

Desarrollo de una gramática para aserciones
simples en español y su implementación en Prolog

Emmanuel Maggiori
emaggiori@alumnos.exa.unicen.edu.ar

Dirección

Silvia Schiaffino, Daniela Godoy {sschia, dgodoy}@exa.unicen.edu.ar

Universidad Nacional del Centro de la Provincia de Buenos Aires

Trabajo de Cátedra: Programación Exploratoria

Desarrollo de una gramática para aserciones simples en español y su implementación en Prolog

Resumen

En este trabajo se buscó explorar las virtudes de Prolog para el Procesamiento del Lenguaje Natural, haciendo uso de las *Definite-Clause Grammars* (DCG) para implementar una gramática del idioma español, que cubra aserciones sencillas, y con particularidades del español rioplatense. Ésta permite determinar, dada una oración, si ésta es o no correcta según la gramática, o bien generar oraciones que sean gramaticales.

Se deseaba que la gramática estuviera documentada paso a paso, registrando las decisiones que se tomaron para poder hacer un uso intensivo de las herramientas que proporcionan las DCG de Prolog para este fin.

El resultado es una gramática de ochenta reglas que permite generar oraciones simples, respetando las reglas elementales de la gramática española, como la concordancia entre el sujeto y el predicado (aun cuando el sujeto es compuesto) y permitiendo la inclusión de frases relativas, entre otras cosas, y que a la vez admite características propias del rioplatense como la redundancia pronominal de los objetos directos.

1. Introducción

Uno de los problemas de la lingüística computacional es el análisis sintáctico o *parsing*, por el cual, dada una oración, se debe descomponer en sus constituyentes sintácticos (en forma de árbol de *parsing*). En este trabajo se usa el análisis sintáctico con el fin de decidir si una oración es gramaticalmente correcta o no, o bien para generar oraciones que lo sean.

La corriente generativista efectúa esta tarea empleando gramáticas formales que intentan capturar las reglas gramaticales de la lengua natural [8].

El lenguaje Prolog nació en gran parte impulsado por el Procesamiento del Lenguaje Natural (NLP), aunque después se expandió y se convirtió en una herramienta útil para aplicaciones de Inteligencia Artificial en general. Una de sus virtudes es la facilidad para implementar *parsers* libres del contexto aumentados, muy útiles para el análisis sintáctico, mediante las *Definite-Clause Grammars* (DCG).

El objetivo de este trabajo es implementar en Prolog una gramática para el idioma español. Se desea contar con una gramática documentada paso a paso, con fines didácticos y para explorar las soluciones que aporta Prolog para esta

tarea en cada problema que se enfrenta. A la vez, se deseaba incluir particularidades de la variante rioplatense del español, que a veces suponen un desafío adicional a la hora de diseñar la gramática. Se buscó que la gramática cubra aseveraciones sencillas, en un número reducido de reglas, pero que se haga en cada regla un uso intensivo de las herramientas que aporta Prolog para el problema del análisis sintáctico.

La gramática completa se encuentra disponible en:
<http://PrologDCG-es.sourceforge.net/>.

El resto de este trabajo se organiza de la siguiente manera: en la sección 2 se introducen las herramientas de Prolog para el análisis sintáctico y, a continuación, la sección 3 describe la gramática implementada. Por último, en la sección 4 se presentan las conclusiones y trabajos futuros.

2. Análisis sintáctico en Prolog

Las llamadas gramáticas generativas [8] emplean gramáticas formales libres de contexto (aunque se ha trabajado con gramáticas más expresivas también [4]), como en el siguiente ejemplo:

```
oracion-->frase-nominal,frase-verbal.
```

Esta gramática podría generar, por ejemplo, YO TENGO FRÍO y NOSOTROS TENEMOS FRÍO, pero también YO TENEMOS FRÍO*. Si se desea que concuerden la frase nominal con la frase verbal, en persona y número, se debe multiplicar el número de reglas. Se contaría, entonces, con 18 reglas: las tres originales multiplicadas por las seis posibles combinaciones de persona y número. Para evitar esto, es muy tentador “anotar” la primera gramática para indicar la concordancia, en vez de multiplicar las reglas:

```
oracion-->frase-nominal(PERS,NUM),frase-verbal(PERS,NUM).
```

Se la conoce como gramática “aumentada” [4] y la implementación de un *parser* en Prolog para éstas es tan directa, que se formalizaron en las llamadas *Definite-Clause Grammars* (DCG) [6] y se soportan en los intérpretes más comunes (como SWI-Prolog, empleado en este trabajo). La regla anterior está escrita en el formato DCG de Prolog.

