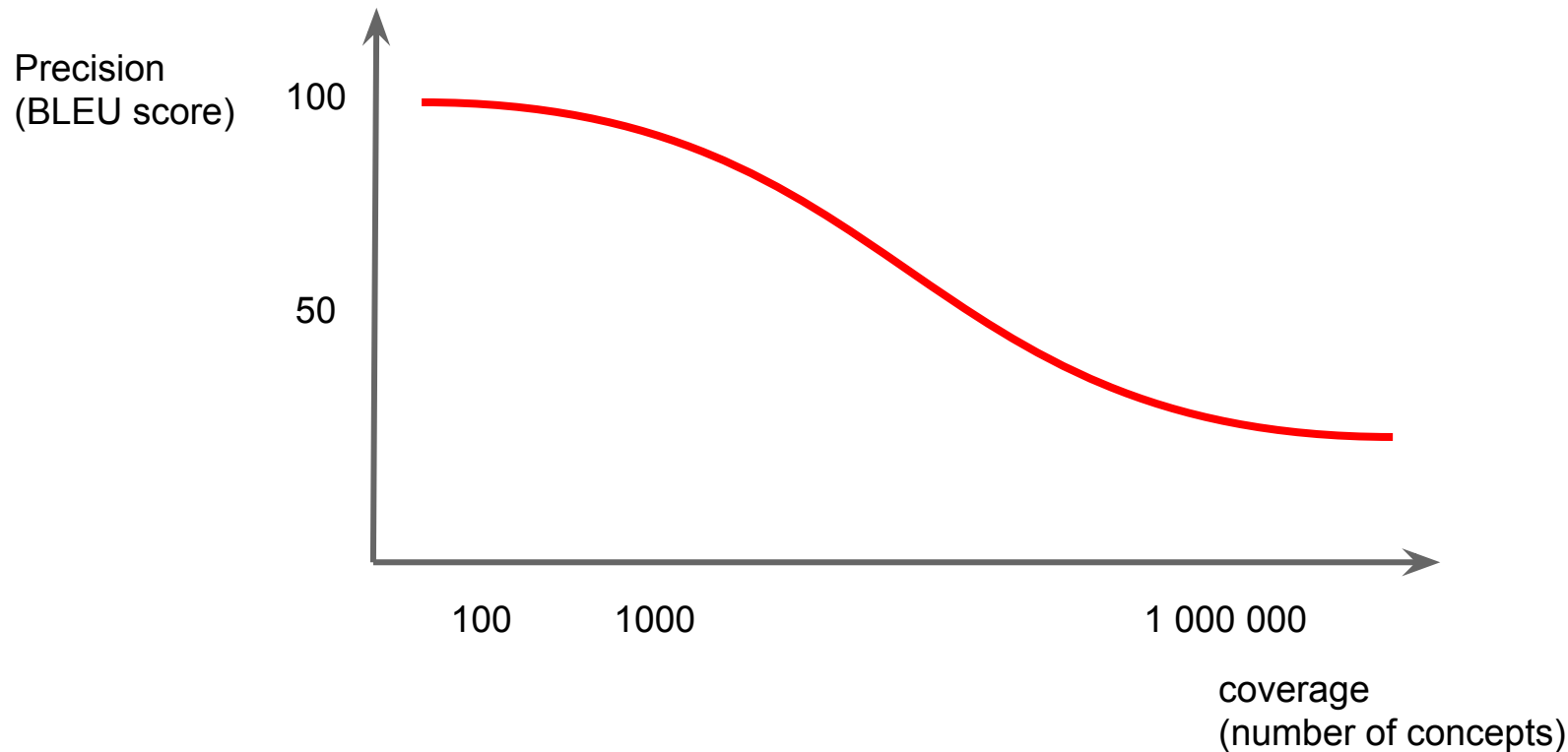
English Swedish French Detect language ▾ ↔ German English Norwegian ▾

Min mor är inte svensk. × Min mor er svensk.
Min mor är svensk. Min mor er svensk.

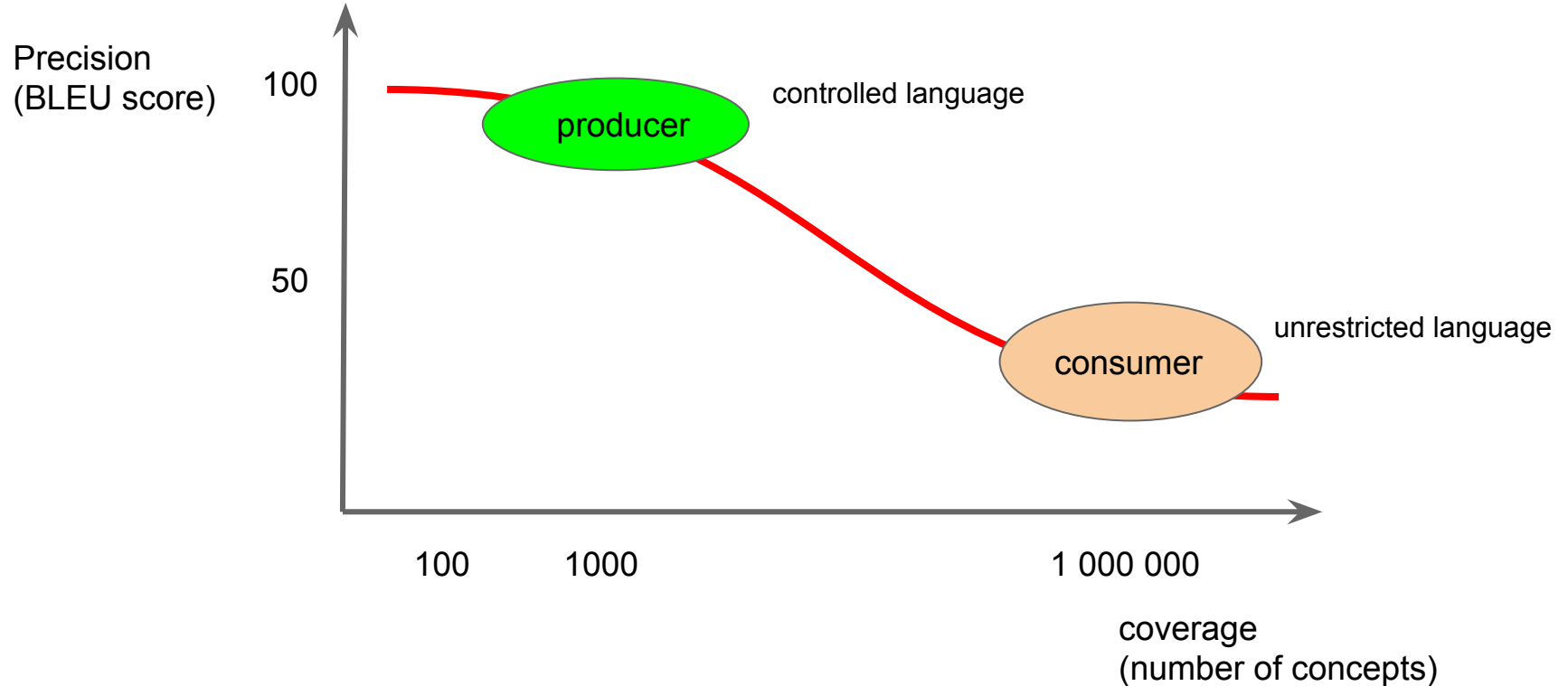
English Finnish Swedish Detect language ▾ ↔ Spanish Persian English ▾ Translate

