Método para la elaboración de un corpus para la GLN

Guadalupe Aguado de Cea
Departamento de Lingüística Aplicada
Facultad de Informática, UPM.
lupe@fi.upm.es

Mª del Socorro Bernardos Galindo
Laboratorio de Inteligencia Artificial
Facultad de Informática, UPM.
sgalindo@delicias.dia.fi.upm.es

Resumen.

Una de las dificultades de la construcción de sistemas de GLN es la especificación de requisitos. Una forma de abordar este problema es usar un corpus para mostrar al cliente el sistema que se pretende construir. En este artículo se describe un método para elaborar ese corpus y se explica cómo se ha utilizado en un proyecto concreto. Este método consta de 5 pasos: recopilación de textos, determinación de las entradas, análisis de los textos recopilados y de las entradas, construcción de un corpus, y extracción de patrones. Finalmente, existe una etapa de búsqueda de similaridades que no forma parte de la elaboración del corpus en sí, pero que está íntimamente relacionada con los patrones y con la que se pretende reducir el tiempo de desarrollo del sistema.

1 Introducción.

La construcción de un sistema de generación de lenguaje natural (GLN) implica la realización de múltiples tareas, como el estudio de los fenómenos lingüísticos que ha de tratar, el diseño de su arquitectura, la construcción de una gramática, etc. Una de las tareas básicas (y de las primeras en abordarse) es realizar un análisis de requisitos para producir una especificación inicial del sistema. Para esto se pueden utilizar técnicas generales de análisis de requisitos aplicadas en ingeniería del software ([Pressman, 1993], por ejemplo), pero puede resultar mucho más interesante utilizar métodos específicos para sistemas de GLN. El más representativo es el de Reiter y Dale [1997], quienes describen un enfoque en el que el desarrollador usa un corpus para describir a los usuarios el sistema que se propone construir. Su ventaja es que es más fácil hablar de la funcionalidad con los usuarios enseñándoles ejemplos que hablando sobre la GLN de modo más abstracto, sobre todo porque muchos usuarios no tienen experiencia previa con la tecnología de GLN. En este artículo se presenta un método para la elaboración de un corpus basado en ese enfoque.

Previo a la descripción del método, conviene explicar qué es lo que se entiende por corpus:

Según el Diccionario de la Real Academia [DRAE 1992], un corpus es un “conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación”. Tomando esta definición como base, y añadiendo matices propios del campo de la GLN (recogidos en [Dale y Reiter, 1997]), un corpus es un conjunto de ejemplos de **textos de salida** y **datos de entrada** que especifica la funcionalidad del sistema de GLN propuesto.

La razón de incluir los datos de entrada es que no tiene sentido que formen parte del corpus de un sistema textos que nunca podrá generar, es decir, textos que no correspondan a ninguna entrada.

La importancia del corpus no radica sólo en que sirve para especificar requisitos del sistema, sino en que puede ser un recurso muy útil en el estudio de los fenómenos lingüísticos a los que se ha de enfrentar el sistema de GLN.

2 Método.

El proceso propuesto aquí para la elaboración del corpus de un sistema de GLN se lleva a cabo en 5 pasos: recopilación de textos de salida, determinación de las entradas, análisis de los textos recopilados y de las entradas, construcción de un corpus, y extracción de patrones. Con los cuatro primeros se crea el corpus, primero se obtiene un conjunto de textos de salida y otro de entradas, y luego se hace un análisis de estos conjuntos que se usa después para dar lugar al corpus, donde cada salida tiene asociada una entrada. La extracción de patrones sólo es útil si se pretende reutilizar el trabajo.

2.1 Recopilación de textos de salida.

Su finalidad es obtener un conjunto de textos, propios del dominio de la aplicación, que cubra el ámbito completo de los textos que se espera que produzca el sistema de GLN. Los textos pueden proceder de libros de consulta, expertos en el dominio a los que se puede pedir que produzcan ejemplos de textos, y documentos o informes hechos en el pasado.

