

Servicios Lingüísticos en Python basados en reglas

Jordi Saludes, UPC

Demo

- proyecto

MOLTO

- Conexión un *sistema de álgebra por ordenador* (CAS)

➡ Sage



Traducción automática

- Estadístico
- Basado en reglas

SMT: *Statistical Machine Translation*

SMT: *Statistical Machine Translation*

- Predominante

SMT: *Statistical Machine Translation*

- Predominante

Pro

SMT: *Statistical Machine Translation*

- Predominante

Pro

 Extensión

SMT: *Statistical Machine Translation*

- Predominante

Pro

☒ Extensión

☒ Robustez

SMT: *Statistical Machine Translation*

- Predominante

Pro

☒ Extensión

☒ Robustez

Contra

SMT: *Statistical Machine Translation*

- Predominante

Pro

☒ Extensión

☒ Robustez

Contra

☐ Poca precisión

SMT: *Statistical Machine Translation*

- Predominante

Pro

☒ Extensión

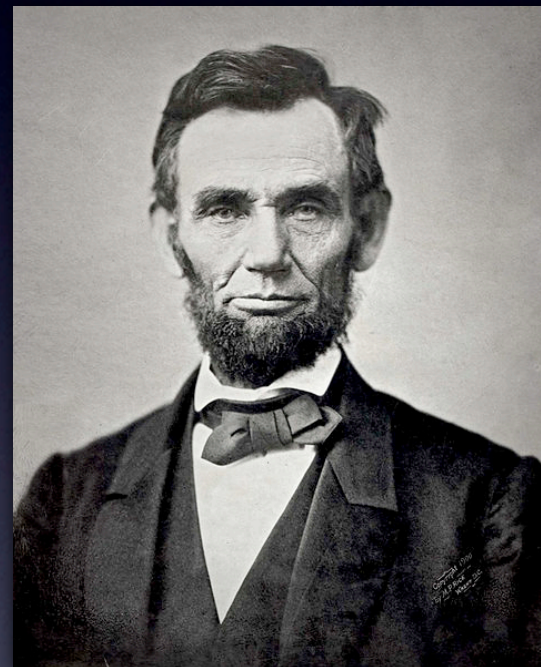
☒ Robustez

Contra

☐ Poca precisión

☐ Lenguas minoritarias

Wikipedia

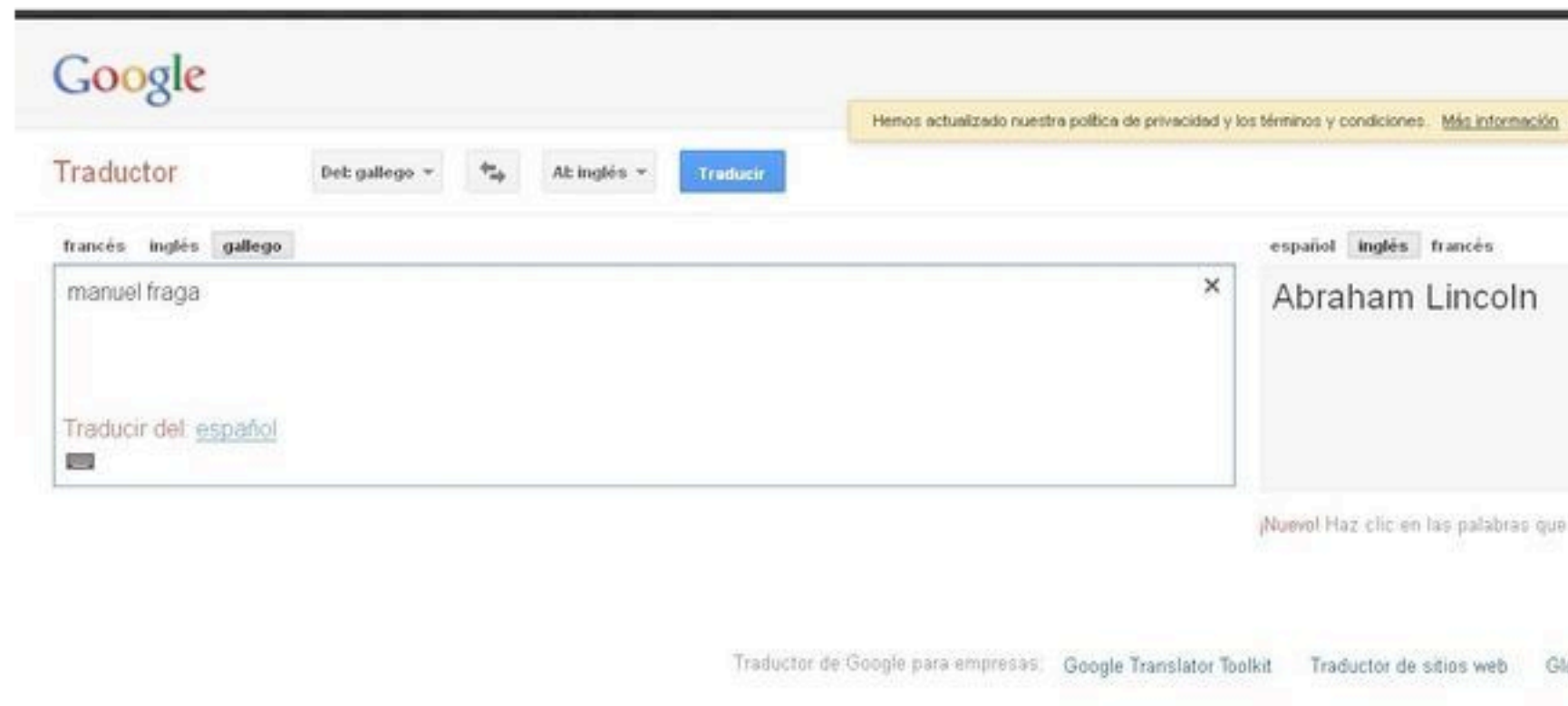


ESPAÑA

El traductor de Google confunde a Manuel Fraga con Abraham Lincoln

E. M.  / MADRID | Día 22/02/2012 - 19.09h

- Si escribes el nombre del fundador del PP y traduces del gallego al inglés, el resultado es el del que fuera presidente de EE.UU., pero «el Manuel Fraga» en español es también el sajón



ABC, 22/2/2012

MENU 16.6.11

PRIMERS

Empedrat de lletlies i cigrons
Amanida verda
Tortellini napolitana

SEGONS

Culxa de pollastre a l'ast
Llom a la graella
Mero al forn amb verdures

Postre

FIRST

Pavement of lentils and chickpeas	
green salad	
Tortellini Neapolitan	

