

# Recursos lingüísticos, libres? Modelos de desenvolvemento

Susana Sotelo Docío

<susana.sotelo@usc.es>

Grupo de Gramática do español

Departamento de Lingua española  
Universidade de Santiago de Compostela

Curso de verán

*O Software Libre e a Lingüística (LINGUA 09)*

Santiago de Compostela, 21-24 de setembro de 2009

# Táboa de contidos

- 1 Recursos lingüísticos
  - Qué son os recursos lingüísticos?
  - Qué son os recursos lingüísticos libres?
  - As licenzas libres
- 2 Modelos de desenvolvemento
  - Modelos libres e privativos
  - Financiamento
  - Conclusión
- 3 Exemplo de modelo baseado no voluntariado: LanguageTool
  - Introducción
  - Estado actual
  - Tarefas pendentes e folia de ruta

# Que é un Recurso lingüístico (RL)?

- 1 Colección de datos lingüísticos preparados para ser usados con aplicacións de software (*data resources*)
- 2 Aplicacións de software que manexan datos lingüísticos (*algorithmic resources*)

Ex. corpus, léxicos, gramáticas, *treebanks*, etc

# Que é un Recurso lingüístico (RL)?

- 1 Colección de datos lingüísticos preparados para ser usados con aplicacións de software (*data resources*)
- 2 Aplicacións de software que manexan datos lingüísticos (*algorithmic resources*)

Ex. corpus, léxicos, gramáticas, *treebanks*, etc

# Que é un Recurso lingüístico (RL)?

- 1 Colección de datos lingüísticos preparados para ser usados con aplicacións de software (*data resources*)
- 2 Aplicacións de software que manexan datos lingüísticos (*algorithmic resources*)

Ex. corpus, léxicos, gramáticas, *treebanks*, etc

# Recursos lingüísticos libres

## Que é Software Libre?

Todo software cunha licenza que garanta as catro liberdades

- 0 Liberdade para usar o programa con calquera propósito
- 1 Liberdade para estudar o programa e adaptalo ás nosas necesidades
- 2 Liberdade de distribuír copias do programa
- 3 Liberdade de mellorar o programa e facer públicas esas melloras

# Recursos lingüísticos libres

## Que é Software Libre?

Todo software cunha licenza que garanta as catro liberdades

- 0 Liberdade para usar o programa con calquera propósito
- 1 Liberdade para estudar o programa e adaptalo ás nosas necesidades
- 2 Liberdade de distribuír copias do programa
- 3 Liberdade de mellorar o programa e facer públicas esas melloras

# As licenzas libres

Contratos que establecen as condicións de uso e distribución de **software** e garanten as catro liberdades

- 1 Tres grandes tipos de licenzas:
  - As obras derivadas teñen que seguir sendo libres (ex. GPL) [apertium, freeling]
  - As obras derivadas teñen que ser libres pero permiten enlazar con software privativo (ex. LGPL) [EP4IR]
  - As obras derivadas poden ser privativas (ex. BSD)
- 2 Diseñadas para a distribución de software

# As licenzas libres

Contratos que establecen as condicións de uso e distribución de **software** e garanten as catro liberdades

- 1 Tres grandes tipos de licenzas:
  - As obras derivadas teñen que seguir sendo libres (ex. GPL) [apertium, freeling]
  - As obras derivadas teñen que ser libres pero permiten enlazar con software privativo (ex. LGPL) [EP4IR]
  - As obras derivadas poden ser privativas (ex. BSD)
- 2 Diseñadas para a distribución de software

# As licenzas libres

Contratos que establecen as condicións de uso e distribución de **software** e garanten as catro liberdades

- 1 Tres grandes tipos de licenzas:
  - As obras derivadas teñen que seguir sendo libres (ex. GPL) [apertium, freeling]
  - As obras derivadas teñen que ser libres pero permiten enlazar con software privativo (ex. LGPL) [EP4IR]
  - As obras derivadas poden ser privativas (ex. BSD)
- 2 Diseñadas para a distribución de software

# A licenza LGPL-LR: conceptos (1/3)

- 1 Deseñada especificamente para establecer condicións de **copia, modificación e distribución** de Recursos Lingüísticos