En esta etapa interesa tener un gran número de textos. No importa tanto su validez, como que den una muestra lo más completa posible de todas las posibilidades que se pueden presentar como texto de salida.

2.2 Determinación de las posibles entradas.

Sirve para determinar el conjunto de entradas a las que ha de hacer frente el sistema de GLN. Basado en la experiencia de las autoras, la entrada de un sistema de GLN consta de 4 componentes: el **objetivo comunicativo** que se quiere lograr, el **modelo de**l **usuario** al que está dirigido el texto, la **historia del discurso** que se ha producido, y la **solicitud** del usuario del sistema.

Un *objetivo comunicativo* se puede definir como lo que se quiere expresar y para qué. Los objetivos dirigen el proceso de generación: determinan lo que se debe seleccionar de la fuente de información, cómo se debe estructurar el texto, qué palabras se deben utilizar si hay más de una posibilidad, etc. [Bañón, 1999].

El *modelo de usuario* es una caracterización del receptor para el que se genera el texto. Entre los muchos aspectos que se pueden incluir en un modelo de usuario están la información sobre el grado de maestría del usuario, su tarea y sus preferencias. Esto permite personalizar el texto que se genere.

La *historia del discurso* es un modelo de las interacciones previas entre el usuario y el sistema de GLN. Se utiliza para que la salida se pueda generar teniendo en cuenta lo que ya se ha generado.

La *solicitud del usuario* está constituida por lo que el usuario del sistema puede introducir en una interacción concreta. Puede variar mucho dependiendo del tipo de sistema. Por ejemplo, en un sistema de generación de documentos, sería la petición de generación de un documento concreto (como el informe meteorológico de un día, una semana o un mes, etc.); en un sistema de consulta, se tendría la petición que se realiza al sistema, etc.

No siempre se dan todos estos componentes, puede haber sistemas en los que no se tenga en cuenta el tipo de usuario que los utiliza, o no importe repetir lo ya expresado, o el objetivo comunicativo sea siempre el mismo, etc.

Esta etapa puede hacerse previa o simultáneamente a la de recopilación de textos.

2.3 Análisis de los textos y de las entradas.

El objetivo de esta etapa es conseguir una comprensión detallada de las correspondencias entre los textos de salida y los datos de la fuente de información (base de datos, base de conocimientos, etc.) de la que se disponga, y entre los textos de salida y las entradas.

Lo primero se logra analizando el contenido de los textos. Para ello, se pueden clasificar las cláusulas de cada texto en las categorías que proponen Reiter y Dale [1997]:

- **Texto inmutable**: Un fragmento de texto que siempre está presente en los textos de salida.
- **Datos disponibles directamente**: Texto con información disponible directamente en los datos de entrada o en una fuente de información asociada.
- **Datos computables**: Texto con información que se puede deducir de los datos de entrada mediante un cálculo o un razonamiento.
- **Datos no disponibles**: Texto con información que no está presente ni se puede deducir de los datos de entrada.

Estos últimos textos originan la mayor parte de los problemas, ya que no se puede obtener la información necesaria para generarlos. Esto se puede solucionar haciendo que el sistema de GLN tenga más información disponible, cambiando los textos para eliminar las partes afectadas, y si el sistema de GLN se está usando como ayuda, esperando que una persona escriba esos fragmentos.

El estudio de las relaciones entre los textos de salida y las entradas, servirá para establecer las correspondencias entre ambos, y para descubrir textos sin ninguna entrada asociada y viceversa.

También hay que analizar la forma de los textos. Es necesario averiguar si para una entrada existen textos con el mismo contenido, pero muy diferentes en su estilo (por ejemplo, porque pertenezcan a distintos expertos) y estudiar cómo son estos estilos. Además, hay que detectar los textos que no son óptimos (por ejemplo, porque son de difícil comprensión) y que pueden mejorarse.

2.4 Construcción de un corpus.

A partir del análisis efectuado en la etapa anterior, y de los criterios que establezcan los desarrolladores y los clientes se determina un corpus con los textos que el sistema de GLN va a ser capaz de generar, junto con la entrada a la que van a estar asociados.

