

# Exercicis de matemàtiques multilingües usant la plataforma *gf*

Jordi Saludes

1 de Juny de 2006

## Resum

Un projecte europeu té per objecte construir un dipòsit de problemes de matemàtiques de primer curs d'universitat que puguin ser traduïts automàticament a totes les llengües europees. Una peça clau del projecte (la generació de llenguatge natural a partir de la representació semàntica del problema) es basa en programari lliure que procedeix de la recerca (*Grammatical Framework*). L'objecte d'aquesta comunicació és presentar aquesta eina lliure per a les tasques de manipulació del llenguatge natural.

## 1 Introducció

*Web Advanced Learning Technologies* [2] és un projecte finançat per la UE dins el programa *eContent* que té com aspecte principal desenvolupar un dipòsit d'exercicis de matemàtiques per a estudiants de primer any d'universitat (Àlgebra lineal i Càlcul) que es puguin generar en qualsevol llengua de la unió (d'entrada només 7 llengües).

Val a dir que un dels objectius del projecte consisteix en engegar una empresa que explotará aquest dipòsit i per tant serà l'encarregada de generar les versions en diferents llengües. Per donar a suport a aquesta tasca s'està desenvolupant una biblioteca de recursos lingüístics. L'eina en que s'està desenvolupant i que genera el llenguatge natural pot ser d'interès a la comunitat de programari lliure i passo a descriure-la a continuació.

## 2 L'entorn *gf*

*Grammatic Framework* [3] és un formalisme per a gramàtiques multilingües amb aplicacions a la traducció i la localització. El sistema estableix un sistema de generació de text multilingüe que és reversible. El paquet (amb llicència GPL) està format per

- Un llenguatge de programació específic

- Un compilador per aquest llenguatge
- Un processador de gramàtiques genèric.

El compilador llegeix els fitxers de gramàtiques de l'usuari i permet que el processador de gramàtiques faci el següent:

- generació de text
- *parsing* (que és l'operació inversa)
- traducció de text d'una llengua a una altra.
- etc.

En particular *gf* s'adreça l'aspecte multi-lingüístic (generació paral·lela del mateix text en llengües diferents), permet el desenvolupament modular i la re-utilització de codi. *Gf* està pensat per a modelar el llenguatge en un domini de discurs ben delimitat i és modular per tal de facilitar la divisió del treball: Els experts en el domini que es vol modelar escriuen la *gramàtica d'aplicació* mentre que els lingüistes escriuen les *gramàtiques de recursos* que connecten la primera amb el llenguatge natural.

## 2.1 Gramàtiques de recursos

Es tracta de regles de combinació per a construir text ben format en una llengua determinada, tenint en compte tots els aspectes lingüístics com morfologia, concordança, etc.

Quan parlem de la biblioteca de recursos de *gf* estem parlant d'una col·lecció de fitxers de gramàtiques que capturen les regles de formació de cada categorial gramatical per a cada llengua, fent abstracció de les característiques comunes. Així, per exemple, hi ha fitxers que modelen els aspectes generals de les llengües romàniques i que després són refinats pels fitxers de gramàtiques de cada llengua particular.

## 2.2 Gramàtiques d'aplicació

El sistema permet la separació funcional dels mòduls d'aplicació per tal de separar-los dels mòduls més lingüístics de les gramàtiques de recursos. D'aquesta manera una gramàtica d'aplicació pot ser escrita sense necessitat de conèixer els detalls lingüístics que hi ha al nivell inferior: Qui escriu una gramàtica d'aplicació ha de ser un expert en el domini de discurs que es vol capturar però no cal que conegui els detalls dels nivells inferiors.

Així una gramàtica d'aplicació per escriure matemàtiques (que és el nostre cas) tindria produccions per a objectes matemàtics i per a maneres de combinar aquests objectes (operacions). La implementació d'aquestes produccions es faria usant les produccions lingüístiques que es troben a les gramàtiques de recursos de cada llengua.

## 2.3 Exemple

Per exemple, la gramàtica *peacekeeping* que ve amb la distribució del *gf* està inspirada en llibres de frases per a cooperants de primers auxilis i emergències. Anem a usar-la dins la *shell* de *gf*.

