Grammatical Framework Overview

Aarne Ranta

GF Summer School 2015, Marsalforn

Mission

The mission of GF is to formalize the grammars of the world and make them usable in computer applications.

Grammar

morphology

syntax

semantics

pragmatics

The world

6000 languages in the world

30 languages in the **GF Resource Grammar Library**

Computer applications

translation

parsing, generation

query systems

dialogue systems

language learning

Usable

executable

programmable

available

Executable

Linux, Mac, Windows

Android, iOS

web services

the GF cloud

Programmable

the GF programming language

the PGF "machine language"

grammar compiler

standard libraries

Available

the GF web page

open source licenses: GPL, LGPL, BSD

commercial uses allowed

Rationale

A popular idea

use statistics and machine learning to eliminate human work

The GF idea

develop programming methods to make human work easier

Example: resource grammar implementation

from years to months

from PhD to Masters level

from professional linguist to computer scientist, mathematician, etc

from native speakers to language learners

Application implementation

days for the first language

hours for additional languages

portable across languages

programmer + native informant

Method: functional programming

instead of copy and paste, write a function!

static type checking: grammar errors become type errors

modules: reusable parts of grammars

functors: functions on modules

Uses

function types: library API

information hiding

collaborative grammar development

Method: multilingual grammars

abstract syntax + concrete syntaxes

type-theoretical structure + compositional mappings

Uses

translation interlingua

shared semantics

interface to backend applications

cross-lingual abstractions

Community

The greatest asset

knowledge of the languages

contributions theory and technology

ideas for new applications

Some figures

170 members of GF Developer list

50 contributors to Resource Grammar Library

40 participants in each Summer School

Our commitments

open-source availability

backward compatibility

help with problem solving

The summer schools

2009 Gothenburg

scale up from 10 languages

establish a community

2011 Barcelona

scale up from 15 languages

MOLTO: EU project on domain-specific precision translation

hybrid GF + statistical translation

CNL (Controlled Natural Language)

2013 Fraueninsel

scale up from 26 languages

funded by Volkswagen Foundation

one result: wide-coverage mobile translation

2015 Marsalforn

scale up from 30 languages

improve existing language resources

improve wide-coverage translation

develop applications

develop theory and technology

Challenges

Multilingual lexicon

66,000 ”word senses” in English

partial mappings to 14 languages

automatic from Wiktionary, Wordnet, etc

manual checking

sense distinctions?

multiwords?

Information sharing

finite-state morphology

Apertium

lexicon as database?

Machine translation quality

proper evaluation and comparison to do

focus on some languages as pilots?

focus on some application areas?

Machine translation productivity

we can add a new language in a few hours

how does this compare with SMT?

what about if there is no RGL?

controlled experiment à la Chanod and Tapanainen?

Language learning

learning by translation

exercise generation + assessment

grammar-based explanation

grammar-based language documentation

Query languages

given: YAQL + partial SPARQL mapping

sought: good coverage of SQL

Semantics and inference

text analysis based on abstract syntax trees

Interaction

dialogue systems

robotics

Industrial applications

MOLTO: Ontotext, BeInformed

Galois

Lingsoft

Talkamatic

Digital Grammars

Resources

the GF book

the "official" GF tutorial

the resource grammar tutorial (LREC 2010)

wanted: GF book Vol. 2