Existen varias posibilidades:

- Descartar o modificar los textos que hagan referencia a datos que no están disponibles de forma manejable.
- Simplificar los textos difíciles o imposibles de generar para que sean más sencillos.
- Añadir entradas para cubrir los textos que no tenían correspondencia con ninguna entrada, o bien no incluirlos en el corpus meta.
- Añadir textos para responder a las entradas que habían quedado sin ningún tipo de texto asociado, o bien no considerar esas entradas en la aplicación.
- Modificar los textos que sean difíciles de comprender o en los que se piense que se pueden añadir mejoras.
- Resolver conflictos, como que los expertos sugieran textos muy diferentes para la misma entrada.

Es conveniente realizar el corpus en colaboración con los expertos (lingüistas y del dominio) y los clientes, cada uno en los momentos que sea adecuado. Los expertos servirán para garantizar que sea correcto (gramaticalmente y dentro del dominio) y la discusión con los clientes servirá para tomar las decisiones relativas a la eliminación, modificación o inclusión de un texto.

En la mayoría de las ocasiones el corpus obtenido no será definitivo, pues la escasa experiencia en el campo de la GLN hace que sea difícil establecer desde el principio qué es lo que se puede conseguir, es decir, qué fenómenos lingüísticos se logrará tratar, en qué contexto y con qué profundidad.

2.5 Extracción de patrones.

A partir del corpus y del análisis efectuado anteriormente se construyen patrones de los textos que se van a generar.

Esta etapa sólo es útil en el caso de que se pretenda que el sistema sea lo más independiente posible del dominio de la aplicación. En estos casos el uso de los patrones como base para la generación servirá para que, además de generarse correctamente los textos del corpus, también se puedan generar textos con la misma estructura aunque correspondan a otros dominios.

En primer lugar hay que establecer la estructura de cada texto que forma parte del corpus sustituyendo las palabras propias del dominio por otras generales como concepto, propiedad, etc. Los textos resultantes pueden ser considerados como los "esqueletos" de esos textos iniciales. Seguidamente, se ha de acordar una notación para representar esos patrones mediante estructuras repetitivas, optativas, partes fijas, etc. Por último se agrupan los distintos "esqueletos" de los textos del corpus utilizando la notación acordada.

Como ya se ha indicado, es muy probable que el corpus meta sufra continuas modificaciones a lo largo de la construcción del sistema. En los casos en los que se han extraído patrones, lo más recomendable es realizar esas modificaciones en los patrones, de este modo además de mantener la independencia que se busca se evita hacer dos trabajos: uno para modificar los textos y otro para reflejar esos cambios en los patrones.

2.6 Búsqueda de similares.

Esta etapa no es parte de la elaboración del corpus propiamente dicha. La razón para mencionarla es que está muy relacionada con los patrones extraídos anteriormente y parece que el momento más adecuado para realizarla es una vez terminado el corpus .

Consiste en examinar los patrones de otros proyectos y ver si se pueden aplicar a los textos del corpus con el que se está trabajando.

Con esta tarea se pretende ahorrar trabajo y tiempo en el desarrollo del sistema. Siempre que se considere conveniente, se podrán aprovechar los resultados de las tareas realizadas en otros proyectos correspondientes a patrones similares. En el mejor de los casos sólo se necesitará incluir el lexicón del nuevo dominio para obtener los textos deseados.

3 Elaboración de un corpus en ONTOGENERATION.

El método descrito anteriormente se aplicó en el proyecto ONTOGENERATION [Aguado *et al.*, 1998] pero adaptándolo a la aplicación particular. En concreto, los aspectos que tuvieron una mayor influencia fueron:

- La fuente de información estaba predeterminada, se utilizó la ontología *Chemicals* [Fernández, 1996]. Esto hizo que el conjunto de textos quedara bastante delimitado y que incluso en un primer momento se pudiera descartar mucho material inútil.
- Se pretendía que el resultado pudiera ser empleado con otras fuentes de información, fundamentalmente con otras ontologías.

