

# **Comprensión de Problemas de Programación en LN**

M.F. Verdejo

K. Sarasola

Informatika Fakultatea  
Euskal Herriko Unibertsitatea

COMPRESION DE PROBLEMAS DE PROGRAMACION EN  
LENGUAJE NATURAL

K. Sarasola

Informatika Fakultatea, Euskal Herriko Unibertsitatea  
649 postapartatua, 20080 Donostia-San Sebastián

M.F. Verdejo

Facultat d'Informàtica, Universitat Politècnica Catalunya  
Pau Gargallo 5, 08028 Barcelona

1. INTRODUCCION.

El sistema CAPRATE tiene como objetivo el tratamiento automático de problemas enunciados en lenguaje natural para expresarlos formalmente mediante un lenguaje de especificación. Nuestro trabajo se enmarca dentro de un proyecto más amplio: un sistema ICAI para la enseñanza de programación (CAPRA [Garijo & Verdejo, 84a]) siguiendo una metodología de refinamientos sucesivos.

La traducción se divide en dos fases consecutivas: la primera toma como entrada textos en español o euskara de los problemas de programación y produce representaciones del significado de los objetos que aparecen en el texto. La segunda fase consiste en la creación de la especificación después de comprobar que la representación obtenida es correcta.

El primer paso se lleva a cabo con ATNs complementadas con tratamiento semántico y relativo al dominio. El intérprete de ATNs ha sido diseñado para aceptar la definición standard dada por [Bates, 78] enriquecida con acciones y predicados creados para manejar explícitamente el conocimiento semántico. Las nuevas acciones permiten tanto buscar referencias usando categorías semánticas e información contextual como construir representaciones de los objetos sus propiedades y relaciones en el lenguaje RDES [Garijo & Verdejo, 84b] basado en redes semánticas extendidas.

El conocimiento del dominio se usa de forma extensiva en el segundo paso de la traducción para tener en cuenta condiciones globales de consistencia. Este proceso completa la definición de los objetos y decide cuáles formarán parte del input de la especificación y cuáles del output.

Tipos similares de aplicaciones en otros dominios son [Burton, 75] y [Mellish, 83]. Nuestro dominio presenta más ambigüedad que el primero y usa diferentes lenguas (español y euskara) y formalismos (para análisis y representación) que el segundo.

## 2. DEFINICION DEL DOMINIO PROBLEMA.

El objetivo del proyecto CAPRA es el desarrollo de un sistema inteligente para la enseñanza de programación a un nivel elemental.

El sistema esta compuesto por los siguientes módulos:

- Un tutor con interfaz en LN.
- Un experto en construcción de programas [Garijo 86]
- Una base de conocimiento que modela el conocimiento y comportamiento del alumno.

La metodología de enseñanza sigue una estrategia standard para diseño de programas, comenzando con una especificación para obtener un algoritmo abstracto y produciendo a continuación un programa por refinamientos sucesivos. El tutor guía el proceso de aprendizaje de los alumnos y controla la adquisición de conocimientos. El módulo experto está gestionado por el tutor con dos objetivos diferentes: como consejero explica qué refinamientos son precisos para escribir un programa estructurado; como crítico sigue la actividad del alumno, haciéndole notar sus errores.

El módulo CAPRATE presentado en este artículo tiene por objeto interpretar los textos de problemas de programación que manejan objetos tales como números, caracteres, secuencias y conjuntos. El añadir la posibilidad de traducción de enunciados desde LN al lenguaje de especificación permite al sistema construir y enriquecer fácilmente la colección de problemas.

Un conjunto de 100 problemas, escritos en español y euskara, representativos del dominio total han sido recogidos como muestra. Para un primer desarrollo se han eliminado los enunciados que poseen cuantificadores y aquellos en los que define alguna función. A continuación se muestran 4 ejemplos de enunciado:

- 1) Dada una secuencia de M caracteres, escribir el número de comas que tiene dicha secuencia.
- 2) Dados 50 números enteros, calcular su media aritmética.
- 3) Dada una secuencia de 20 números enteros, obtener otra con los elementos pares de S y determinar su longitud.
- 4) Sea S una secuencia de enteros y M un entero. Diseñar un algoritmo para escribir todos los números de S que acaben en M y sean múltiplos de dicho número.