ACCORDING

Roast Leg of chicken ||||
Grilled Loin ||
Baked Grouper with vegetables |||||

Dessert

MENÚ 16.6.11

PRIMERS

Empedrat de llenties i cigrons
Amanida verda
Tortellini napolitana

SEGONS

Guixa de pollastre a l'ast
Llom a la graella
Mero al forn amb verdures

Postre

FIRST

Pavement of lentils and chickpeas || || || ||
green salad || || || || || ||
Tortellini Neapolitan || ||

ACCORDING

Roast Leg of chicken (|| || ||
Grilled Loin || ||
Baked Grouper with vegetables || || || || || || || || || ||

Dessert

Google translate

Google translate

Google Traductor

De: alemany ▾



↔

A: català ▾

Tradueix



anglès castellà **alemany**

selig sind, die da Leid tragen; denn sie sollen getröstet

català anglès

Feliços els qui plorar, perquè ells són consolats

Nou! Feu clic a les paraules que hi ha més amunt per visualitzar-ne traduccions alternatives. [Omet](#)

Google translate

Google Traductor

De: alemany ▾



↔

A: català ▾

Tradueix



anglès castellà **alemany**

selig sind, die da Leid tragen; denn sie sollen getröstet

català anglès

Feliços els qui plorar, perquè ells són consolats

Nou! Feu clic a les paraules que hi ha més amunt per visualitzar-ne traduccions alternatives. [Omet](#)

Google translate

Google Traductor

De: alemany ▼

↔

A: català ▼

Tradueix

anglès castellà **alemany**

selig sind, die da Leid tragen; denn sie sollen getröstet

×

⌨

🔊

català anglès

Feliços els qui plorar, perquè ells són consolats

🔊 ✓

Nou! Feu clic a les paraules que hi ha més amunt per visualitzar-ne traduccions alternatives. [Omet](#)