A continuación se explica el modo en el que se procedió.

3.1 Recopilación de textos.

En primer lugar se examinaron distintos libros sobre química con el fin de recoger textos referidos a los elementos químicos y sus propiedades.

En un primer examen, se observó que en muchos casos los textos eran demasiado extensos e incluían mucha información que no estaba ni iba a estar disponible en *Chemicals*. Por eso se decidió no utilizar los textos de los libros tal como estaban, sino que se optó por eliminar algunas partes y crear textos nuevos a partir de otros.

Cada miembro del equipo preparó un conjunto de posibles textos y después se hizo una puesta en común de los distintos grupos de textos elaborados. Dado que esos textos no eran exactamente los que aparecían en los documentos y libros que se habían utilizado, se acudió a un experto para que validara su adecuación y corrección dentro del dominio de la química.

3.2 Determinación de las posibles entradas.

Se determinó la siguiente entrada:

- *Objetivos comunicativos*: El sistema tiene un único objetivo comunicativo: proporcionar toda la información pedida por el usuario.
- *Modelos de usuario*: No existen variaciones en los textos generados debido a las diferencias entre distintos usuarios.
- *Historia del discurso*: No existe. Cada vez que se haga una consulta se supone que ésta está aislada de las anteriores.
- *Solicitudes del usuario*: Este componente de la entrada está formado por todas las consultas que el usuario puede hacer al sistema de GLN. Estas consultas pueden ser relativas a cualquiera de los conocimientos incluidos en la ontología *Chemicals* (conceptos, instancias, atributos, axiomas y funciones).

Como se ve, la entrada del sistema de generación estará formada únicamente por la consulta realizada. Esto es lo que se estudió con más detalle.

Se dividieron las consultas en 5 categorías:

- Descripción de los grupos químicos.
- Descripción de las fórmulas.
- Descripción de los elementos químicos.
- Comparación de los elementos químicos.
- Descripción de las propiedades de los elementos químicos.

A partir de esas 5 categorías se definieron los tipos de peticiones, 13 en total. Las consultas concretas no se tomaron hasta que se estableció la correspondencia entre los textos de salida y las entradas.

3.3 Análisis de los textos y de las entradas.

En esta etapa para cada tipo de consulta, se llevó a cabo una discusión sobre cuáles de los textos recopilados debían incluirse como respuesta a la misma.

A cada texto se le podía asociar más de una petición. Así, a la oración “Los metales de no transición se dividen en alcalinos y alcalinotérreos”, se asociaron las consultas “Definición de un concepto”, “Clasificación de un concepto” y “Descomposición de un concepto”.

Por otro lado, el análisis del contenido de los textos reveló que, con respecto a la clasificación de Reiter y Dale [1997] se tenía lo siguiente:

- **Textos inmutables**: No hay cláusulas enteras que estén siempre presentes de la misma manera.
- **Datos disponibles directamente** y **datos computables**: No es necesario realizar cálculos sobre los datos de *Chemicals*, excepto en expresiones que hacen referencia a datos incluidos en la descripción de los distintos términos, como “El argón fue descubierto por Lord Rayleigh y Sir William Ramsay en 1894.” o “El arsénico es venenoso.” Estas oraciones son sólo un ejemplo de la enorme cantidad de cláusulas que presentan estas características. Cabe destacar, que los textos que las contienen no están incluidos como respuesta a ninguna de las entradas determinadas anteriormente.

- **Datos no disponibles**: Aparecen textos como "Los no reactivos están clasificados según su característica metálica, así se tendrán no metales, semimetales y metales.", que son imposibles de generar debido a que en la ontología *Chemicals* no existen atributos de concepto (sólo contiene atributos de instancia). Esto impide acceder a un atributo que, por ejemplo, haga referencia a la "característica metálica".

  Por otro lado, también se detectaron datos no disponibles en el caso de las oraciones que expresan el valor cualitativo de una propiedad, como por ejemplo, "La electronegatividad del bromo es muy alta y tiene un valor de 2,8". Dado que la ontología *Chemicals* no proporciona datos que sirvan para poder afirmar si un valor es "alto", "bajo" o "medio".