RL: Colección de datos lingüísticos preparados para ser usados con aplicacións de software (lexicóns, gramáticas, corpus, etc)

- 2 Outros conceptos importantes

**Traballo Derivado (TD):** todo traballo que conteña o RL ou un anaco del, con ou sen modificacións

**Traballo que usa o LR (TRL):** programa que traballa cos datos contidos no RL

# A licenza LGPL-LR: conceptos (1/3)

- 1 Deseñada especificamente para establecer condicións de **copia, modificación e distribución** de Recursos Lingüísticos

**RL:** Colección de datos lingüísticos preparados para ser usados con aplicacións de software (lexicóns, gramáticas, corpus, etc)

- 2 Outros conceptos importantes

Traballo Derivado (TD): todo traballo que conteña o RL ou un anaco del, con ou sen modificacións

Traballo que usa o LR (TRL): programa que traballa cos datos contidos no RL

# A licenza LGPL-LR: conceptos (1/3)

- 1 Deseñada especificamente para establecer condicións de **copia, modificación e distribución** de Recursos Lingüísticos

**RL:** Colección de datos lingüísticos preparados para ser usados con aplicacións de software (lexicóns, gramáticas, corpus, etc)

- 2 Outros conceptos importantes

**Traballo Derivado (TD):** todo traballo que conteña o RL ou un anaco del, con ou sen modificacións

**Traballo que usa o LR (TRL):** programa que traballa cos datos contidos no RL

# A licenza LGPL-LR: condicións (2/3)

- 1 Permite copiar e distribuír o RL acompañado do texto da licenza
- 2 Permite copiar e distribuír modificacións feitas ao RL sempre que...
  - ... o resultado da modificación siga sendo un RL
  - ... se deixe constancia documental da modificación
  - ... a modificación se distribúa baixo a licenza LGPL-LR
- 3 Os TRL non están cubertos por esta licenza, pero cando se combinan co RL nun paquete...
  - ... o paquete considérase un traballo derivado (por tanto LGPL-LR)
  - ... a distribución do RL no paquete debe ser en formato electrónico

Exemplos: XIADA, Spanish Resource Grammar, RL de Unitex

# A licenza LGPL-LR: condicións (2/3)

- 1 Permite copiar e distribuír o RL acompañado do texto da licenza
- 2 Permite copiar e distribuír modificacións feitas ao RL sempre que...
  - ... o resultado da modificación siga sendo un RL
  - ... se deixe constancia documental da modificación
  - ... a modificación se distribúa baixo a licenza LGPL-LR
- 3 Os TRL non están cubertos por esta licenza, pero cando se combinan co RL nun paquete...
  - ... o paquete considérase un traballo derivado (por tanto LGPL-LR)
  - ... a distribución do RL no paquete debe ser en formato electrónico

Exemplos: XIADA, Spanish Resource Grammar, RL de Unitex..

# A licenza LGPL-LR: condicións (2/3)

- 1 Permite copiar e distribuír o RL acompañado do texto da licenza
- 2 Permite copiar e distribuír modificacións feitas ao RL sempre que...
  - ... o resultado da modificación siga sendo un RL
  - ... se deixe constancia documental da modificación
  - ... a modificación se distribúa baixo a licenza LGPL-LR
- 3 Os TRL non están cubertos por esta licenza, pero cando se combinan co RL nun paquete...
  - ... o paquete considérase un traballo derivado (por tanto LGPL-LR)
  - ... a distribución do RL no paquete debe ser en formato electrónico

Exemplos: XIADA, Spanish Resource Grammar, RL de Unitex..  

## A licenza LGPL-LR: condicións (2/3)

- 1 Permite copiar e distribuír o RL acompañado do texto da licenza
- 2 Permite copiar e distribuír modificacións feitas ao RL sempre que...
  - ... o resultado da modificación siga sendo un RL
  - ... se deixe constancia documental da modificación
  - ... a modificación se distribúa baixo a licenza LGPL-LR
- 3 Os TRL non están cubertos por esta licenza, pero cando se combinan co RL nun paquete...
  - ... o paquete considérase un traballo derivado (por tanto LGPL-LR)
  - ... a distribución do RL no paquete debe ser en formato electrónico

Exemplos: XIADA, Spanish Resource Grammar, RL de Unitex..  