Hon är en professor i akutvård. × She is an acupuncture professor.

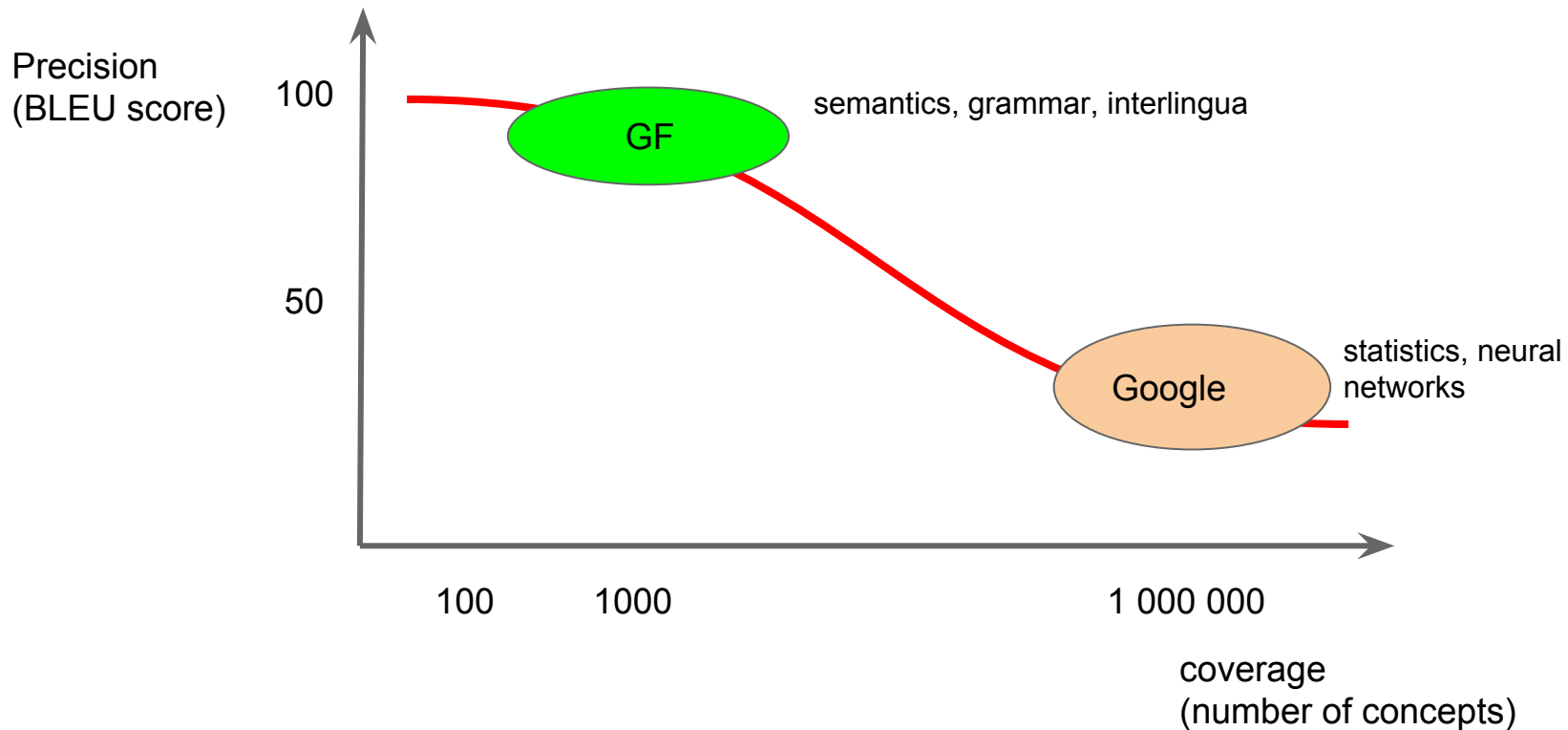
Precision vs. coverage



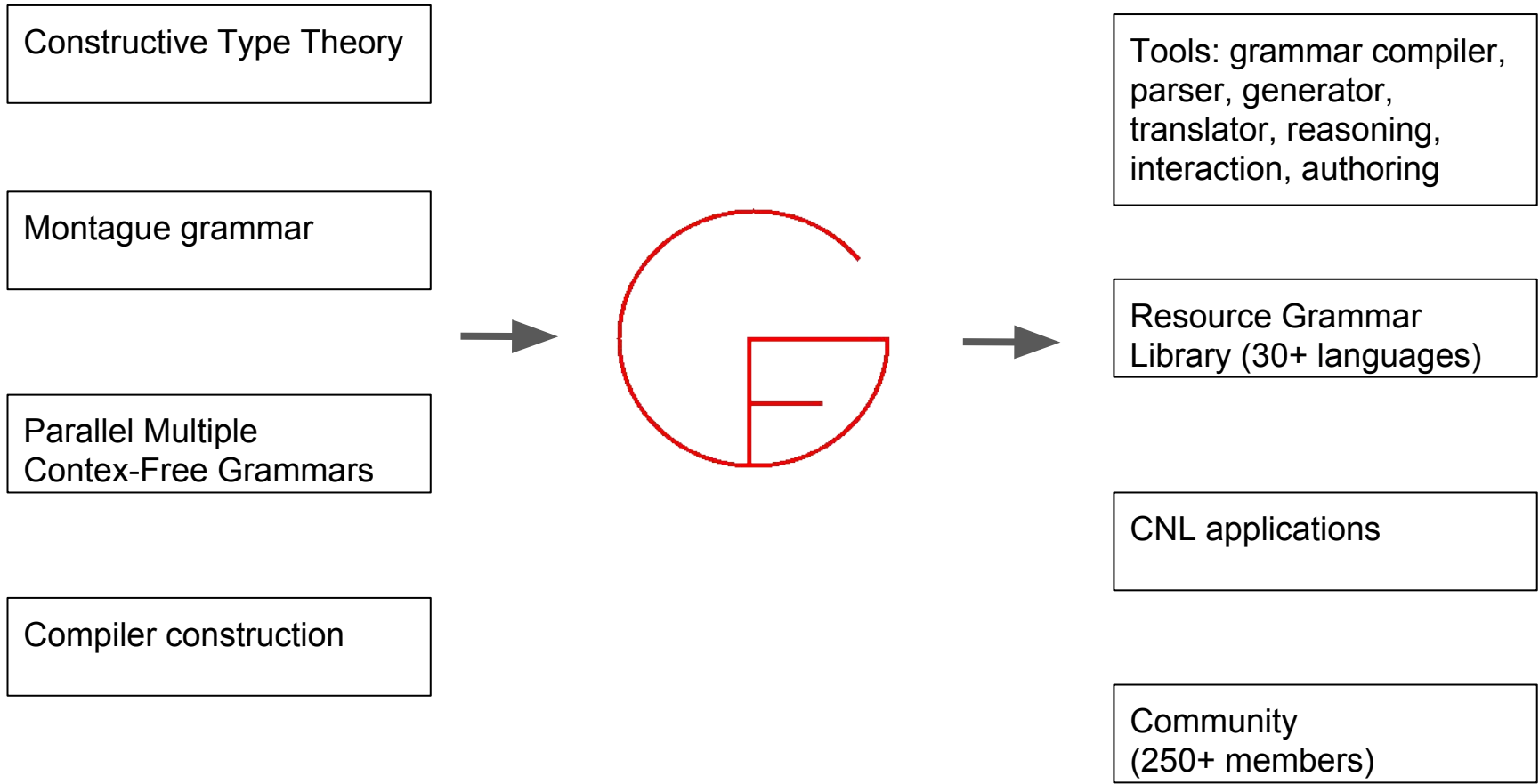
Precision vs. coverage

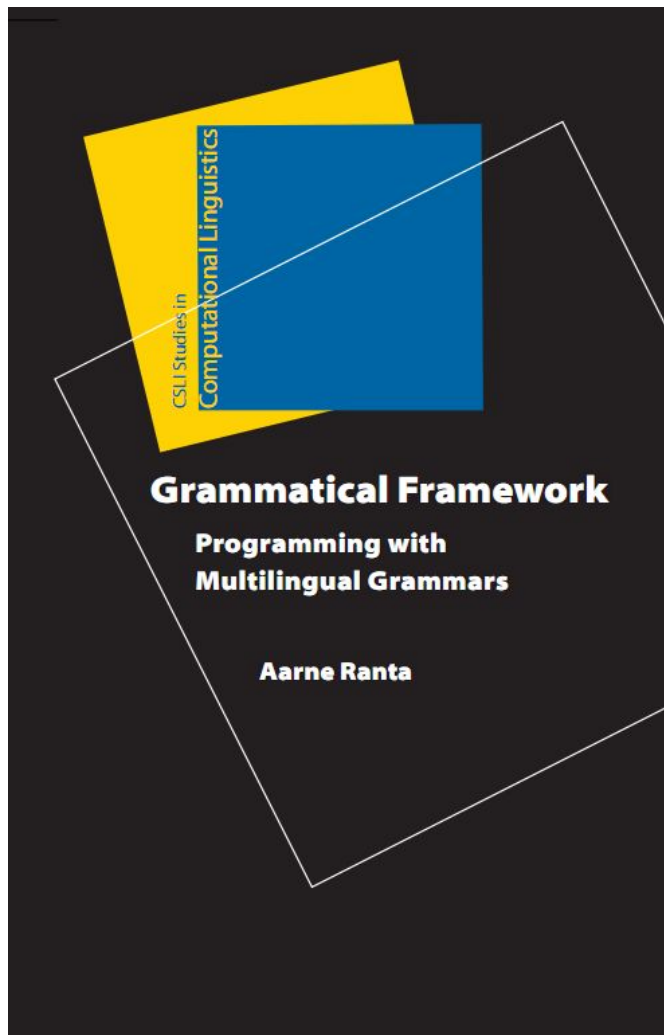


Precision vs. coverage



The role of GF





A. Ranta. *Grammatical Framework: Programming with Multilingual Grammars*, CSLI, Stanford, 2011. Chinese translation by Prof. Yan Tian: *语法框架 为多种自然语言语法编程*, Shanghai Jiao Tong University Press, 2014.

The GF Resource Grammar Library

Morphology: complete paradigms

Basic syntax: 80 categories, 200 functions

Function word lexicon: 100 lemmas

Content word test lexicon: 300 lemmas

32 languages: Afrikaans, Bulgarian, Catalan, Chinese (simplified), Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hindi, Icelandic, Japanese, Italian, Latvian, Maltese, Mongolian, Nepali, Norwegian (bokmål), Norwegian (nynorsk), Persian, Polish, Punjabi, Romanian, Russian, Sindhi, Spanish, Swedish, Thai, Urdu

