

Conocimiento y Razonamiento Automatizado: Implementación de una Gramática de Clausulas Definidas en Prolog

Carlos Javier Tacón Fernández

12 de abril de 2018

1. Introducción

El lenguaje natural, en cualquier idioma, se compone de muchas estructuras gramaticales y sintácticas diferentes, con reglas flexibles para la formación de oraciones y concordancia en diferentes aspectos morfológicos de las palabras.

En esta práctica se explora una forma de traducción basada en la evaluación de estructuras sintácticas en oraciones usando gramáticas en español e inglés, haciendo la correspondencia entre unas estructuras u otras dependiendo de las características específicas del idioma.

2. Grupos Gramaticales

Tanto en español como en inglés se han definido los mismos grupos gramaticales, pero estos idiomas no tienen la misma estructura así que se han manejado dos gramáticas diferentes, que aún teniendo prácticamente las mismas definiciones, su desarrollo es diferente:

- **Oración.** Es la estructura sintáctica más global que se va a analizar, se compone de un grupo nominal y un grupo preposicional en su forma más común. En español podemos omitir el sujeto en todas las ocasiones, pero en inglés no podemos omitirlo salvo excepciones (oraciones coordinadas, por ejemplo), así que las reglas varían en este caso.
- **Oración relativo (subordinada).** Una oración que es parte de un grupo nominal más amplio, añadida con una conjunción de subordinación, que en inglés se puede omitir, así que en inglés tendremos dos reglas diferentes para esta estructura, con y sin conjunción. En castellano no omitimos esta conjunción así que solo escribiremos una regla.

- **Oración conjuntiva (coordinada).** Una oración que es precedida un nexo o conjunción coordinada y que forma parte de otra oración, en este caso es similar en ambos idiomas.
- **Grupo nominal.** Formado por un nombre, pudiendo también tener un determinante y un adjetivo. También puede ser un pronombre, que hace referencia a algo sustituyéndolo. El orden de un adjetivo que acompañe a un nombre es opuesto en inglés y español, pero el resto de casos son muy parecidos
- **Grupo preposicional.** Un grupo nominal precedido por una preposición. En el caso de la preposición *a*, esta puede ser omitida en inglés en algunas ocasiones así que este idioma tendrá una regla específica para la posibilidad de omitirla.
- **Grupo verbal.** Simboliza la acción del grupo nominal sujeto, tiene la misma estructura en ambos idiomas, pudiendo ser solo un verbo o verbo más un complemento que puede ser un grupo nominal, preposicional, adjetival o adverbial.
- **Grupo adjetival.** Define las cualidades de un nombre, en inglés esta estructura se localiza antes del nombre y en castellano después, las reglas son diferentes. Puede ser precedido de un grupo adverbial.
- **Grupo adverbial.** Modifica un nombre o un adjetivo, en ambos idiomas tiene la misma estructura, en este caso recursiva ya que puede haber varios adverbios que modifiquen a un solo objeto.

3. Reglas de Concordancia

Las reglas de concordancia morfológica presentan algunas diferencias entre los idiomas español e inglés. Las características que estos dos lenguajes comparten es la concordancia de persona entre el sujeto y el verbo (en este caso el sujeto es un grupo nominal que será tercera persona a no ser que esté formado por un pronombre, que tiene implícita la representación de una persona específica), además de la concordancia en número.

El castellano es un lenguaje segregado por género, tanto los determinantes como los pronombres, nombres, adverbios y adjetivos tienen género además de número. Esta diferenciación de género no está presente en inglés así que en el programa, solo el castellano tendrá que comprobar esa concordancia.

- Misma persona y número entre el grupo nominal con el grupo verbal que representan sujeto y predicado.
- Mismo género y número entre determinante, nombre y adjetivos.

El inglés tiene la particularidad de tener un determinante artículo indeterminado diferente dependiendo de si la palabra siguiente empieza por vocal o

consonante, esto no ocurre en castellano, así que solo tendremos que comprobarlo en inglés.