# A licenza LGPL-LR: exemplo (3/3)

## Exemplo práctico

Léxico XIADA (v2.2)	RL	LGPL-LR
XIADAbis (v0.1) (baseado en XIADA)	TD	LGPL-LR
FreeLingDICT	RL	GPL
Paquete XIADA + FreeLingDICT (adición)	RL+RL	LGPL-LR + GPL
NovoDICT (XIADAbis + FreeLingDICT)	TD	LGPL-LR
FreeLing	TRL	GPL

<http://sanskrit.inria.fr/DATA/LGPLLR.pdf>

# Modelos de desenvolvemento

## 1 Modelo privativo

- Modelo pechado
- Modelo aberto ou pseudo-aberto

- Baseado na difusión *controlada* dos recursos producidos
- Mundo empresarial e unha parte importante do académico

## 2 Modelo libre

- Académico: empresas + universidades e centros I+D
- Voluntario: comunidade de SL

- O resultado é o mesmo, cambian enfoque e sistema de traballo

# Modelos de desenvolvemento

## 1 Modelo privativo

- Modelo pechado
- Modelo aberto ou pseudo-aberto

- Baseado na difusión *controlada* dos recursos producidos
- Mundo empresarial e unha parte importante do académico

## 2 Modelo libre

- Académico: empresas + universidades e centros I+D
- Voluntario: comunidade de SL

- O resultado é o mesmo, cambian enfoque e sistema de traballo

# Modelos de desenvolvemento

## 1 Modelo privativo

- Modelo pechado
- Modelo aberto ou pseudo-aberto

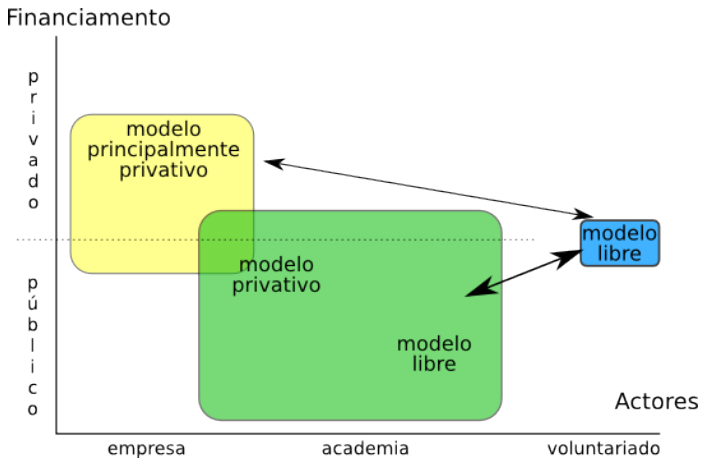
- Baseado na difusión *controlada* dos recursos producidos
- Mundo empresarial e unha parte importante do académico

## 2 Modelo libre

- Académico: empresas + universidades e centros I+D
- Voluntario: comunidade de SL

- O resultado é o mesmo, cambian enfoque e sistema de traballo

# Financiamento



# Conclusión

## Os recursos académicos deben ser libres

- 1 (social) Financiamento público: retorno da inversión
- 2 (científico) Non existe posibilidade de reproducir os experimentos

➤ O caso islandés: iceNLP

➤ Hai que comezar a falar de coñecemento libre

The Open Knowledge Foundation: <http://www.okfn.org/>

# Conclusión

## Os recursos académicos deben ser libres

- 1 (social) Financiamento público: retorno da inversión
- 2 (científico) Non existe posibilidade de reproducir os experimentos

➤ O caso islandés: iceNLP

➤ Hai que comezar a falar de coñecemento libre

The Open Knowledge Foundation: <http://www.okfn.org/>

# Conclusión

## Os recursos académicos deben ser libres

- 1 (social) Financiamento público: retorno da inversión
- 2 (científico) Non existe posibilidade de reproducir os experimentos

➤ O caso islandés: iceNLP

➤ Hai que comezar a falar de coñecemento libre

The Open Knowledge Foundation: <http://www.okfn.org/>



# Conclusión

## Os recursos académicos deben ser libres

- 1 (social) Financiamento público: retorno da inversión
  - 2 (científico) Non existe posibilidade de reproducir os experimentos
- O caso islandés: iceNLP