El lenguaje de especificación objeto de esta traducción ofrece un formalismo estructurado para definir los objetos dando su nombre tipo y propiedades. A modo de ejemplo, la especificación correspondiente al último de los enunciados propuestos es la siguiente:

ENTRADA: Sea S secuencia\_de ent, M ent

SALIDA : Escribir R secuencia\_de ent definida: R S formado\_por todos\_los x de S tal\_que ( mod(x,M)=0 & mod(x,10)=M ).

### 3. COMPONENTES DEL SISTEMA.

Los componentes del sistema pueden ser clasificados como sigue:

- a) Herramientas generales útiles para cualquier otra aplicación.
  - RDES es el lenguaje de representación del conocimiento referente al dominio en el proyecto CAPRA. Se basa en el manejo de redes semánticas particionadas.
  - El intérprete de ATNs está basado en la descripción standard de [Bates,78]. La estrategia seguida por el intérprete explora todos los análisis posibles intentando eliminar cuanto antes las posibilidades incorrectas, utilizando para ello información semánticas. Los núcleos de las unidades de significado (verbos en oraciones, nombres en grupos nominales, ...) crean expectativas sobre los requisitos semánticos de sus complementos. Un proceso de filtrado seleccionará para cada expectativa entre todos los posibles candidatos, con el objeto de llevar a cabo esta labor de lanzamiento de posibilidades se han definido 3 nuevos tipos de acciones y un predicado, todos ellos utilizables por las ATNs.
- b) Conocimiento no lingüístico relativo al dominio de aplicación.
  - Conjunto de funciones para el manejo de la representación de los objetos. Creación de un objeto, Añadir una propiedad, nombre o tipo a un objeto, obtención o comprobación de una característica sobre un objeto ya representado, y obtención de una lista con los objetos ya representados que son de un tipo o cumplen una característica concreta.
  - Reglas de inferencia que expresan el conocimiento sobre la completitud de la representación final.
- c) Conocimiento relativo al dominio ligado con conocimiento lingüístico.
  - Diccionario con información sintáctica y semántica.
  - Analizador léxico. Toma como entrada un enunciado completo y devuelve una secuencia de unidades lexicales con información extra no accesible desde el diccionario. Hay cinco tipos de unidades lexicales: Signos de puntuación, Identificadores, Expresiones, Formas verbales y Palabras normales con entrada en el diccionario.
  - ATNs sintáctico-semánticas para el análisis del español y euskara.

#### 4. JUSTIFICACION POR MEDIO DE EJEMPLOS.

Esta sección presenta algunos tratamientos llevados a cabo por el sistema:

a) Interpretación guiada por información basada en casos semánticos.

... el productorio desde 1 hasta 10 de  $(i+1)*(i+2)/i$   
Esta frase genera cinco posibles análisis sintácticos correctos dependiendo de las diferentes asociaciones de los grupos preposicionales. El nombre "productorio" tiene definido en el diccionario un único uso semántico:

```
[productorio ...[sem [[[ GPREP desde] [elem ent] [let x1]]  
                    [[[ GPREP hasta] [elem ent] [let x2]]  
                    [[[ GPREP de]      [expr num] [let x3]]] ...  
]
```

"productorio" necesita tres grupos preposicionales. Uno representando un entero introducido con el caso "desde", otro representado por un entero introducido con el caso "hasta" y un tercero representando una expresión numérica introducida por el caso "de". Solo uno de los 5 análisis sintácticos correctos será filtrado por el intérprete.

```
[ el productorio [de 1] [hasta 10] [de  $(i+1)*(i+2)/i$  ] ]
```

b) Interpretación guiada por información del dominio sobre propiedades.

... calcular el número de comas que tiene dicha secuencia.  
En este caso resulta insuficiente pedir una categoría semántica determinada para el GPREP de "número".