Per traduir de l'anglès al finès usem la comanda:

```
> t Peace_Eng Peace_Fin "Where do you see the
landmine?"
Missä te näette maamiinan?
Missä sinä näet maamiinan?
```

Per a traduir de l'anglès al suec només hem canviar el llenguatge d'arribada:

```
> t Peace_Eng Peace_Swe "Where do you see the
landmine?"
Var ser ni landminan?
Var ser du landminan?
```

Veiem que hi ha dues possibles traduccions. Per tal d'apreciar la diferència entre les dues, fem separatament cada un dels dos passos en que consisteix l'operació de traducció: El primer pas (*parsing*) converteix el text anglès en la seva representació abstracta:

```
> p -lang=Peace_Eng "Where do you see the landmine?"
Written (PhrQuest (QuestIAdv_V2 see_V2 (UsePron youPl_Pron)
(DetCN theSg_Det (UseN landmine_N)) where.IAdv))
Written (PhrQuest (QuestIAdv_V2 see_V2 (UsePron youSg_Pron)
(DetCN theSg_Det (UseN landmine_N)) where.IAdv))
```

Hem obtingut no una sinó dues representacions abstractes en forma d'arbre. Aquests dos arbres es diferencien només en el node `UsePron` que en un cas és `UsePron youPl_Pron` i en l'altre és `UsePron youSg_Pron`. La llengua anglesa no diferencia entre singular i plural pel pronom "you". Això explica que hi hagin dos arbres amb significats diferents.

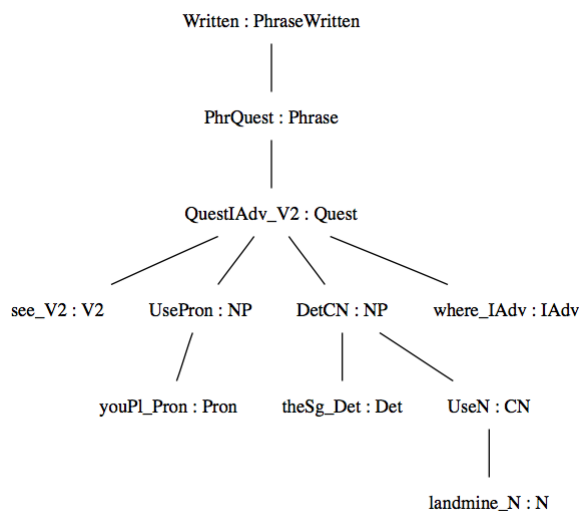
L'estructura de l'arbre ens diu que és una frase interrogativa amb el verb transitiu "veure" (`see_V2`) i un subjecte pronominal ("tu" o "vosaltres"), on es demana pel lloc referit a un sintagma nominal (NP: "la mina").

Apliquem el segon pas de la traducció (linealització) per a convertir la representació abstracta al suec, i obtenim:

```
> l -lang=Peace_Swe Written (PhrQuest (QuestIAdv_V2
see_V2 (UsePron youSg_Pron) (DetCN theSg_Det (UseN
landmine_N)) where_IAdv))
Var ser du landminan?
```

Fem el mateix però al finès:

Figura 1: Estructura de l'arbre abstracte per a: *Where do you see the landmine?*



```

> 1 -lang=Peace_Fin Written (PhrQuest (QuestIAdv_V2
see_V2 (UsePron youSg_Pron) (DetCN theSg_Det (UseN
landmine_N)) where_IAdv))
Missä sinä näet maamiinan?
  
```

### 3 webALT i gf

Com he dit al començament, la part multilingüe del projecte webALT consisteix en la generació de problemes de matemàtiques de primer curs de carrera en diferents llengües a partir de la representació abstracta de la semàntica. Des d'aquest punt de vista es tracta d'operacions de linealització en cada una de les set llengües a les que s'adreça el projecte (anglès, finès, suec, francès, espanyol, italià i neerlandès). Les nostres gramàtiques d'aplicacions parlen d'objectes i operacions matemàtiques. L'estàndard *OpenMath* [1] pretén recollir tots els objectes i operacions matemàtiques i donar-ne una codificació uniforme (basada en xml). Cada símbol (objecte o operació) ve acompanyat d'una descripció i d'una signatura en un sistema de tipus (*Small Type System*) que hem seguit per tal de definir les categories de la nostra gramàtica d'aplicacions. Les gramàtiques estan dividides en dues capes

**OpenMath** Per a símbols matemàtics: determinant, producte, número 3,...

**Operations** Per a accions que es demana a l'estudiant de fer (i que no estan recollides a l'estàndard OpenMath) com ara: calcula, troba, determina, ...

Per a introduir nous problemes al dipòsit, s'ha escrit un editor predictiu que, conduit per la gramàtica corresponent, permet a un autor compondre nous problemes en la pròpia llengua i obtenir-ne la representació abstracta. Aquest editor utilitza un nou *top-down parser* que ha estat desenvolupat inicialment com a programari privatiu però que finalment s'afegirà a la distribució de *gf* com a contribució lliure en retorn dels beneficis que el projecte ha obtingut per l'ús de *gf*.

## 4 Conclusions

S'han exposat les característiques d'un programari que pot ser d'interés general per a tasques de traducció o localització de programari i la manera en que s'usa per obtenir exercicis de matemàtiques en diverses llengües.

## Referències

- [1] OpenMath <http://openmath.activemath.org/>
- [2] webALT <http://www.webalt.net>
- [3] Grammatical Framework <http://www.cs.chalmers.se/~aarne/GF/>