Además, se pueden incluir predicados comunes de Prolog en las reglas si estos se encierran entre llaves, que deben ser verdaderos para que se pueda aplicar la regla.

Como trabajo previo se encontró GATOS [9], que es una gramática implementada en Prolog, mediante DCG y muy abarcativa (demasiado a comparación de los fines de este trabajo), aunque genera en exceso. Por ejemplo, no evalúa la concordancia (admite YO VA AL CINE*). No se encontraron otros trabajos en los que se documente paso a paso el proceso de desarrollo de una gramática para el español.

También se encontró una gramática empleada en el proyecto Freeling [10], de la cual se cuenta con su código fuente. Es libre del contexto convencional (no aumentada) y no está implementada en Prolog.

3. Gramática implementada

En esta sección se describe la gramática. Se comienza por el léxico y, a continuación, se explican las reglas para derivar oraciones. Finalmente, se detallan las reglas para cada uno de los sintagmas (nominal, adjetival, proposicional, verbal y adverbial).

3.1. Léxico

La entrada del algoritmo de *parsing* consiste en una serie de etiquetas, que identifican a cada palabra morfológicamente (mediante un previo etiquetado o *PoS Tagging*). Las etiquetas empleadas en este trabajo son parcialmente coincidentes en sus categorías con las etiquetas Eagles [2] empleadas en el proyecto Freeling [10]. En el trabajo original se puede ver la lista completa de etiquetas. A continuación se presentan algunos ejemplos:

```
determinante(femenino,plural) --> [las].
determinante(masculino,singular) --> [el].
```

```
pro_personal(masculino,plural,primera,_) --> [nosotros].
pro_personal(_,singular,primera,nominativo) --> [yo].
pro_personal(_,singular,primera,oblicuo) --> [mi].
```

El último elemento corresponde al caso gramatical, que puede ser nominativo u oblicuo para los pronombres personales.

```
verbo(indicativo,_,plural,primera,comer) --> [comemos].
verbo(gerundio,_,_,_,comer)-->[comiendo].
diatesis(comer,transitivo).
copulativo(ser).
```

```
adjetivo(masculino,singular) --> [lindo].
adjetivo(_,singular) --> [grande].
```

```
adverbio(cantidad) --> [muy].
adverbio(afirmacion) --> [si].
```

3.2. Estructura general de la oración

La estructura de la oración se definió mediante las siguientes dos reglas:

```
oracion(AUSENCIA) -->
    sn_coord(GENERO,NUMERO,PERSONA,nominativo),
    sv_coord(indicativo,GENERO,NUMERO,PERSONA,AUSENCIA).
oracion(AUSENCIA) -->
    sv_coord(indicativo,_,_,_,AUSENCIA).
```

La primer regla expresa que una oración se compone de un sintagma nominal (o frase nominal) coordinado, seguido de un sintagma verbal coordinado. Se contempla la coordinación, es decir que el sintagma pueda ser compuesto: JUAN, MARÍA Y PEDRO COMEN POCHOCOLOS o JUAN O SU HIJO VAN AL CINE. Además, el caso del sintagma nominal debe ser nominativo por tratarse del sujeto: YO TENGO FRÍO y no MÍ TENGO FRÍO*.

La segunda regla contempla el sujeto tácito (la omisión del sintagma nominal): TENEMOS FRÍO en vez de NOSOTROS TENEMOS FRÍO. Es importante destacar que esta construcción no es admisible en otros idiomas como el inglés o el francés: AM HERE*, REVIENDRAI*.

La funcionalidad del parámetro AUSENCIA, tiene que ver con los *gaps* y se explica más adelante.