- Misma persona y número entre el grupo nominal con el grupo verbal que representan sujeto y predicado.
- Mismo número entre determinante y nombre.
- Dependiendo de si la siguiente palabra es vocal o consonante se utilizará *a* o *an*.

4. Traducción de oraciones de ejemplo

Al traducir oraciones de español a inglés, como hemos hecho dos gramáticas que funcionan de forma bidireccional, podemos observar que en la gramática española tenemos esta regla:

```
g_nominal_basico(gnb(Nom) , 3 , Genero , Numero) —>
determinante(det(art_det) , _ , Genero , Numero) , nombre
(Nom, Genero , Numero, c) .
```

En la que un artículo determinado más un nombre puede funcionar solo como nombre en inglés [el,alumno,ama,la,universidad]=[the,student,loves,university] esto hace que muchas frases omitan el artículo de su sujeto (no es erróneo, los titulares de periódicos tienen esta estructura), haciendo que haya algunas traducciones menos naturales, incluso erróneas en algunos casos.

Como tenemos que tener en cuenta casos específicos, para resolver un problema en un caso generamos muchas más opciones de traducción, algunas alternativas correctas, pero otras gramaticalmente erróneas. Tendríamos que hacer muchas reglas para controlar todo.

Las frases traducidas de español a inglés quedan de esta manera (en negrita la mejor):

- 01 [el,hombre,come,una,manzana]
 - A - [man,eats,an,apple]
 - B - [**the,man,eats,an,apple**]
- 02 [ellos,comen,manzanas]
 - A - [**they,eat,some,apples**]
 - B - [they,eat,apples]
- 03 [tu,comes,una,manzana,roja]
 - A - [**you,eat,a,red,apple**]

- 04 [juan,ama,a,maria]
 - A - [john,loves,mary]
- 05 [el,gato,grande,come,un,raton,gris]
 - A - [the,big,cat,eats,a,greymouse]
 - B - [the,large,cat,eats,a,greymouse]
- 06 [juan,estudia,en,la,universidad]
 - A - [john,studies,at,university]
 - B - [john,studies,in,university]
 - C - [john,studies,at,the,university]
 - D - [john,studies,in,the,university]
- 07 [el,alumno,ama,la,universidad]
 - A - [student,loves,university]
 - B - [student,loves,the,university]
 - C - [the,student,loves,university]
 - D - [the,student,loves,the,university]
- 08 [el,perro,persiguió,un,gato,negro,en,el,jardín]
 - A - [dog,chased,a,black,cat,at,garden]
 - B - [dog,chased,a,black,cat,in,garden]
 - C - [dog,chased,a,black,cat,at,the,garden]
 - D - [dog,chased,a,black,cat,in,the,garden]
 - E - [the,dog,chased,a,black,cat,at,garden]
 - F - [the,dog,chased,a,black,cat,in,garden]
 - G - [the,dog,chased,a,black,cat,at,the,garden]
 - H - [the,dog,chased,a,black,cat,in,the,garden]
- 09 [la,universidad,es,grande]
 - A - [university,is,big]
 - B - [university,is,large]
 - C - [the,university,is,big]
 - D - [the,university,is,large]
- 10 [el,hombre,que,vimos,ayer,es,mi,vecino]
 - A - [man,saw,yesterday,is,my,neighbour]
 - B - [man,we,saw,yesterday,is,my,neighbour]