➤ Hai que comezar a falar de coñecemento libre

The Open Knowledge Foundation: <http://www.okfn.org/>



# Conclusión

## Os recursos académicos deben ser libres

- 1 (social) Financiamento público: retorno da inversión
- 2 (científico) Non existe posibilidade de reproducir os experimentos

➤ O caso islandés: iceNLP

➤ Hai que comezar a falar de coñecemento libre

The Open Knowledge Foundation: <http://www.okfn.org/>

## Parte II.

# Un exemplo de modelo libre baseado no voluntariado: LanguageTool

# LanguageTool (1/3)

**Que é?** Un corrector lingüístico sensíbel ao contexto baseado en regras

≠ corrector ortográfico

## Componentes

dicionario + regras + etiquetador + segmentador + tokenizador  
(opt) regras de desambigüación + regras de falsos amigos

## As regras

- XML [▶ mostra](#)
- Java
  - Xenéricas: WORD\_REPEAT\_RULE, DOUBLE\_PUNCTUATION, UNPAIRED\_BRACKETS, UPPERCASE\_SENTENCE\_START, etc
  - Específicas: [▶ ElwithFemRule](#)

# LanguageTool (1/3)

**Que é?** Un corrector lingüístico sensíbel ao contexto baseado en regras  
≠ corrector ortográfico

## Componentes

dicionario + regras + etiquetador + segmentador + tokenizador  
(opt) regras de desambigüación + regras de falsos amigos

## As regras

- XML [▶ mostra](#)
- Java
  - Xenéricas: WORD\_REPEAT\_RULE, DOUBLE\_PUNCTUATION, UNPAIRED\_BRACKETS, UPPERCASE\_SENTENCE\_START, etc
  - Específicas: [▶ ElwithFemRule](#)

# LanguageTool (1/3)

**Que é?** Un corrector lingüístico sensíbel ao contexto baseado en regras

≠ corrector ortográfico

## Componentes

dicionario + regras + etiquetador + segmentador + tokenizador  
(opt) regras de desambigüación + regras de falsos amigos

## As regras

- **XML** [▶ mostra](#)
- **Java**
  - Xenéricas: WORD\_REPEAT\_RULE, DOUBLE\_PUNCTUATION, UNPAIRED\_BRACKETS, UPPERCASE\_SENTENCE\_START, etc
  - Específicas: [▶ ElwithFemRule](#)



# LanguageTool (2/3)

## Que tenta corrixir?

- A nivel de palabra, o que Hunspell/MySpell non resolve
  - Homónimos: grúa/guindastre
  - Castellanismos léxicos:
  - Falsos amigos: (gl) real → (en) real, royal
- En niveis superiores todo o que se poida expresar mediante patróns (con limitacións).
  - Gramática:
  - Estilo: palabras repetidas, dobre puntuación, etc.

# LanguageTool (2/3)

## Que tenta corrixir?

- A nivel de palabra, o que Hunspell/MySpell non resolve
  - Homónimos: grúa/guindastre ▶ regra
  - Castelanismos léxicos: ▶ regra
  - Falsos amigos: (gl) real → (en) real, royal ▶ regra
- En niveis superiores todo o que se poida expresar mediante patróns (con limitacións).
  - Gramática:  
  - Estilo: palabras repetidas, doble puntuación, etc.

# LanguageTool (2/3)

## Que tenta corrixir?

- A nivel de palabra, o que Hunspell/MySpell non resolve
  - Homónimos: grúa/guindastre ▶ regra
  - Castelanismos léxicos: ▶ regra
  - Falsos amigos: (gl) real → (en) real, royal ▶ regra
- En niveis superiores todo o que se poida expresar mediante patróns (con limitacións).
  - Gramática: ▶ regra ▶ regra
  - Estilo: palabras repetidas, dobre puntuación, etc.