GF grammar building effort



abstract syntax: weeks

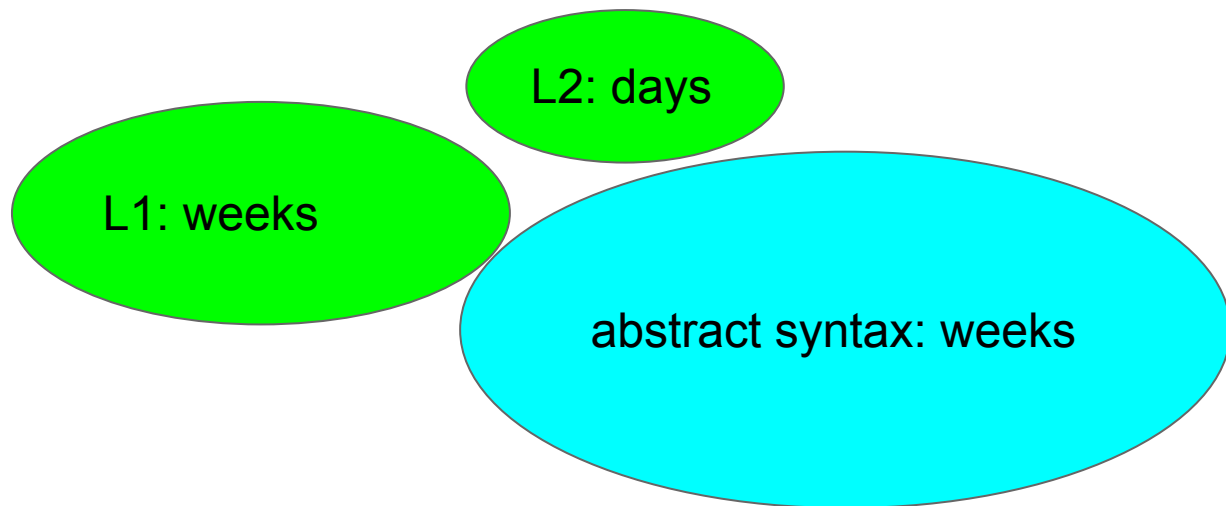
GF grammar building effort



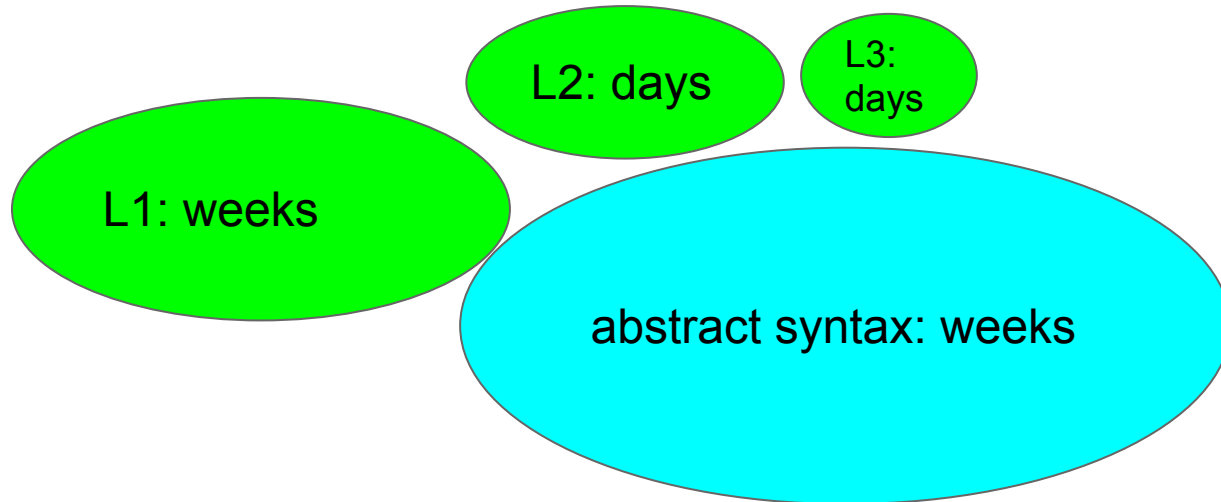
L1: weeks

abstract syntax: weeks

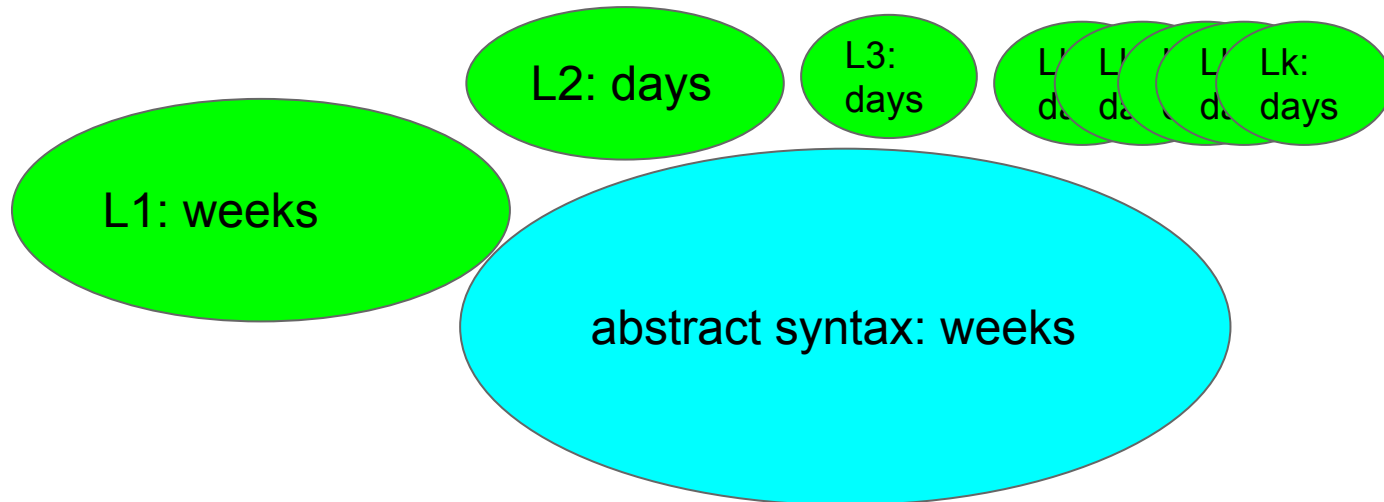
GF grammar building effort



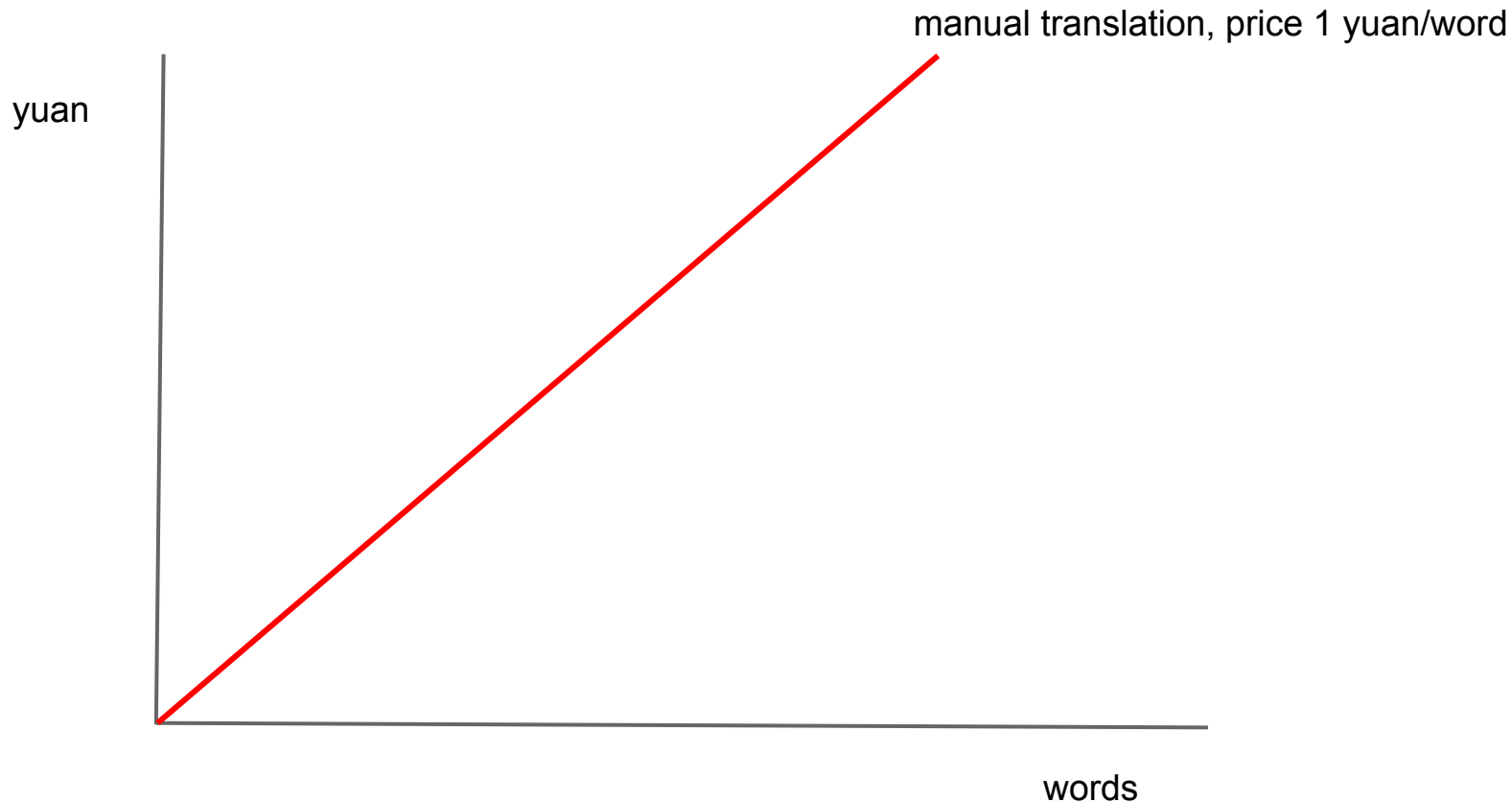
GF grammar building effort