- C - [man,that,saw,yesterday,is,my,neighbour]
 - D - [man,that,we,saw,yesterday,is,my,neighbour]
 - E - [the,man,saw,yesterday,is,my,neighbour]
 - F - [**the,man,we,saw,yesterday,is,my,neighbour**]
 - G - [the,man,that,saw,yesterday,is,my,neighbour]
 - H - [**the,man,that,we,saw,yesterday,is,my,neighbour**]
- 11 [el,canario,amarillo,canta,muy,bien]
 - A - [**the,yellow,canary,sings,very,well**]
 - 12 [juan,toma,un,cafe,y,lee,el,periodico]
 - A - [john,has,a,coffee,and,reads,newspaper]
 - B - [john,has,a,coffee,and,he,reads,newspaper]
 - C - [john,has,a,coffee,and,she,reads,newspaper]
 - D - [john,has,a,coffee,and,it,reads,newspaper]
 - E - [**john,has,a,coffee,and,reads,the,newspaper**]
 - F - [john,has,a,coffee,and,he,reads,the,newspaper]
 - G - [john,has,a,coffee,and,she,reads,the,newspaper]
 - H - [john,has,a,coffee,and,it,reads,the,newspaper]
 - 13 [juan,es,delgado,y,maria,es,alta]
 - A - [**john,is,thin,and,mary,is,tall**]
 - 14 [oscar,wilde,escribio,el,fantasma,de,canterville]
 - A - [**oscar,wilde,wrote,the,canterville,ghost**]

En el caso de traducir de español a inglés nos encontramos con otros problemas, el más común es poner la preposición *a* donde no debería escribirse. En la gramática inglesa hemos escrito esta regla:

$$g_preposicional_en(gp(pre(a), GN)) \longrightarrow g_nominal_en(GN, -, -, s).$$

Así podemos soportar [john,loves,mary]=[juan,ama,a,maria]. El problema es que añadimos errores a muchas otras frases, la solución podría ser concretar más esa regla para que tenga concordancia solo con algunos verbos o estructuras.

Las frases traducidas de inglés a español quedan de esta manera (en negrita la mejor):

- 01 [the,man,eats,an,apple]
 - A - [**el,hombre,come,una,manzana**]

- B - [el,hombre,come,a,una,manzana]
- 02 [they,eat,some,apples]
 - A - [comen,manzanas]
 - B - [comen,las,manzanas]
 - C - [comen,algunas,manzanas]
 - D - [ellos,comen,manzanas]
 - E - [ellos,comen,las,manzanas]
 - F - [ellas,comen,manzanas]
 - G - [ellas,comen,las,manzanas]
 - H - [ellos,comen,algunas,manzanas]
 - I - [ellas,comen,algunas,manzanas]
- 03 [you,eat,a,red,apple]
 - A - [comes,una,manzana,roja]
 - B - [comes,a,una,manzana,roja]
 - C - [tu,comes,una,manzana,roja]
 - D - [tu,comes,a,una,manzana,roja]
- 04 [john,loves,mary]
 - A - [juan,ama,maria]
 - B - [juan,ama,a,maria]
- 05 [the,big,cat,eats,a,grey,mouse]
 - A - [el,gato,grande,come,un,raton,gris]
 - B - [el,gato,grande,come,a,un,raton,gris]
- 06 [john,studies,at,the,university]
 - A - [juan,estudia,en,la,universidad]
- 07 [the,student,loves,university]
 - A - [el,alumno,ama,universidad]
 - B - [el,alumno,ama,la,universidad]
 - C - [el,alumno,ama,a,universidad]
 - D - [el,alumno,ama,a,la,universidad]
- 08 [the,dog,chased,a,black,cat,in,the,garden]
 - A - [el,perro,persiguió,un,gato,negro,en,el,jardin]

- 09 [the,university,is,large]
 - A - [la,universidad,es,grande]
- 10 [the,man,we,saw,yesterday,is,my,neighbour]
 - A - [el,hombre,que,vimos,ayer,es,mi,vecino]
 - B - [el,hombre,que,vimos,ayer,es,a,mi,vecino]
 - C - [el,hombre,que,nosotros,vimos,ayer,es,mi,vecino]
 - D - [el,hombre,que,nosotras,vimos,ayer,es,mi,vecino]
 - E - [el,hombre,que,nosotros,vimos,ayer,es,a,mi,vecino]
 - F - [el,hombre,que,nosotras,vimos,ayer,es,a,mi,vecino]
- 11 [the,yellow,canary,sings,very,well]
 - A - [el,canario,amarillo,canta,muy,bien]
- 12 [john,has,a,coffee,and,reads,the,newspaper]
 - A - [juan,toma,un,cafe,y,lee,el,periodico]
 - B - [juan,toma,un,cafe,y,lee,a,el,periodico]
 - C - [juan,toma,a,un,cafe,y,lee,el,periodico]
 - D - [juan,toma,a,un,cafe,y,lee,a,el,periodico]
- 13 [john,is,thin,and,mary,is,tall]
 - A - [juan,es,delgado,y,maria,es,alta]
- 14 [oscar,wilde,wrote,the,canterville,ghost]
 - A - [oscar,wilde,escribio,el,fantasma,de,canterville]
 - B - [oscar,wilde,escribio,a,el,fantasma,de,canterville]