3.3. Sintagma nominal

3.3.1. Coordinación

En este trabajo se realiza la coordinación de forma recursiva, donde la condición de corte contiene los últimos dos elementos, separados por una conjunción. Los llamados recursivos van separando con comas (representadas por la palabra *coma* en esta gramática) hasta llegar a la condición de corte: JUAN, MARÍA, ..., PEDRO Y ADRIÁN.

Este problema en sintagmas nominales es más desafiante ya que se debe determinar el género y la persona correspondientes a la coordinación. En castellano, si todos los elementos son femeninos, la coordinación será femenina: MARÍA Y JUANA SON ALTAS. Caso contrario, será masculina: MARÍA Y PEDRO SON ALTOS.

Lo que introduce aun mayor dificultad es la persona. Si alguno de todos los elementos es primera persona, se conjuga en primera persona: MARÍA YO SOMOS GRANDES y no MARÍA Y YO SON GRANDES*. Si no hay ningún elemento en primera persona, basta con que haya uno en segunda persona para que se conjugue en segunda persona: ÉL Y TÚ SOIS GRANDES (contemplando el uso general del vosotros, inexistente en el español rioplatense). En cualquier otro caso (todo tercera persona), se conjuga en tercera: ÉL Y ELLA SON GRANDES.

En este problema, se debe hacer el control simultáneo de dos parámetros, género y persona, en las dos reglas (la recursiva y el corte). La primer solución que se implementó consistía en escribir el par de reglas varias veces, para las distintas posibles combinaciones: presencia de la primera persona y presencia de un masculino, presencia de la segunda persona y presencia de un masculino, etc.

Finalmente se recurrió a una solución muy simple. Primero se escribieron hechos de Prolog que representan el conocimiento previamente mencionado:

```
gen(masculino,_,masculino).
gen(femenino,X,X).
```

```
pers(primer,_,primera).
```

```

pers(segunda, primera, primera) .
pers(segunda, segunda, segunda) .
pers(segunda, tercera, segunda) .
pers(tercera, X, X) .

```

Luego se hizo uso de la posibilidad de incluirlos en las DCG entre llaves (éstos deben ser verdaderos para que se pueda aplicar la regla y, sin embargo, no forman parte de ésta). De esta manera se conservan únicamente las mismas dos reglas y se “calculan” la persona y el género, por ejemplo:

```

sn_coord(GENERO, plural, PERSONA, CASO) -->
  sn(GEN1, _, PER1, CASO) , y , sn(GEN2, _, PER2, CASO) ,
  {gen(GEN1, GEN2, GENERO) , pers(PER1, PER2, PERSONA)} .

```

Es interesante notar que de lo que hubieran sido unas decenas de reglas, se logró reducir el número a dos, gracias al uso de las llaves de las DCG. Además, es una notación muy natural, ya que se mantiene intacta la regla recursiva que genera los elementos de la coordinación y, aparte, se adiciona la información de la coordinación.

3.3.2. Estructura general

El sintagma nominal puede tomar tres formas distintas. La primera está formada por un determinante (la, una, unos, mis, nuestros, etc.), seguido de un sustantivo. A continuación pueden haber, opcionalmente, una oración de relativo, un sintagma adverbial, un sintagma proposicional y una aposición. Una oración que incluye todos estos elementos, podría ser LA CASA ENORME DE PABLO NERUDA QUE VISITÉ AYER, LA SEBASTIANA, ES IMPACTANTE, donde ES IMPACTANTE constituye el sintagma verbal.

La segunda involucra al pronombre personal (yo, él, etc.). Es necesario especificar el caso gramatical, ya que estos pronombres varían de acuerdo a él: YO ME DIJE A MÍ MISMO, pero no YO ME DIJE A YO MISMO*.

La tercera regla es el nombre propio (*María*).

La aposición (LA SEBASTIANA) es en este caso, simplemente, un sintagma nominal en caso nominativo, entre comas. Se incluyó para generar oraciones como JUAN, EL PORTERO, ES... o MI HIJO, MARTÍN, ES...

En el caso particular del sintagma proposicional, la regla se deriva permitiendo múltiples sintagmas: LA CASA DE MARÍA EN BUENOS AIRES ES...

3.3.3. Oraciones de relativo y uso de *gaps*

En los sintagmas nominales puede aparecer una oración de relativo. Se decidió implementar en este trabajo por ser un problema distinto al resto y así explorar la solución que aporta Prolog para este problema.