# LanguageTool (3/3)

## Modelo

- Naceu como proxecto Fin de Carreira de Daniel Naber
- Sistema de traballo baseado no voluntariado para software e recursos lingüísticos (agás islandés)

## Como usalo?

- Como extensión en OpenOffice.org; integrado en LyX
- De xeito independente: liña de ordes e GUI ▶ mostra

Versión 1.0-dev (CVS)

<http://gramatica.usc.es/~sdocio/downloads/LanguageTool-1.0.0-dev.oxt>



# LanguageTool (3/3)

## Modelo

- Naceu como proxecto Fin de Carreira de Daniel Naber
- Sistema de traballo baseado no voluntariado para software e recursos lingüísticos (agás islandés)

## Como usalo?

- Como extensión en OpenOffice.org; integrado en LyX
- De xeito independente: liña de ordes e GUI [▶ mostra](#)

Versión 1.0-dev (CVS)

<http://gramatica.usc.es/~sdocio/downloads/LanguageTool-1.0.0-dev.oxt>



# LanguageTool (3/3)

## Modelo

- Naceu como proxecto Fin de Carreira de Daniel Naber
- Sistema de traballo baseado no voluntariado para software e recursos lingüísticos (agás islandés)

## Como usalo?

- Como extensión en OpenOffice.org; integrado en LyX
- De xeito independente: liña de ordes e GUI [▶ mostra](#)

Versión 1.0-dev (CVS)

<http://gramatica.usc.es/~sdocio/downloads/LanguageTool-1.0.0-dev.oxt>

# Estado actual

## 1 Versión oficial (v0.9.9)

Lingua	XML	Java	Lingua	XML	Java
francés	1673	1	ruso	67	1
polaco	912	2	islandés	39	0
inglés	464	2	sueco	26	1
romanés	404	0	esloveno	21	0
holandés	265	0	eslovaco	13	1
alemán	79	7	ucraíno	8	1
italiano	77	0	español	3	0

## 2 Próxima versión (v1.0-dev, agora no CVS)

- Galego (32/0)
- Español (4/1)

# Estado actual

## 1 Versión oficial (v0.9.9)

Lingua	XML	Java	Lingua	XML	Java
francés	1673	1	ruso	67	1
polaco	912	2	islandés	39	0
inglés	464	2	sueco	26	1
romanés	404	0	esloveno	21	0
holandés	265	0	eslovaco	13	1
alemán	79	7	ucraíno	8	1
italiano	77	0	español	3	0

## 2 Próxima versión (v1.0-dev, agora no CVS)

- Galego (32/0)
- Español (4/1)

# Tarefas pendentes (ToDo)

- 1 **Diccionario:** integración de novas fontes; determinación de cobertura (opción -t)
- 2 **Segmentador:** regras SRX para galego e español
- 3 **Falsos amigos**
  - Obtención semiautomática mediante diccionarios bilingües e medidas de similitude: *adept* (experto) e *adepto*
  - Adaptación da funcionalidade dos falsos amigos para comprobación terminolóxica en 110n
- 4 **Regras**
  - Paquetes de regras adaptados a necesidades de proxectos de 110n (estilo e escollas libres)
  - Regras que verifiquen a consistencia nas escollas libres

# Tarefas pendentes (ToDo)

- 1 **Diccionario:** integración de novas fontes; determinación de cobertura (opción -t)
- 2 **Segmentador:** regras SRX para galego e español
- 3 **Falsos amigos**
  - Obtención semiautomática mediante diccionarios bilingües e medidas de similitude: *adept* (experto) e *adepto*
  - Adaptación da funcionalidade dos falsos amigos para comprobación terminolóxica en 110n
- 4 **Regras**
  - Paquetes de regras adaptados a necesidades de proxectos de 110n (estilo e escollas libres)
  - Regras que verifiquen a consistencia nas escollas libres

# Tarefas pendentes (ToDo)

- 1 **Diccionario:** integración de novas fontes; determinación de cobertura (opción -t)
- 2 **Segmentador:** regras SRX para galego e español
- 3 **Falsos amigos**
  - Obtención semiautomática mediante diccionarios bilingües e medidas de similitude: *adept* (experto) e *adepto*
  - Adaptación da funcionalidade dos falsos amigos para comprobación terminolóxica en 110n
- 4 **Regras**
  - Paquetes de regras adaptados a necesidades de proxectos de 110n (estilo e escollas libres)
  - Regras que verifiquen a consistencia nas escollas libres