Google Traductor

De: alemany ▼

↔

A: català ▼

Tradueix

anglès castellà **alemany**

selig sind, die da Leid tragen; denn sie sollen getröstet werden

×

⌨

🔊

català anglès

Feliços els qui plorar, Déu els consolarà

🔊 ✓

Nou! Feu clic a les paraules que hi ha més amunt per visualitzar-ne traduccions alternatives. [Omet](#)

Marzo 2012

Google translate

Google Traductor

De: alemany ▼

↔

A: català ▼

Tradueix

anglès castellà **alemany**

selig sind, die da Leid tragen; denn sie sollen getröstet

×

⌨

🔊

català anglès

Feliços els qui plorar, perquè ells són consolats

🔊 ✓

Nou! Feu clic a les paraules que hi ha més amunt per visualitzar-ne traduccions alternatives. [Omet](#)

Google Traductor

De: alemany ▼

↔

A: català ▼

Tradueix

anglès castellà **alemany**

selig sind, die da Leid tragen; denn sie sollen getröstet werden

×

⌨

🔊

català anglès

Feliços els qui plorar, Déu els consolarà

🔊 ✓

Nou! Feu clic a les paraules que hi ha més amunt per visualitzar-ne traduccions alternatives. [Omet](#)

Marzo 2012

Google translate

Google Traductor

De: alemany ▼

↔

A: català ▼

Tradueix

anglès castellà **alemany**

selig sind, die da Leid tragen; denn sie sollen getröstet

×

⌨

🔊

català anglès

Feliços els qui plorar, perquè ells són consolats

🔊 ✓

Nou! Feu clic a les paraules que hi ha més amunt per visualitzar-ne traduccions alternatives. [Omet](#)

Google Traductor

De: alemany ▼

↔

A: català ▼

Tradueix

anglès castellà **alemany**

selig sind, die da Leid tragen; denn sie sollen getröstet werden

×

⌨

🔊

català anglès

Feliços els qui plorar, Déu els consolarà

🔊 ✓

Nou! Feu clic a les paraules que hi ha més amunt per visualitzar-ne traduccions alternatives. [Omet](#)

Marzo 2012

RB: *Rule-based*

RB: *Rule-based*

- Gramáticas

RB: *Rule-based*

- Gramáticas

Pro

RB: *Rule-based*

- Gramáticas

Pro

 Precisión

RB: *Rule-based*

- Gramáticas

Pro

☒ Precisión

☒ Uniformidad

RB: *Rule-based*

- Gramáticas

Pro

- ☒ Precisión
- ☒ Uniformidad
- ☒ Representación abstracta

RB: *Rule-based*

- Gramáticas

Pro

- ☒ Precisión
- ☒ Uniformidad
- ☒ Representación abstracta

Contra

RB: *Rule-based*

- Gramáticas

Pro

- ☒ Precisión
- ☒ Uniformidad
- ☒ Representación abstracta

Contra

- ☐ Lenguaje controlado

RB: *Rule-based*

- Gramáticas

Pro

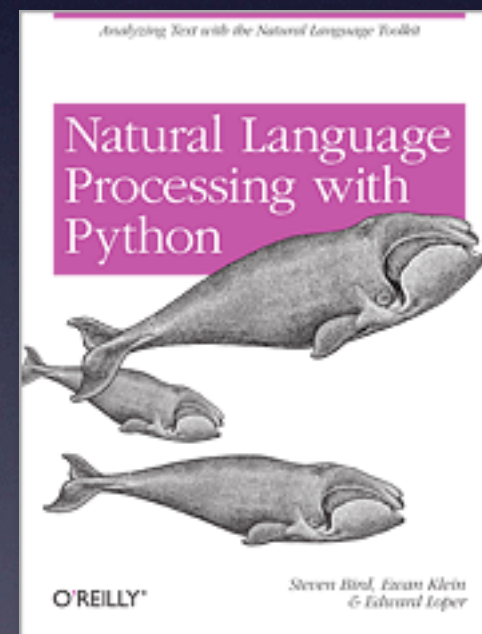
- ☒ Precisión
- ☒ Uniformidad
- ☒ Representación abstracta

Contra

- ☐ Lenguaje controlado
- ☐ Fràgil

Lenguaje Natural y Python

- NLTK: *Natural Language Toolkit*
- <http://nltk.org>





Grammatical Framework

- <http://www.grammaticalframework.org/>
- Lenguaje GF para gramáticas multilíngües
- Abstracto \leftrightarrow concreto
- Reciente: “bindings” de Python

Representación abstracta

- Qué puede expresarse
- *Interlingua*
- Única

Representación concreta

- **Cómo** debe expresarse
- Múltiples
 - una por lengua.