Una oración de relativo tiene un pronombre relativo (que, donde) seguido de otra frase: LA CASA QUE CONOZCO ES GRANDE, donde QUE CONOZCO es la oración de relativo. Se puede ver que ésta puede tener las características de

cualquier oración completa: LA CASA QUE MARÍA CONOCIÓ EN MADRID ES GRANDE. Sin embargo, existe la prohibición de que esta oración tenga un objeto directo, porque al ser relativa ya está haciendo referencia a un objeto anterior: si bien la oración MARÍA CONOCIÓ LA CASA EN MADRID es correcta, no lo es LA CASA QUE MARÍA CONOCIÓ LA CASA EN MADRID...*.

Este tipo de característica se denomina de largo alcance, ya que involucra parámetros muy alejados en el árbol de *parsing*. Una posible solución en Prolog involucra el uso de *gaps* [1] [3]. Un *gap* es un parámetro adicional que indica si es obligatoria o, al contrario, está prohibida la inclusión de cierto constituyente, como en este caso el objeto directo. Por esa razón la regla general de oración mencionada anteriormente incluye una variable *AUSENCIA*, que puede tomar los valores *presencia* o *ausencia*, siendo este último el que se emplea en la regla para relativos:

```
rel(GENERO,NUMERO)-->pron_rel(GENERO,NUMERO),oracion(ausencia).
```

En cualquier otro contexto se deberá usar *oracion(presencia)*. La regla para las oraciones, como se vio anteriormente, se deriva como un sintagma nominal y un sintagma verbal, y a este último se le “pasa” el parámetro *AUSENCIA* que se tratará más adelante en el desarrollo del sintagma verbal.

Ésta es una solución muy simple que aporta Prolog para tratar rasgos de largo alcance.

3.4. Sintagma adjetival

La coordinación de sintagmas adjetivales es análoga a aquella del sintagma nominal: incluye una condición de corte con una conjunción, y otra regla recursiva que separa con comas. Sin embargo, no presenta las dificultades que había en el caso anterior, ya que es obligatorio que los sintagmas adjetivales coordinados tengan el mismo género y número: FEROS Y MUY AGRESIVO y no FEROS Y MUY AGRESIVOS* (no se pueden “mezclar” como en MARÍA, ÉL Y NOSOTROS). Además del adjetivo, se admite opcionalmente un adverbio de cantidad: MUY LOCO.

3.5. Sintagma proposicional

La coordinación de sintagmas preposicionales se admite siempre que éstos tengan la misma preposición: DE JUAN, DE PEDRO Y DE MARÍA o EN BUENOS AIRES, EN PARÍS Y EN NUEVA YORK. Para eso se incluyó una variable *PREP* en las reglas de la coordinación, para controlar que todos los sintagmas preposicionales que forman parte de la coordinación, estén introducidos por la misma preposición. La única regla para el sintagma preposicional, más allá de las reglas para coordinaciones, es

```
sp(PREP)-->preposicion(PREP),sn_coord(_,_,_ ,oblicuo).
```

Es interesante ver que se pide el caso oblicuo, ya que rige éste ante una preposición (PARA MÍ y no PARA YO*).

La gramática genera DE JUAN Y DE MARÍA, pero también DE JUAN Y MARÍA, ya que en este último caso se trataría de una coordinación a nivel sintagma nominal, mientras que en el primero a nivel sintagma proposicional.

3.6. Sintagma verbal

3.6.1. Preadverbios

Se incluyó la posibilidad de que haya un “adverbio” opcional antes de los verbos, si éste es de negación o afirmación, para generar, por ejemplo, YO TAMBIÉN VOY, pero no YO RÁPIDAMENTE VOY (que no sería la forma más natural de decirlo).

3.6.2. Predicativo subjetivo obligatorio

Una de las reglas para el sintagma verbal tiene que ver con aquellas frases con un verbo copulativo (*ser, parecer, asemejar*). En esta caso, se admite luego del verbo un sintagma adjetival como complemento: LA CASA ES FRÍA. Este tipo de derivación no sería admisible si el verbo no fuera copulativo: MARÍA TIENE FRÍA*. En el léxico se puede incluir la regla *copulativo(LEXEMA)* para este fin (por ejemplo, *copulativo(ser)*), y la regla sólo se puede aplicar si es verdadera la evaluación de *copulativo(LEXEMA)*, encerrado entre llaves en la regla.