Además de la categoría (elem simple), el grupo preposicional habrá de cumplir dos propiedades generales (valor definido y miembro de una secuencia):

```
[número ... [sem [[[GPREP de] [elem ] [let x]  
                  [valor_definido x] [miembro_sec x]  
                  [let S [sec_prop x]] ...]
```

donde *valor\_definido* es un predicado que verifica si *x* tiene un valor definido previamente, *miembro\_sec* verifica que *x* tiene la propiedad de pertenencia y *sec\_prop* es una función que devuelve la secuencia que contiene el elemento *x*. Todas estas funciones pueden ser definidas fácilmente usando el conjunto de funciones primitivas provistas por el lenguaje de representación.

c) Pronombres y resolución de referencias.

...de dicha secuencia.

El adjetivo "dicha" es una referencia a un objeto que ha sido definido previamente. En el análisis se desencadenará la búsqueda de un objeto previamente representado y perteneciente a la misma categoría semántica del nombre. ("secuencia" en este caso).

... su media aritmética.

El adjetivo "su" define una referencia no para el objeto sino para un complemento suyo. El uso semántico de "media aritmética" establece que se necesita un GPREP introducido por el caso "de" representando una secuencia de números. La aparición del determinante "su", que referencia a un objeto previo, desencadena la búsqueda de un objeto perteneciente a la categoría del GPREP necesariamente introducido con el caso "de".

d) Elipsis.

... calcular la media aritmética.

De una forma similar al ejemplo anterior de resolución de referencias, cuando se detecta que el uso semántico está incompleto (falta la secuencia), pero el grupo nominal es determinado se busca entre los objetos ya representados alguno que sea de la categoría semántica precisa para rellenar los complementos necesarios.

e) Examen de la correcta definición de objetos.

Dada la secuencia de M enteros ...

Si M no se especifica posteriormente, al final su valor aparecerá como indeterminado. Criterios pragmáticos permiten inferenciar al término del análisis que M será un valor dato de entrada.

## 5. ESTADO ACTUAL Y CONCLUSIONES.

El sistema CAPRATE comprende problemas de programación elementales escritos en español y vasco. Actúa como un módulo de adquisición del tutor para construir una colección de problemas. Esta herramienta es general pudiendo ser utilizada también en la construcción de la interfaz de LN que conduce el diálogo entre un alumno y el tutor.