# Tarefas pendentes (ToDo)

- 1 **Diccionario:** integración de novas fontes; determinación de cobertura (opción -t)
- 2 **Segmentador:** regras SRX para galego e español
- 3 **Falsos amigos**
  - Obtención semiautomática mediante diccionarios bilingües e medidas de similitude: *adept* (experto) e *adepito*
  - Adaptación da funcionalidade dos falsos amigos para comprobación terminolóxica en 110n [▶ mostra](#)
- 4 **Regras**
  - Paquetes de regras adaptados a necesidades de proxectos de 110n (estilo e escollas libres)
  - Regras que verifiquen a consistencia nas escollas libres

# Tarefas pendentes (ToDo)

- 1 **Diccionario:** integración de novas fontes; determinación de cobertura (opción -t)
- 2 **Segmentador:** regras SRX para galego e español
- 3 **Falsos amigos**
  - Obtención semiautomática mediante diccionarios bilingües e medidas de similitude: *adept* (experto) e *adepito*
  - Adaptación da funcionalidade dos falsos amigos para comprobación terminolóxica en 110n [▶ mostra](#)
- 4 **Regras**
  - Paquetes de regras adaptados a necesidades de proxectos de 110n (estilo e escollas libres)
  - Regras que verifiquen a consistencia nas escollas libres

# Folla de ruta de LanguageTool

## 1 A curto e medio prazo

- Regras de usuario (engadidas)
- Posibilidade de traballar con ficheiros XLIFF e PO
- Tokenización interna das palabras (clífticos)
- Simplificación do proceso para engadir novas linguas
- Bitext rules (reglas aplicables a textos paralelos)
- Integración dun identificador lingüístico para evitar falsos positivos

## 2 A medio e longo prazo

- Named-Entity Recognition para xerar índices alfabéticos en OO.org
- Clasificación de dominio

# Folla de ruta de LanguageTool

## 1 A curto e medio prazo

- Regras de usuario (engadidas)
- Posibilidade de traballar con ficheiros XLIFF e PO
- Tokenización interna das palabras (clíticos)
- Simplificación do proceso para engadir novas linguas
- Bitext rules (regras aplicables a textos paralelos)
- Integración dun identificador lingüístico para evitar falsos positivos

## 2 A medio e longo prazo

- Named-Entity Recognition para xerar índices alfabéticos en OO.org
- Clasificación de dominio

# Folla de ruta de LanguageTool

## 1 A curto e medio prazo

- Regras de usuario (engadidas)
- Posibilidade de traballar con ficheiros XLIFF e PO
- Tokenización interna das palabras (clíticos)
- Simplificación do proceso para engadir novas linguas
- Bitext rules (regras aplicables a textos paralelos)
- Integración dun identificador lingüístico para evitar falsos positivos

## 2 A medio e longo prazo

- Named-Entity Recognition para xerar índices alfabéticos en OO.org
- Clasificación de dominio

# Fin

◀ Volver

LanguageTool 1.0.0-dev

Ficheiro Axuda

En O Morrazo hai moi boas praias. Eu tamén vou a aportar algo ao proxecto en seguida. A grúa ía cara Vigo.

1. Liña 1, columna 0  
**Mensaxe:** Algunhas preposicións seguidas de topónimos con artigo determinado contraen con este. Quería vostede dicir **no Morrazo**?  
**Corrección:** No Morrazo  
**Contexto:** En O Morrazo hai moi boas praias. Eu tamén vou a apo...

2. Liña 1, columna 43  
**Mensaxe:** A perífrase formada polo verbo "ir" máis un infinitivo constrúese en galego sen preposición: **vou aportar**  
**Corrección:** vou aportar  
**Contexto:** ...O Morrazo hai moi boas praias. Eu tamén **vou a aportar** algo ao proxecto en seguida. A grúa ía ...

3. Liña 1, columna 49  
**Mensaxe:** Posíbel castelanismos: o verbo "aportar" significa en galego "chegar a porto". A forma correcta en galego para o sentido de "proporcionar" ou "presentar" é **achegar** ou **facer unha achega**.  
**Corrección:** achegar, facer unha achega  
**Contexto:** ...azo hai moi boas praias. Eu tamén vou a **aportar** algo ao proxecto en seguida. A grúa ía ...