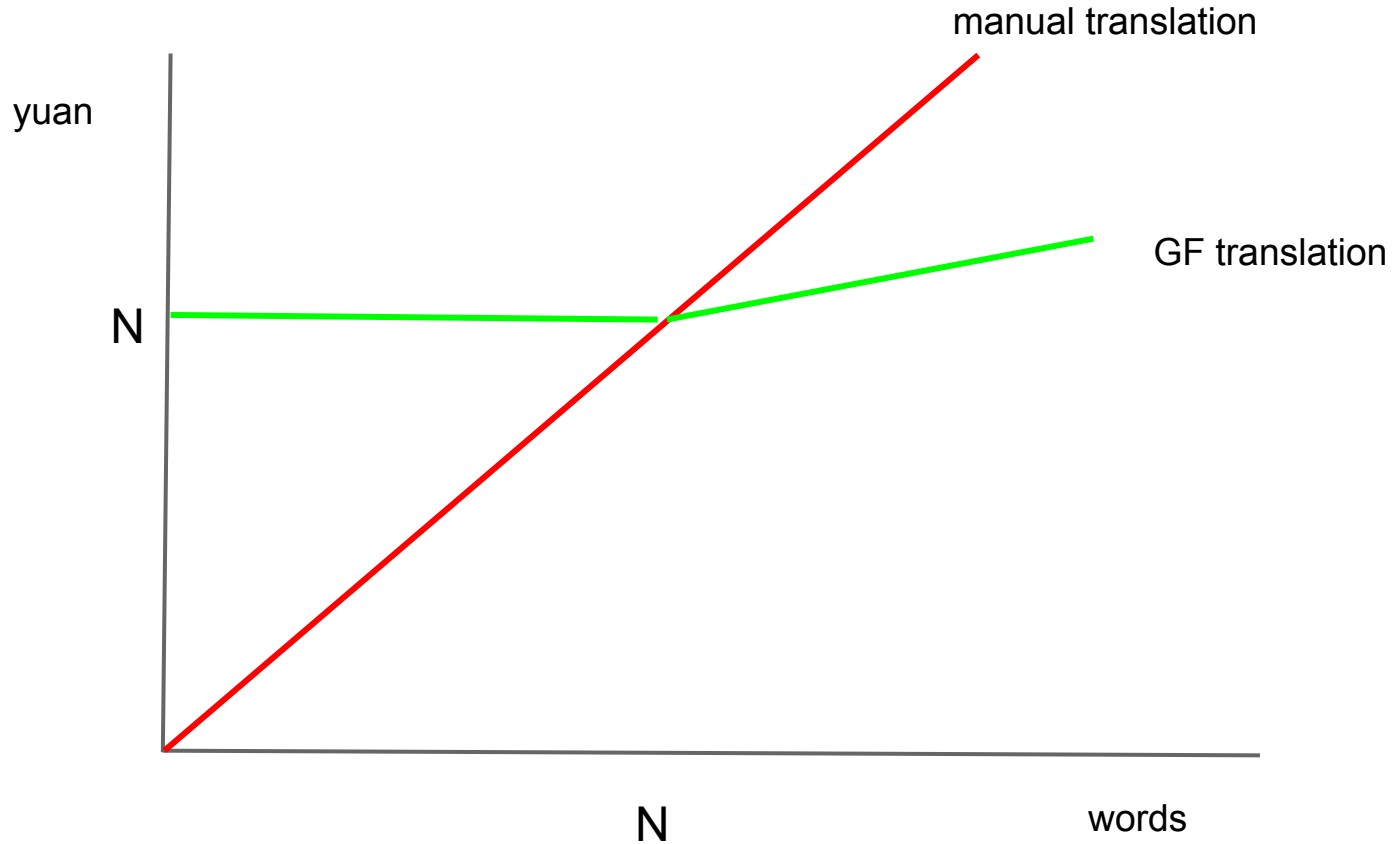
GF grammar building effort



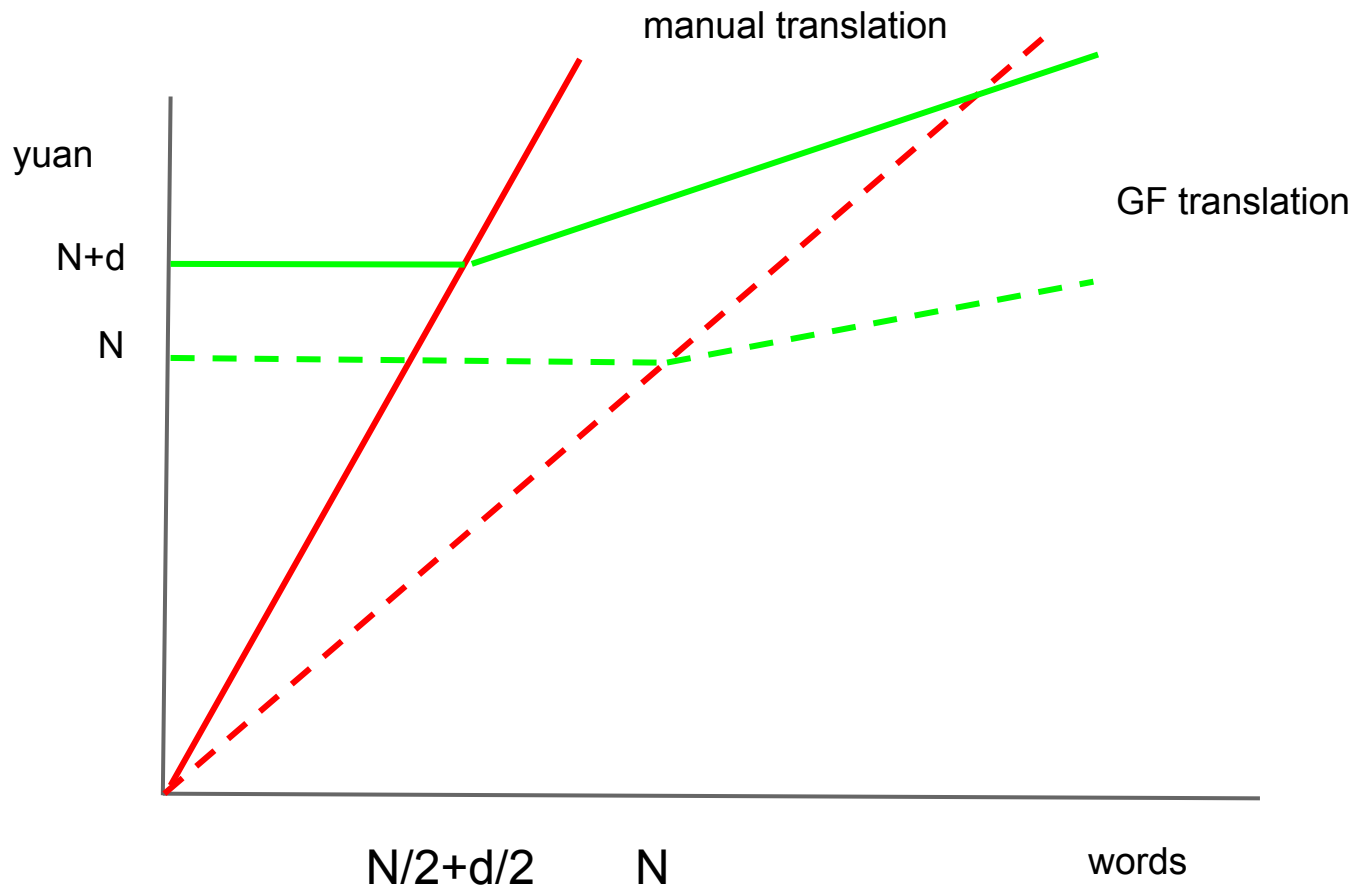
Price of translation, 1 target language



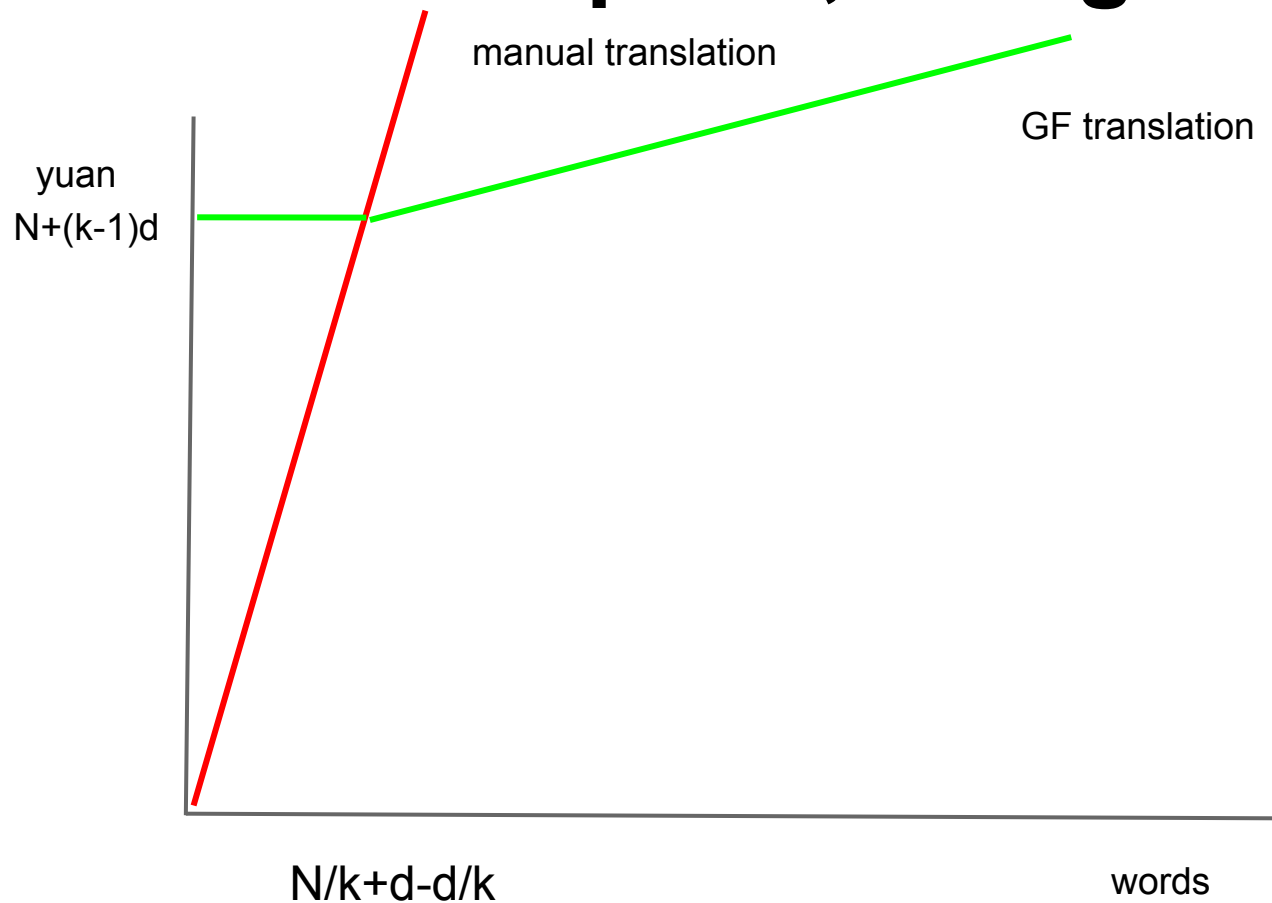
Break-even point, 1 target language



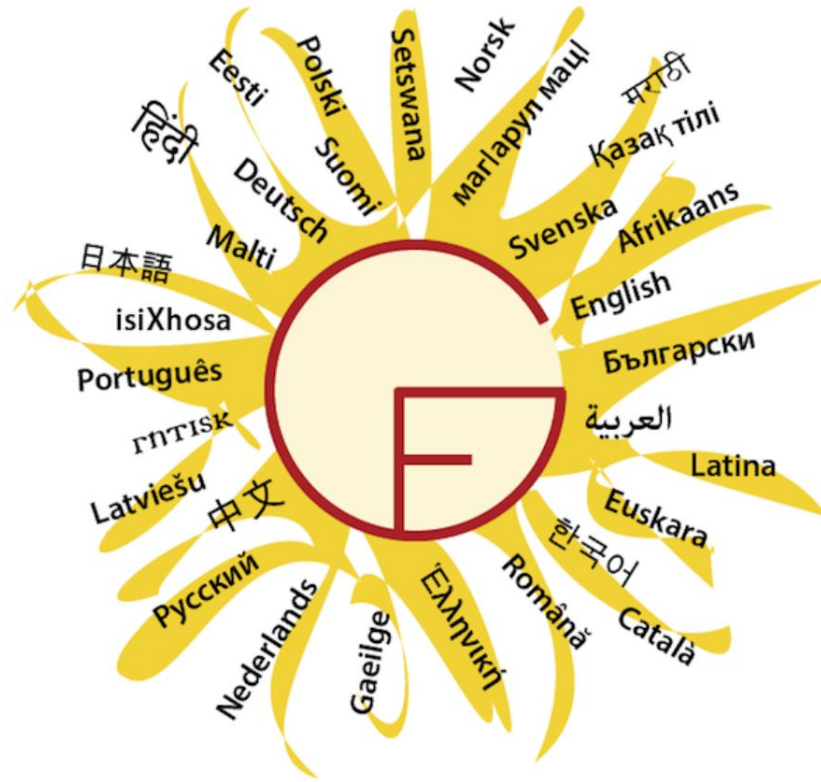
Break-even point, 2 target languages



Break-even point, k target languages