3.6.3. Acerca de los objetos

La diátesis de un verbo indica el número de objetos que debe o puede recibir. En este trabajo se consideraron las siguientes posibilidades:

Intransitivo No se admite ningún objeto. Por ejemplo, *morir, vivir*. Algunos verbos normalmente considerados intransitivos pueden no serlo estrictamente, por ejemplo *caminar* (JUAN CAMINA LAS CALLES DE ROMA).

Transitivo Admiten un único objeto, llamado objeto directo (que representa sobre quién recae la acción que se ejecuta). Por ejemplo, *comer* (JUAN COME UNA TORTA). Sin embargo, esta categoría admite que el objeto se omita (JUAN COME). Se representa la opcionalidad de un objeto directo.

Transitivo Estricto Es obligatorio el objeto directo, como en *tener*.

Ditransitivo El verbo ditransitivo obligatoriamente pide dos objetos, uno directo y uno indirecto, siendo este último el beneficiario de la acción, o quien recibe las consecuencias. Por ejemplo, *dar* (JUAN LE DA EL LIBRO A MARÍA).

Existen otras situaciones de diátesis (como verbos que piden únicamente un objeto indirecto), pero son situaciones muy particulares y por eso no se incluyeron en este trabajo. Un ejemplo es el verbo *quedar*: EL PANTALÓN ME QUEDA BIEN.

Un objeto se puede sustituir por un pronombre: JUAN LO VIO, JUAN LE ENVIÓ UNA CARTA. En el caso de los objetos directos se sustituye por *lo, la, los, las, nos, etc* (aunque existen excepciones en algunas variantes del español). En los objetos indirectos se sustituye por *le, les, nos, os, etc*.

En el caso de que se sustituyan ambos objetos a la vez, el pronombre que reemplaza al objeto indirecto se transforma en *se*, en caso de ser en tercera persona. Por ejemplo, es correcto SE LO DIJE y no LE LO DIJE* (cacofonía).

En las siguientes secciones se presentan las reglas para cada una de las combinaciones posibles de objetos.

3.6.4. Ausencia de objetos

```
sv(MODO, GENERO, NUMERO, PERSONA, AUSENCIA) -->
  pre_adv, auxiliares(MODO, GENERO, NUMERO, PERSONA, _, PIDE),
  verbo(PIDE, GENERO, NUMERO, PERSONA, LEXEMA), op_sa,
  {diatesis(LEXEMA, transitivo); diatesis(LEXEMA, intransitivo)}.
```

Mediante esta regla se generan frases verbales que carecen de objetos, como MARÍA MIRA, y se admiten únicamente ante verbos intransitivos y transitivos (ya que en éstos últimos el complemento es opcional). Estas condiciones se expresan entre las llaves de la regla.

3.6.5. Presencia de un objeto

Para frases con un único objeto se escribieron tres reglas, análogas a la anterior, donde la condición es que el verbo sea transitivo o transitivo estricto.

En la primer regla se adiciona *od(AUSENCIA)* para repretar al objeto directo, y con la posibilidad de que éste deba ser obligatoriamente inexistente (un *gap* para oraciones de relativo, como se vio anteriormente). Mediante esa regla se reconoce MARÍA MIRA LA CASA.

En la segunda regla se agregó *pron_od(AUSENCIA)* para generar oraciones como MARÍA LA MIRA, donde el objeto está reemplazado por un pronombre de objeto directo.

La tercera y última regla se incluyó para representar un fenómeno del español rioplatense que no es estándar en otras variedades, en donde se efectúa una redundancia incluyendo tanto al objeto como al pronombre: MARÍA LA MIRÓ A LA CASA. Se incluyó por ser una particularidad importante del español en esta región. El fenómeno tiene algunas particularidades semánticas que no se pudieron capturar en la gramática: en el caso particular de Argentina, la redundancia no se permite cuando el objeto no es identificado y conocido (MARÍA LA MIRÓ A ALGUNA CASA*) [5].