Math bar online

From: Eng To: All Clear Random

the empty set is a proper subset of the set with only element pi

, and intersection minus or over plus raised times union

Abstract: $\text{mkProp}(\text{prsubset emptyset}(\text{set_base nums1_pi}))$

Bul: празното множество е грозно подмножество на множеството с единствен елемент пи

Cat: el conjunt buit és un subconjunt propi del conjunt amb únic element pi

Eng: the empty set is a proper subset of the set with only element pi

Fin: tyhjä joukko on joukon ainoan alkion pii kanssa aito osajoukko

Fre: l' ensemble vide est un sous-ensemble propre de l' ensemble avec de l' unique élément pi

Ger: die leere Menge ist eine echte Teilmenge der Menge mit einzigem Element Pi

Hin: रिक्त समुच्चय एक एकल अवयव pi के साथ समुच्चय का उचित sub समुच्चय है

Ita: l' insieme vuoto è un sottoinsieme proprio dell' insieme con dell' unico elemento pi

LaTeX: $\text{\emptyset} \subset \{\pi\}$

Pol: pusty zbiór jest właściwym podzbiorem zbioru z pojedynczym elementem pi

Ron: mulțimea vidă este o submulțime proprie mulțimii cu unic element pi

Rus: пустое множество собственное subмножество множества с единственным элементом Пиом

Spa: el conjunto vacío es un subconjunto propio del conjunto con único elemento pi

Swe: den tomma mängden är en äkta delmängd av mängden med enda element pi

Urd: مجموعیہ ہے sub کے ساتھ مجموعیہ کا اصلی pi خالی مجموعیہ ان صرف عنصر

$$\emptyset \subset \{\pi\}$$

Eng

The sum of 3, 4 and 5

plus [n 3, n 4, n 5]



Eng

The sum of 3, 4 and 5

plus [n 3, n 4, n 5]



Eng

The sum of 3, 4 and 5



Spa

la suma de 3, 4 y 5

plus [n 3, n 4, n 5]

Sage

sum([3, 4, 5])

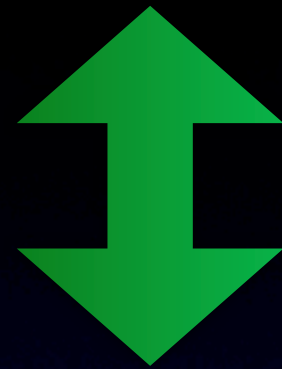
Eng

The sum of 3, 4 and 5

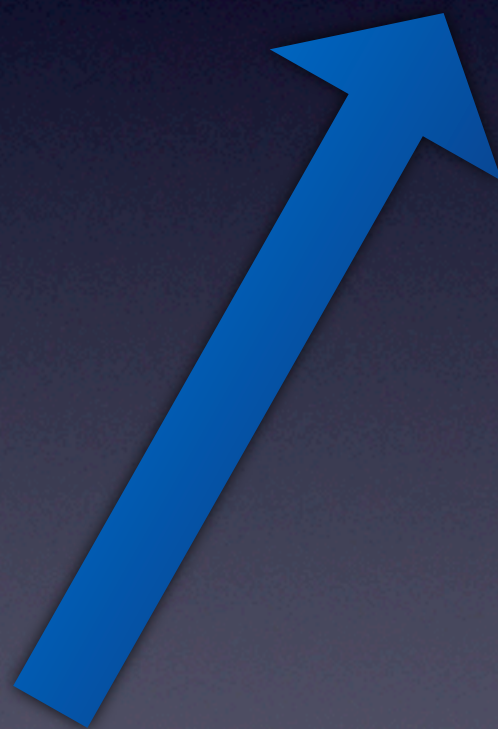
Spa

la suma de 3, 4 y 5

bin_plus n 3 (bin_plus n 4 n 5)



plus [n 3, n 4, n 5]