Sixth GF Summer School 2018



Stellenbosch, South Africa

3rd–14th December 2018

Some CNLs built using GF

AccessibilityDB
5 x 1200

Textual 8 x 250

GDPR 5 x 3500

Ambulance 7 x 350

Inlife 5 x 1800

Phrasebook 24 x 300

MathBar 15 x 250

AceWiki 25 x 150

Foods 35 x 30

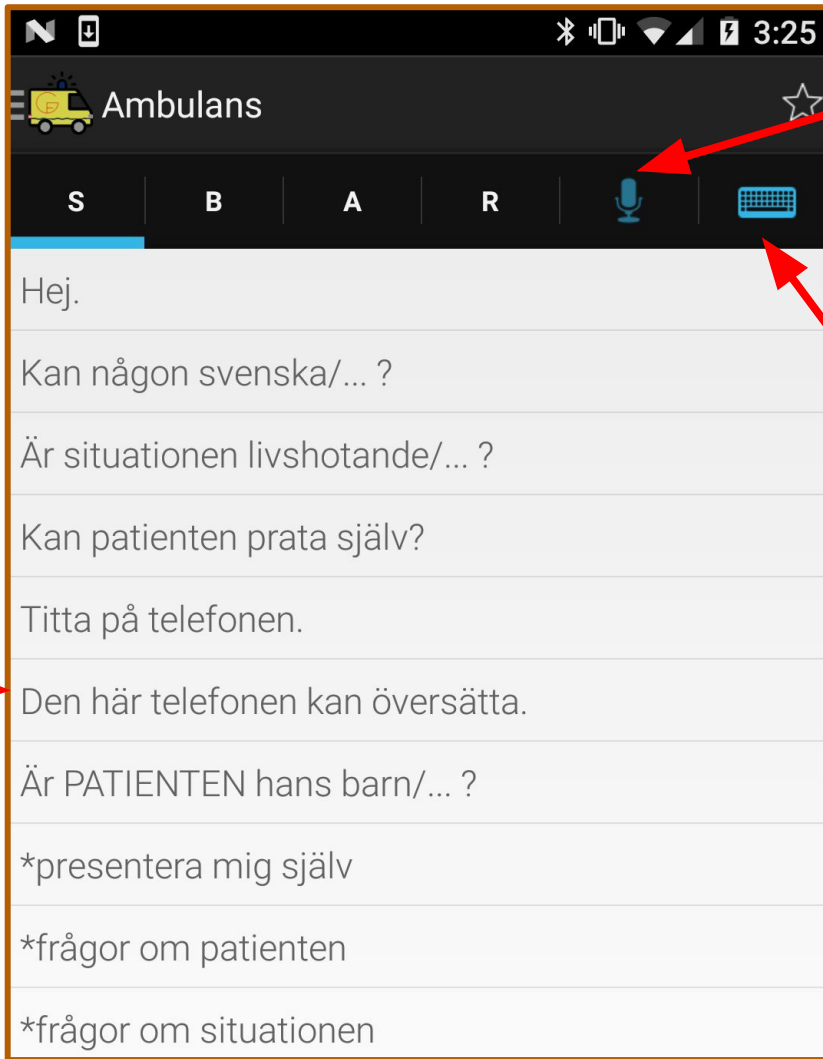