Además de agregar los terminales correspondientes al pronombre y al objeto, éste último tiene parámetros adicionales (*od(AUSENCIA, GEN, NUM, PERS)*) para que exista una concordancia requerida por la redundancia: MARÍA LA MIRA A LA CASA y no MARÍA LO MIRA A LA CASA*.

Los predicados para el objeto directo se derivan como una coordinación de sintagmas nominales en caso oblicuo, con la opcionalidad de ser precedidos por la

preposición *a*. Se decidió que fuera opcional porque es muy difícil determinar si ésta corresponde o no, siendo la elección de naturaleza semántica. En el español se usa la preposición cuando el objeto tiene carácter de persona (aunque muchas veces las cosas se personifican y las personas se cosifican): MIRO A MARÍA, o bien BUSCO UN TAXISTA/BUSCO A UN TAXISTA.

Además, se adicionó una regla para cuando el objeto debe tener concordancia con el pronombre (para la regla del español rioplatense), en la cual la preposición *a* es obligatoria ya que así es en esta región cuando se produce la redundancia: MARÍA LA MIRA A LA CASA pero nunca MARÍA LA MIRA LA CASA*.

3.6.6. Presencia de dos objetos

Cuando la diátesis es ditransitiva, se requiere la presencia de dos objetos. Las reglas incluidas son análogas a las anteriores, y permiten reconocer las frases MARÍA LEE EL LIBRO A JUAN, MARÍA LE LEE EL LIBRO y la redundancia también típica del rioplatense MARÍA LE LEE EL LIBRO A JUAN, en la cual también es necesaria la concordancia entre el objeto y el pronombre. Además, se incluyeron reglas para reconocer MARÍA SE LO LEE y la versión redundante MARÍA SE LO LEE A JUAN. Como se mencionó anteriormente, en estos dos últimos casos se emplea *se* en lugar de *le/les*, por lo que se incluyó una regla intermedia *pron_oi* para representar esta situación:

```
pron_oi(GENERO,NUMERO,tercera)-->[se].
pron_oi(GENERO,NUMERO,PERSONA)-->
  pro_pers_atono(GENERO,NUMERO,PERSONA,dativo),
  {not(PERSONA = tercera)}.
```

El objeto indirecto se deriva como una coordinación de sintagmas nominales, en caso oblicuo, siempre precedida por la preposición *a*:

3.6.7. Auxiliares

En la gramática se decidió incluir la posibilidad de adjuntar una serie de auxiliares o modales a los verbos: MARÍA QUERRÍA HABER PODIDO IR.

En las reglas se incluye información de qué “pide” el auxiliar anterior, para obligar a que el verbo en cuestión tenga el modo adecuado. La información de qué “piden” los verbos auxiliares según su categoría se incluyó en unos hechos que reflejan ese conocimiento:

```
pide(modal,infinitivo).
pide(haber,participio).
pide(estar,gerundio).
```

Esto expresa que los verbos modales piden que el verbo siguiente esté en infinitivo (QUIERE IR) mientras que el auxiliar *haber* pide participio (HA IDO) y *estar* pide gerundio (ESTÁ YENDO).

De esta manera se anidan los distintos auxiliares, cada uno limitando el modo del siguiente y a la vez evitándose la repetición de verbos de la misma

categoría. Se consideraron como categorías las tres mencionadas anteriormente: *estar*, *haber* y *modal*.

3.7. Sintagma adverbial

Para el sintagma adverbial se consideraron reglas que involucran la presencia de un adverbio de modo (opcionalmente precedido por uno de cantidad), de tiempo o lugar. De esta manera se pueden generar, por ejemplo, RÁPIDAMENTE, MUY RÁPIDAMENTE, AYER y AHÍ.

4. Conclusiones y trabajos futuros

La gramática implementada cuenta con ochenta reglas. Se lograron capturar generalidades de la gramática del español y particularidades del rioplatense.

Prolog permite agregar nuevas reglas de manera flexible y así aumentar fácilmente la expresividad de la gramática. Un posible uso de esto podría ser para incluir perífrasis verbales, como en SE PUSO A LLORAR.