Eng

The sum of 3, 4 and 5



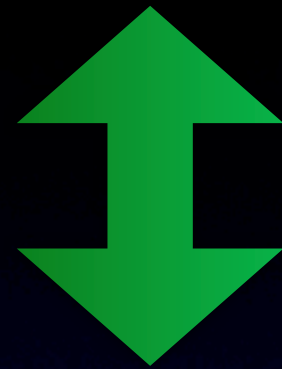
Sage

sum([3, 4, 5])

Spa

la suma de 3, 4 y 5

bin_plus n 3 (bin_plus n 4 n 5)



plus [n 3, n 4, n 5]

3 + 4 + 5

LaTeX

sum([3, 4, 5])

Sage

Eng

The sum of 3, 4 and 5

Spa

la suma de 3, 4 y 5

Pasos

- Módulo abstracto
- Módulos concretos para cada lengua
- Compilar para PGF:
 - *Parallel Grammar Format*
 - usable en Python

GF bindings

- modulo `pgf`
 - Clases:
 - `PGF`, `Concr`, `Expr`, ...
 - `readPGF`: leer un fichero PGF
 - `PGF.languages`: Concretos
 - `Concr.parse`: Análisis
 - `Concr.linearize`: Síntesis
 - `Expr.unpack`

Una gramàtica de matemàtiques

- Para docencia
 - Secundaria y 1r año de carrera
 - Cálculo y Álgebra



Modulo gfsage

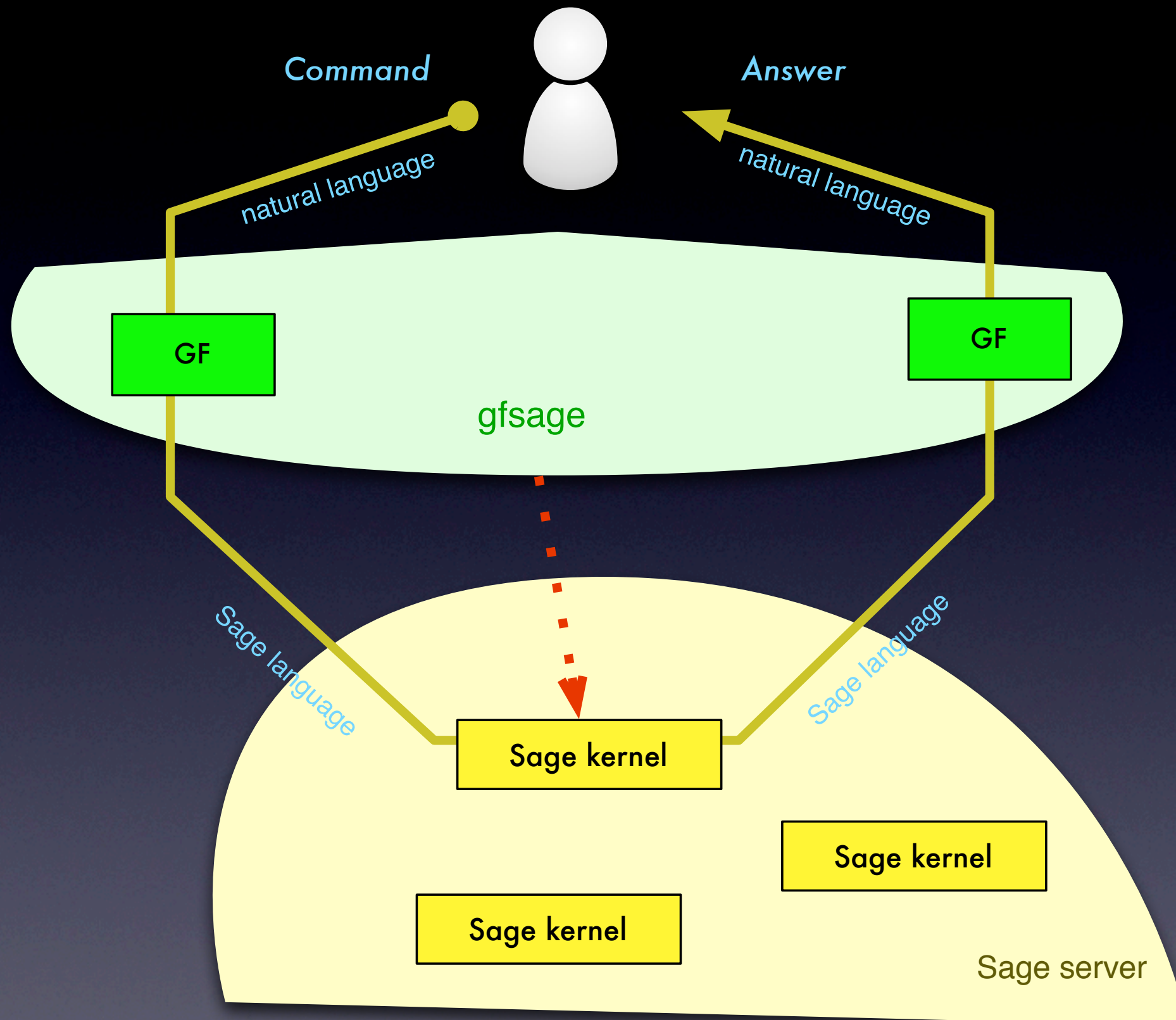
`gfsage`: 1 instancia

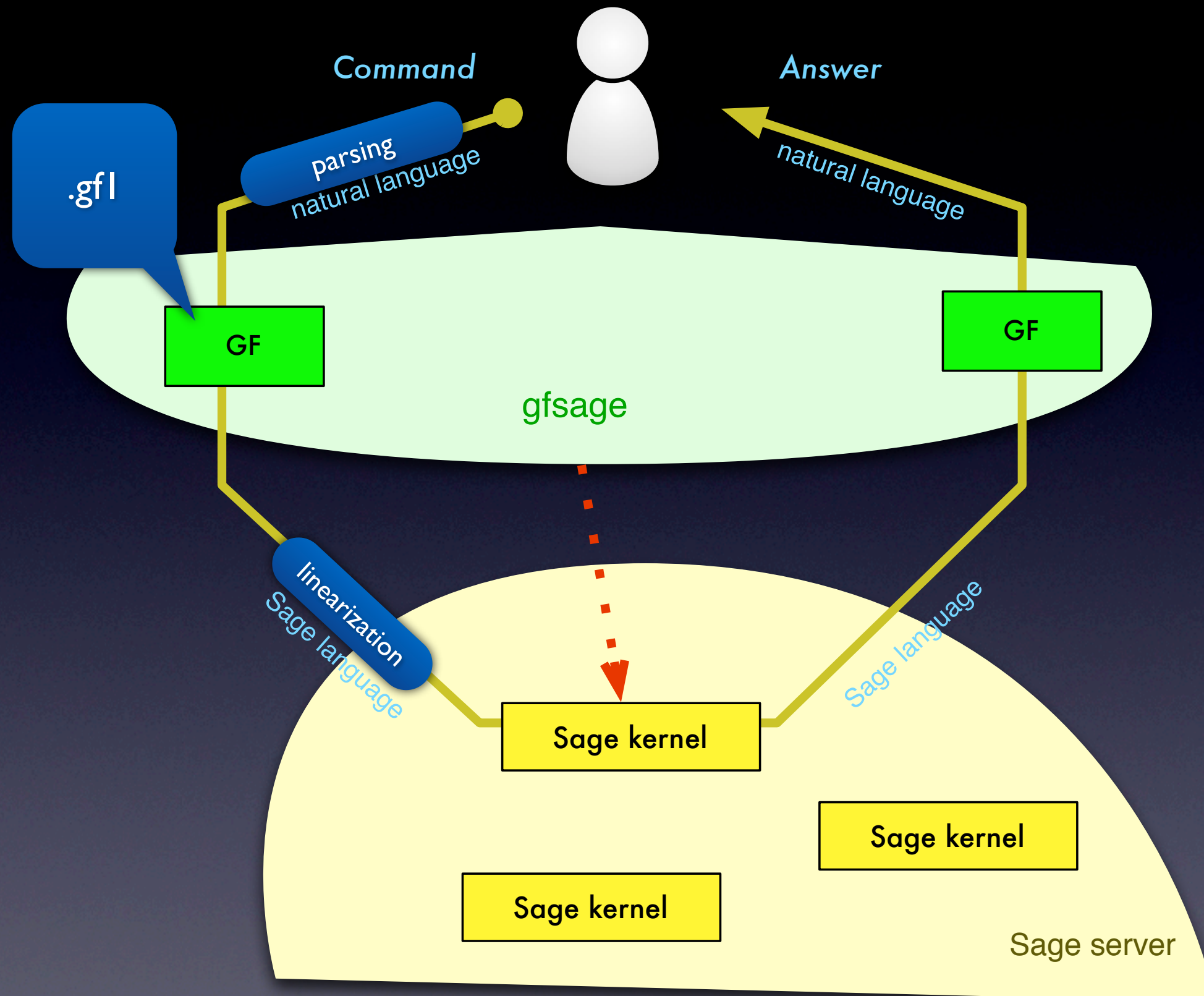