4. Liña 1, columna 74  
**Mensaxe:** As formas "en" e "seguida" forman en galego unha locución que se escribe unida. Quería vostede dicir **enseguida**?  
**Corrección:** enseguida  
**Contexto:** ...Eu tamén vou a aportar algo ao proxecto **en seguida**. A grúa ía cara Vigo.

5. Liña 1, columna 88  
**Mensaxe:** Posíbel castelanismos: o substantivo "grúa" fai referencia a un paxaro. Para denominar ao aparato empregado para mover ou remolcar grandes pesos emprégase preferiblemente **guindastre** (masculino).  
**Corrección:** guindastre  
**Contexto:** ... aportar algo ao proxecto en seguida. A **grúa** ía cara Vigo.

6. Liña 1, columna 96  
**Mensaxe:** A preposición "cara" vai seguida de 'a' en galego. Quería vostede dicir **cara a**?  
**Corrección:** cara a  
**Contexto:** ...aloo ao proxecto en seguida. A grúa ía **cara** Vigo.

Por favor, teclee ou pegue o texto a corrixir na área superior

Comprobar o texto Lingua do texto: galego

# Falsos amigos (1/3)

## Glosario de TRASNO

### **address**

1. n.m. enderezo (recomendado)

Non se aceptan 'dirección' nin 'dirección'

### **add**

1. v.t. engadir

Non son correctos 'añadir' nin 'adicionar'

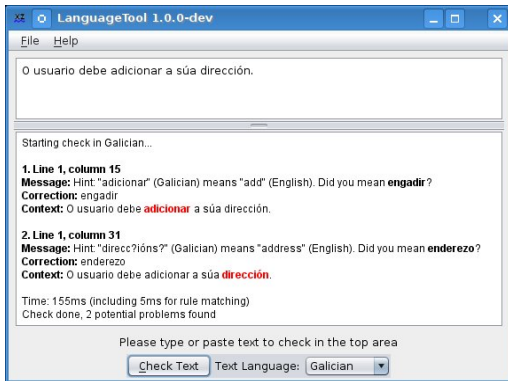
## Falsos amigos (2/3)

```
<rulegroup id="ADDRESS">
  <rule>
    <pattern lang="en">
      <token>address</token>
    </pattern>

    <translation lang="gl">endereço</translation>
  </rule>

  <rule>
    <pattern lang="gl">
      <token regexp="yes">direcc?ións?</token>
    </pattern>
    <translation lang="en">address</translation>
  </rule>
</rulegroup>
```

◀ Volver



# Regras de LT. Mostra de regra simple

[◀ Volver](#)

```
<rule id="ID" name="nome">
  <pattern>
    <token regexp="yes">patrón regular</token>
    <token postag="NC.*" postag_regexp="yes"/>
    <token inflect="yes">forma</token>
  </pattern>
  <message>Mensaxe para o usuario relativo a "<match no="1"/>" e
<suggestion>unha suxestión</suggestion>.</message>
  <example type="incorrect">Un erro: <marker>erro</marker>.</example>
  <example type="correct">Corrección: <marker>correcto</marker>.</example>
</rule>
```

# Regras de LT. Java

◀ Volver

```
[...]
@Override
public String getDescription() {
    return "Use of 'el' with some feminine nouns";
}

@Override
public RuleMatch[] match(final AnalyzedSentence text) {
    final List<RuleMatch> ruleMatches = new ArrayList<RuleMatch>();
    final AnalyzedTokenReadings[] tokens = text.getTokensWithoutWhitespace();
    String prevToken = "";
    int prevPos = 0;
    for (int i = 1; i < tokens.length; i++) {
        String token = tokens[i].getToken();
        boolean doesRequireEl = false;

        token = token.replaceAll("[^a-záéíóúñüA-ZÁÉÍÓÚÑÜ0-9\\.\']", "");
        if (StringTools.isEmpty(token)) {