5. Ajustes para soportar estructuras naturales

Como ya hemos visto en el apartado anterior, al generar muchas reglas específicas para casos concretos podemos acabar contaminando otras respuestas. Una solución que podemos pensar a priori es utilizar dos gramáticas de cada lenguaje, una que se usaría para enviar y otra para recibir, donde podamos definir con más soltura algunas reglas estructurales que pueden servir para esquematizar frases pero en ningún caso para construir frases válidas.

En el caso concreto anterior, si eliminamos la regla de la gramática española que añade el determinante, al traducir de español a inglés no tendremos esos sujetos omitidos en los resultados. Al igual que si eliminamos la regla que añade la preposición *a* en la gramática inglesa, al traducir español no tendríamos esa

preposición creando estructuras erróneas.

Otra solución es trabajar las reglas para conseguir una concordancia más estricta, que podamos usar para definir gran parte de las reglas gramaticales del idioma y evitar errores. Podemos imaginar que sería muy costoso ser totalmente riguroso para trabajar con todas las reglas gramaticales y sintácticas de un lenguaje natural.

En el caso concreto que se plantea, si en inglés el verbo *comer* en presente se utiliza en la forma continua podemos añadir esas palabras al diccionario y construir reglas de concordancia para que grupos verbales que usen esos verbos en español en tiempo presente, usen tiempo continuo en inglés, además de controlar la concordancia de las formas del verbo ser o estar.

Otra solución más fácil (aunque peor) es añadir el verbo ser al verbo comer en forma continua en el diccionario, de esta manera, realizando los cambios:

```
verbo(v(v_3), presente, 3, s) —> [come].
verbo(v(v_4), presente, 2, s) —> [comes].
verbo(v(v_5), presente, 3, p) —> [comen].

verbo_en(v(v_2), presente, 3, s) —> [sings].
verbo_en(v(v_3), presente, 3, s) —> [is,eating].
verbo_en(v(v_4), presente, 2, s) —> [are,eating].
verbo_en(v(v_5), presente, 3, p) —> [are,eating].
```

- 01 [el,hombre,come,una,manzana]
 - A - [the,man,is,eating,an,apple]
- 02 [ellos,comen,manzanas]
 - A - [they,are,eating,some,apples]
 - A - [they,are,eating,apples]
- 03 [tu,comes,una,manzana,roja]
 - A - [you,are,eating,a,red,apple]
- 05 [el,gato,grande,come,un,raton,gris]
 - A - [the,big,cat,is,eating,a,grey,mouse]
 - A - [the,large,cat,is,eating,a,grey,mouse]

6. Conclusiones

Podemos observar que este método, aunque pueda funcionar con más o menos soltura para un diccionario específico aplicado a unas frases de prueba, definir con reglas todas las posibles variantes del lenguaje natural para consultas de traducción más complejas puede ser costoso y difícil, ya que añadir una regla para una traducción específica puede perjudicar a otras reglas que antes funcionaban devolviendo resultados mejores, en estos ejemplos ha ocurrido. Si la solución es poner reglas muy específicas para casos concretos perdemos la eficiencia del sistema, tenemos que tener en cuenta que si una traducción busca objetividad, muy pocas reglas deben satisfacerse, así que si el programa engorda habrá muchas más que no se validarán hasta encontrar alguna correcta.