➡ `gf`: 2 instancias

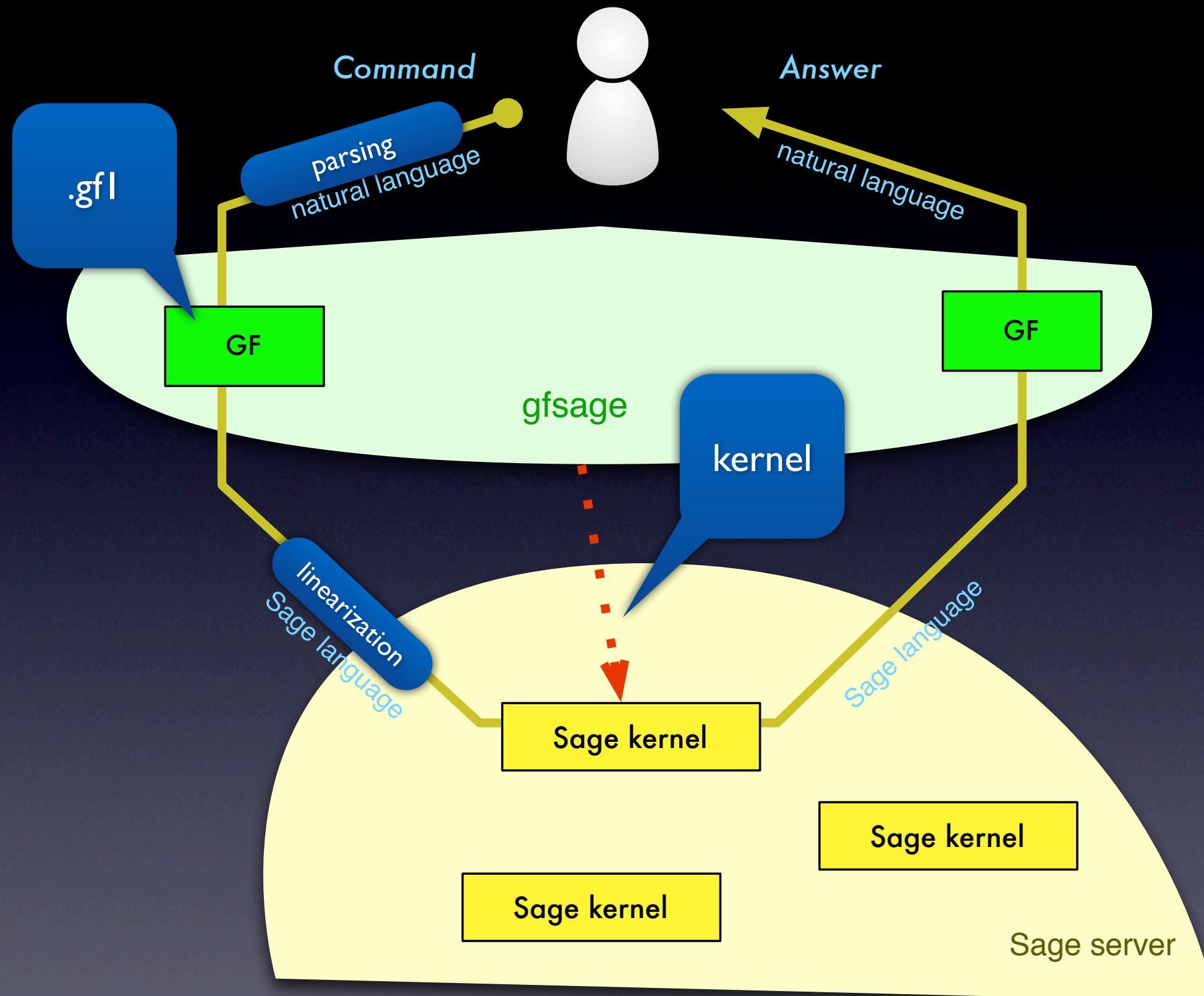
▶ Pregunta

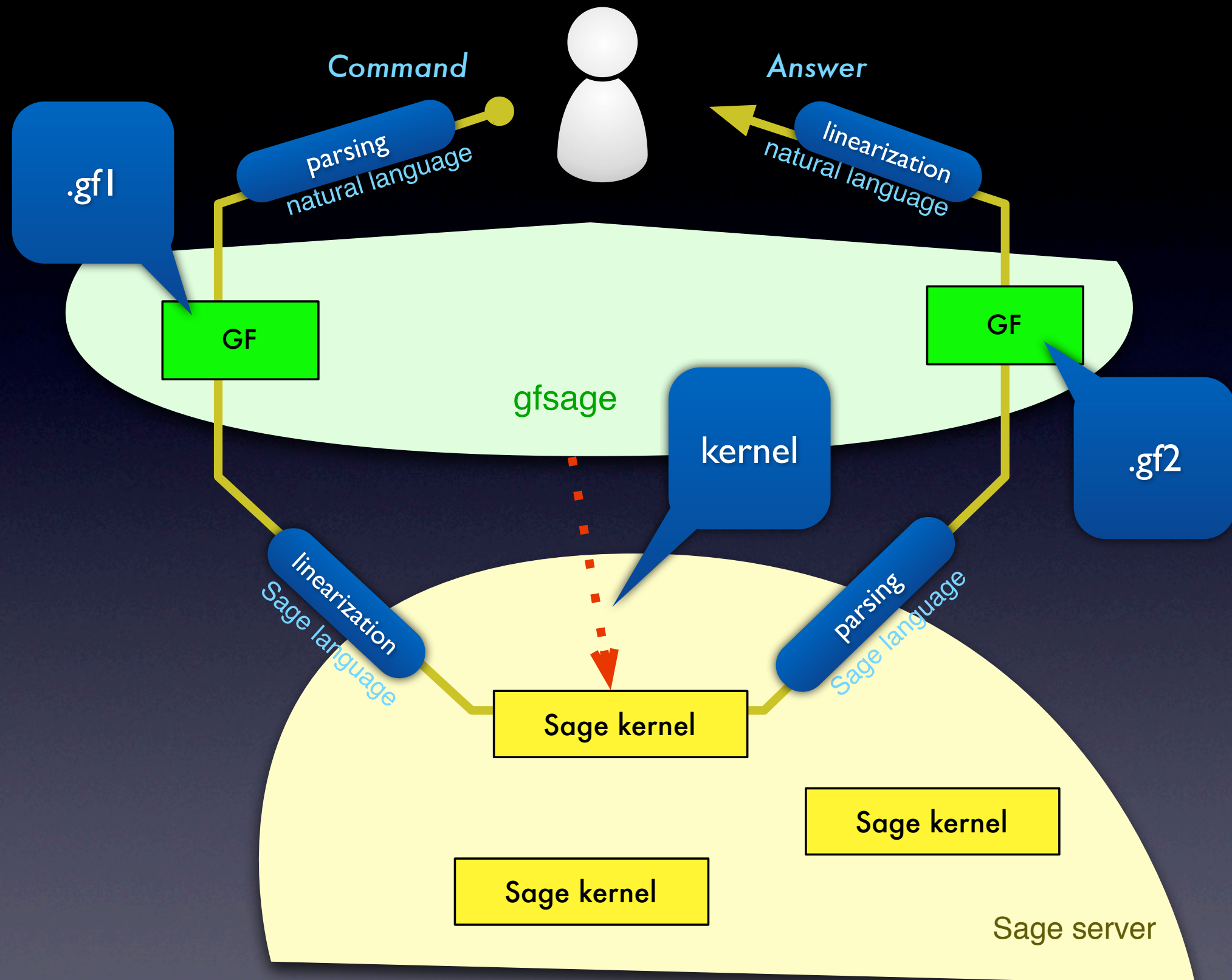
▶ Respuesta

➡ `pgf`









¿Dónde?

- GF:

<http://www.grammaticalframework.org/download/index.html>

- gfsage:

<svn://molto-project.eu/wp6/mgl>