No se incluyeron verbos reflexivos y pronominales, aunque se podrían incluir fácilmente adicionando algunas reglas en la derivación de los sintagmas verbales.

También se podría aumentar la expresividad de la gramática mediante reglas inspiradas en las gramáticas transformacionales [7], identificando determinada estructura de una oración, transformándola y repreguntando por la oración transformada. Por ejemplo, para incluir la voz pasiva se puede escribir una regla que transforme las oraciones pasivas en activas.

La gramática fracasa en cubrir constituyentes complejos. Por ejemplo, JUAN ESTÁ AHÍ DONDE TERMINA LA CALLE, que cuenta con un sintagma adverbial con una estructura mucho más compleja que la que cubre la gramática implementada, pero no por eso es agramatical, o la frase adjetival de ES UN PROBLEMA HISTÓRICAMENTE DIFÍCIL DE RESOLVER.

Por ahora el etiquetado de las palabras se supone hecho, y se incluyen las reglas en la gramática para el vocabulario que se desee trabajar. Se podría integrar con algún corpus lingüístico para realizar el etiquetado.

En este trabajo se exploró a Prolog como una herramienta para hacer NLP y se conocieron y aprovecharon sus virtudes para este fin. Sus características declarativas hacen que la tarea de diseñar sea sencilla y las cuestiones técnicas de implementación queden encapsuladas.

La gramática resultante cuenta con ochenta reglas y cubre aserciones simples, que cuentan con constituyentes coordinados (JUAN, MARÍA Y PEDRO...), permiten oraciones de relativo (LA CASA QUE YO CONOZCO...) y admiten redundancia pronominal como es típico en el español rioplatense (MARÍA LO QUIERE A JUAN), entre otras cosas.

Numerosos problemas se pudieron solucionar de manera sencilla haciendo un buen uso de la sintaxis de las DCG de Prolog, reduciendo el número de reglas necesarias.

A medida que se avanza, se empieza a enfrentar el problema de tener que adicionar muchas reglas para cubrir casos muy particulares, por lo que la gramática se vuelve poco poderosa para oraciones complejas. Por esta razón en la actualidad la mayoría de los problemas de NLP se resuelven mediante técnicas probabilísticas, como las PCFG donde las reglas se aprenden de un corpus de árboles de sintaxis.

Se espera que este trabajo pueda servir como referencia para comprender la utilidad de Prolog para NLP y a la vez contar con una gramática generativa concisa y documentada del idioma español.

Referencias

- [1] V. Dahl. Hiding complexity from the casual writer of parsers. In *Proceedings of the First International Workshop on Natural Language Understanding and Logic Programming (Rennes, France)*, ed. by Veronica Dahl and Patrick Saint-Dizier, pages 1–19, 1985.
- [2] Proyecto Freeling (Universitat Politècnica de Catalunya). Introducción a las etiquetas eagles (v. 2.0). Disponible en <http://nlp.lsi.upc.edu/freeling/>.
- [3] W. Frey. Noun phrases in lexical functional grammar. In *Natural Language Understanding and Natural Language Understanding Workshop*, pages 121–137, 1985.
- [4] R. Grishman. *Introducción a la lingüística computacional*. Visor, 1991.
- [5] Justo Fernández López. Duplicado del complemento directo e indirecto – la redundancia pronominal. Disponible en <http://hispanoteca.eu/>.
- [6] F.C.N. Pereira and D.H.D. Warren. Definite clause grammars for language analysis—a survey of the formalism and a comparison with augmented transition networks. *Artificial intelligence*, 13(3):231–278, 1980.
- [7] M. Pilleux and H.U. Cárdenas. *Introducción a la gramática transformacional en español*. Facultad de Letras y Educación, Universidad Austral de Chile, 1976.
- [8] N. Ruwet, E. Bombín, and M.M. Hernández. *Introducción a la gramática generativa*. Gredos, 1974.
- [9] CLIMENT S. and MORÉ J. (dc)gatos. a definite clause grammar to parse corpora of spanish. Technical Report 2, Reports de Recerca de la Universitat de Barcelona, 1997.
- [10] Montserrat Civit Tortuella. Gramática para freeing. Código fuente disponible en <http://nlp.lsi.upc.edu/freeling/>.