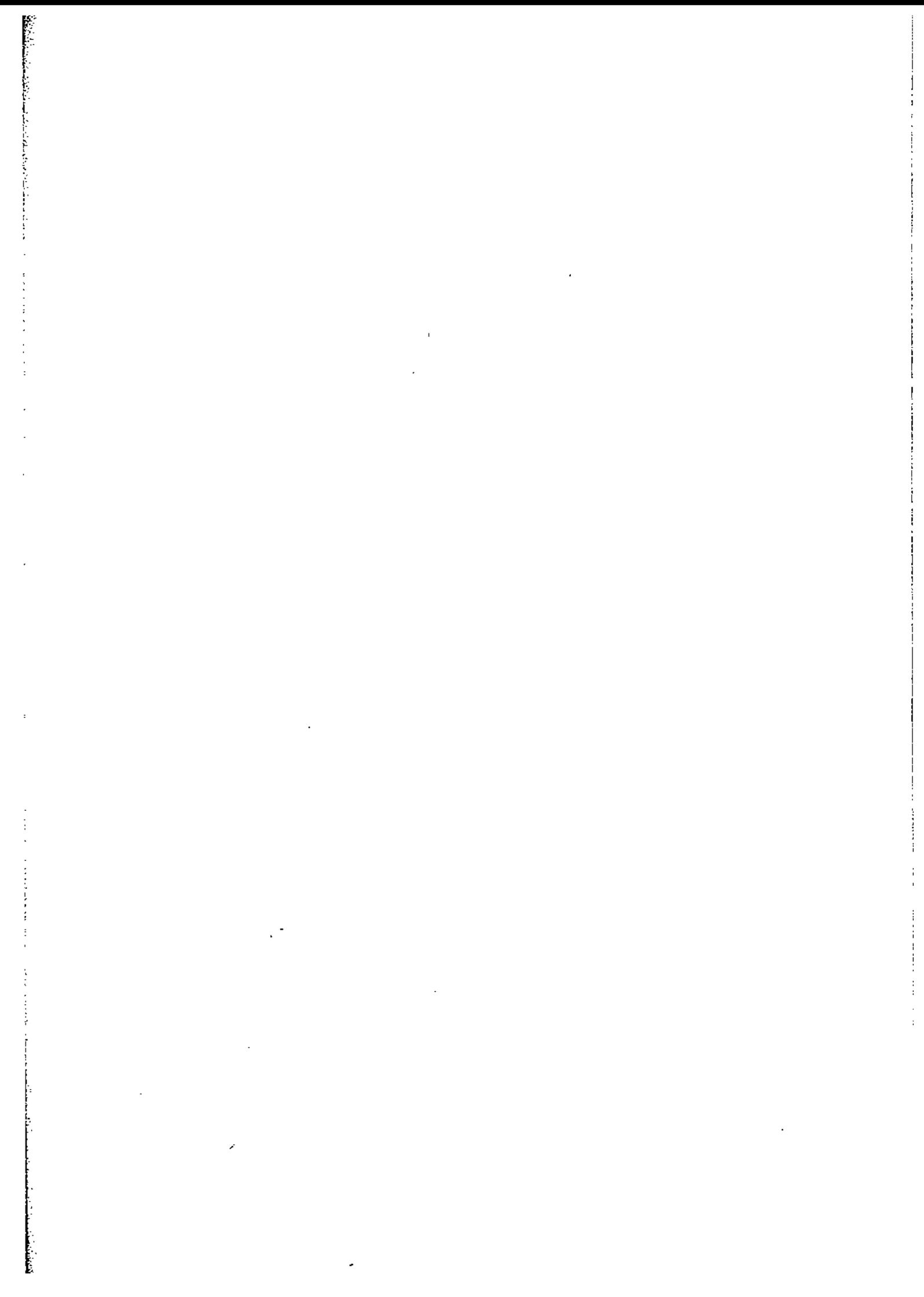
El sistema interpreta textos que presentan referencias, elipsis y ambigüedad. Nuestra aproximación integra conocimientos sintácticos, semánticos, contextuales y del dominio para conseguir un análisis incremental. Un intérprete de ATNs extendidos ha sido implementado.

Un estudio posterior habrá de realizarse sobre los enunciados que poseen cuantificadores y definición de nuevas funciones.

### Referencias:

- [Bates 78] "The theory and practice of Augmented Transition Network Grammars", in: Bolc, L.(ed), "Natural Language Communication with Computers", Berlin 1978.
- [Burton 76] "Semantic grammar": an engineering technique for constructing natural language understanding systems. Report 3422, Bolt Beranek and Newman Inc, Cambridge Mass.

- [Garijo and Verdejo, 84a] "CAPRA: an intelligent system for teaching programming concepts". Internal report. Facultad de Informática de San Sebastián. FISS-S-S.1-IT/P.
- [Garijo and Verdejo, 84b] "Knowledge representation for teaching programming in an ICAI environment". First IEEE conference on Artificial Intelligence Applications. Denver. Colorado.
- [Garijo and, Hernández, Inchausti, Verdejo, 86] "SYNTALAB: an expert system for the synthesis of abstract algorithms. The 6th Int. W. on expert system & their applications. Avignon.
- [Mellish 83] "Incremental semantic interpretation in a modular parsing system." in Automatic Natural Language Parsing, K. Sparck Jones & Y. Wilks eds. Ellis Horwood Limited. pg 148-155



# INSCRIPCION

## SOCIEDAD ESPAÑOLA PARA EL PROCESAMIENTO DEL LENGUAJE NATURAL

D.....

Dirección Postal

Calle:.....

Código postal:.....Ciudad.....

Teléfono.....

Centro de trabajo.....

Area de investigación o interés.....

Solicita su inscripción como socio de la S.E.P.L.N., para lo cual adjunta talón nominativo por valor de 1500 pst, correspondiente a la cuota del presente año.

Fecha.....Firma

ENVIAR A:

Jorge Garcia Sanz  
Secretaria SEPLN  
Clavel 12, bajo 3  
Rivas-vaciamadrid  
Madrid