```

# Regras de LT. Léxico::Homónimos

◀ Volver

```
<rule id="GRUA" name="grua (guindastre)">
  <pattern>
    <token inflected="yes">grúa</token>
  </pattern>
  <message>Posíbel castelanismo: o substantivo "grúa" fai referencia
a un paxaro. Para denominar ao aparato empregado para mover ou remolcar
grandes pesos emprégase preferibelmente
<suggestion>guindastre</suggestion> (masculino).</message>
  <short>Posíbel forma incorrecta</short>
  <example type="incorrect">A <marker>grúa</marker> levoulle o
automóbil.</example>
  <example type="correct">O <marker>guindastre</marker> levoulle o
automóbil.</example>
</rule>
```

# Regras de LT. Léxico::Castelanismos

[← Volver](#)

```
<rule id="ACERA" name="acera (beirarrúa)">
  <pattern>
    <token regexp="yes">aceras?</token>
  </pattern>
  <message>A palabra "<match no="1"/>" é un castelanismo. Empregue no seu
sitio <suggestion>beirarrúa</suggestion>.</message>
  <short>Castelanismo léxico</short>
  <example type="incorrect">As <marker>aceras</marker> da vila están en
mal estado.</example>
  <example type="correct">As <marker>beirarrúas</marker> da vila están
en mal estado.</example>
</rule>
```

# Regras de LT. Falsos amigos

[← Volver](#)

```
<rulegroup id="REAL">
  <rule>
    <pattern lang="en">
      <token>real</token>
    </pattern>
    <translation lang="gl">verdadeiro</translation>
    <translation lang="gl">auténtico</translation>
    <example type="incorrect">The Real Academy of Sciences.</example>
  </rule>

  <rule>
    <pattern lang="gl">
      <token inflected="yes">real</token>
    </pattern>
    <translation lang="en">royal</translation>
    <translation lang="en">real</translation>
    <example type="correct">The Royal Academy of Sciences.</example>
  </rule>
</rulegroup>
```

# Regras de LT. Gramática

[← Volver](#)

```
<rule id="AMBOS" name="ambos, ambas e substantivo">
  <pattern>
    <token regexp="yes">(entr)?amb[oa]s</token>
    <!-- HACK to avoid [ao]s as noun (barely used) (FIXME) -->
    <token postag="NC.P000" postag_regexp="yes"><exception postag="NC.*"
regexp="yes">[oa]s</exception></token>
  </pattern>
  <message>A forma "<match no="1"/>" require polo xeral do artigo
determinado cando vai diante dun substantivo. Quería vostede dicir
<suggestion><match no="1" regexp_match="( ?iu)amb([ao])s"
regexp_replace="ámb$1l$1s"/> \2</suggestion> ou <suggestion>
<match no="1" regexp_match="( ?iu)amb([ao])s" regexp_replace="amb$1s $1s"/>
\2</suggestion>?</message>
  <example type="incorrect">Visitou <marker>ambas cidades</marker>.</example>
  <example type="correct">Visitou <marker>ámbalas cidades</marker>.</example>
  <example type="correct">Visitou <marker>ambas as cidades</marker>.</example>
</rule>
```

# Regras de LT. Gramática

[← Volver](#)

```
<rule id="IR_A_INF" name="ir e infinitivo">
  <pattern>
    <token inflected="yes">ir</token>
    <token>a</token>
    <token postag="VMN0000|VSN0000" postag_regexp="yes" />
  </pattern>
  <message>A perífrase formada polo verbo "ir" máis un infinitivo
  constrúese en galego sen preposición: <suggestion>\1 \3</suggestion>
</message>

  <example type="incorrect"><marker>vou a saltar</marker> con ela.</example>
  <example type="correct"><marker>vou saltar</marker> con ela.</example>
</rule>
```

# LanguageTool: diccionario

[← Volver](#)

## Compilado con FSA (Finite state automata)

```
...  
ganglioma      ganglioma      NCMS000  
gangliomas     ganglioma      NCMP000  
ganglionar     ganglionar     AQOCS0  
ganglionares   ganglionar     AQOCPO  
ganglios       ganglio        NCMP000  
gangrena       gangrena       NCF5000  
gangrenas      gangrena       NCFP000  
gangrenosa     gangrenoso     AQOFS0  
gangrenosas    gangrenoso     AQOFPO  
gangrenoso     gangrenoso     AQOM50  
...
```