Numerals 88 x 23

Attempto 8 x 240

Mobile app and its 3 use modes

Menu

build phrases
interactively



Microphone

speak

Keyboard

write

Conclusions

Take home

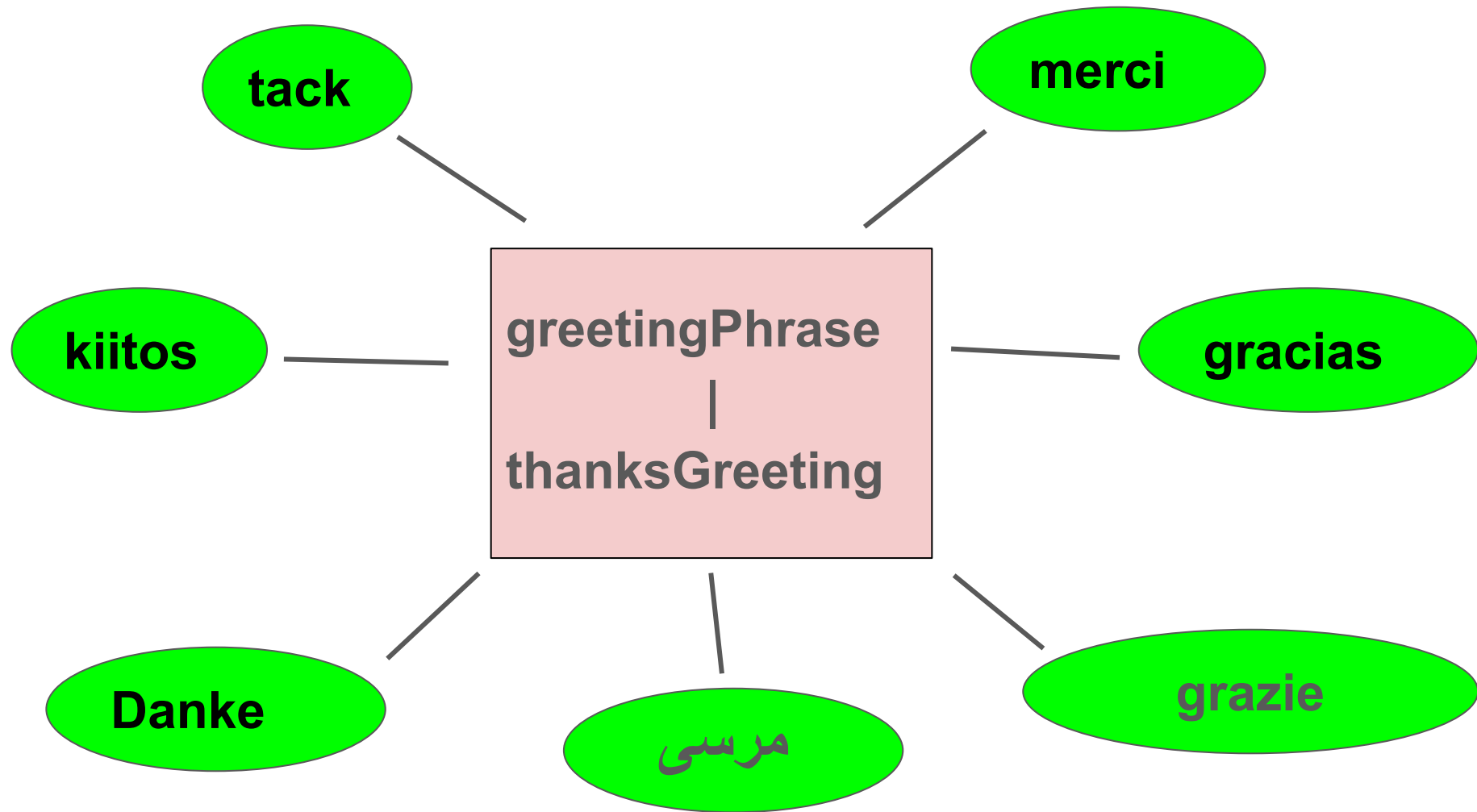
Multilingual CNL = CNL based on shared abstract syntax

GF = Grammatical Framework, designed for building multilingual CNLs

- 30+ languages
- 250+ developers
- 15+ companies

Digital Grammars AB

- started by the core developers of GF to build commercial CNL systems
- health care: mobile translation app for ambulance staff
- legal translation and analysis: GDPR lexicon, contract reasoning



tack

merci

greetingPhrase

thanksGreeting

gracias

grazie

مرسی

Danke

kiitos