3.4 Construcción de un corpus.

El primer corpus contenía el conjunto de textos recopilados en el primer paso, clasificados según los tipos de consultas establecidos en el segundo paso, mediante el análisis del tercer paso al que se eliminaron los textos que :

- Hacían referencia a atributos de conceptos.
- Contenían datos computables relacionados con el campo de descripción.

Según el análisis efectuado, el otro tipo de cláusulas que podía presentar problemas era el que expresaba valores cualitativos de una propiedad de un elemento químico. Tras un estudio de la ontología y de las distintas opciones, se decidió incluir en la ontología baremos que permitieran generar ese tipo de expresiones.

3.5 Extracción de patrones.

Como se quería que el resultado final fuera independiente de la fuente de información utilizada, se decidió construir patrones de modo que los textos se pudieran generalizar a otras fuentes, fundamentalmente ontologías.

A → B	A tiene la estructura indicada en B
[A]	Elemento opcional
[A | B | C]	A o B o C al menos uno de éstos
$\{A\}^m_n$	A se repite entre n y m veces
A → B | C	A tiene la estructura indicada en B o en C
negrita	palabras fijas
cursiva	palabras variables

Tabla 1: Notación de los patrones.

A partir de la notación reflejada en la Tabla 1 se definieron 95 patrones, la mayoría correspondientes a más de una oración. Por ejemplo, con el patrón:

> *Artículo_det nombre_concepto* {, *artículo_det nombre_concepto*}$^n_0$ **y** *artículo_det nombre_concepto* {**forman** *artículo_det* | **forman parte de** *artículo_det* | **forman un tipo general llamado** | **se incluyen en** *artículo_det* | **forman parte de un grupo general llamado** | **componen una** [**sub**]**división de** *artículo_det*} *nombre_concepto_padre*.

se pueden generar frases como:

- "La primera serie de transición, la segunda serie de transición, la tercera serie de transición y los actínidos forman los metales de transición."
- "Los halógenos se incluyen en los no metales."

3.6 Evolución del corpus.

Una vez elaborados los patrones, se trabajó con ellos directamente cada vez que se realizaba algún cambio.

La primera modificación fue básicamente de organización y motivada por el uso del entorno de generación KPML [Bateman, 1997] para construir la gramática[1]. KPML genera oraciones, no párrafos[2]. Así, los patrones se clasificaron en dos clases: oraciones y párrafos. Los párrafos se dividieron en oraciones. La consecución de estas oraciones se puede considerar como una meta intermedia que serviría para alcanzar el párrafo final.

Otros cambios se debieron a la relación entre textos/patrones y consultas. Tal y como se definieron los patrones en un primer momento, un mismo patrón podía estar asociado a varias peticiones de usuario. Para simplificar la selección de los patrones que debían responder a

[1] Para construir la gramática se siguió la teoría sistémico-funcional de Halliday [1985].

[2] Existen varias definiciones de párrafo y de oración. A efectos de este trabajo, con el término oración se hará referencia a la palabra o conjunto de palabras con que se expresa un sentido gramatical completo [DRAE, 1992] y se entenderá párrafo como cada una de las divisiones de un escrito señaladas por letra mayúscula al principio del renglón y punto y aparte al final del trozo de escritura [DRAE, 1992].

cada petición de usuario, se decidió asociar una única consulta a cada patrón.

Conviene destacar que las modificaciones del corpus se realizaron en continua interacción con los expertos del dominio (los expertos en lingüística ya formaban parte del equipo de desarrollo) para que dieran su conformidad, señalaran aquello con lo que no estaban de acuerdo o comentaran aquellos aspectos que se podían mejorar, para realizar las correcciones oportunas.

En resumen, los cambios producidos en el corpus inicial se debieron a:

1. Correcciones de los expertos.
2. Imposibilidad técnica de generar el texto.
3. Decisiones de diseño.

El corpus final puede verse en [Olalla, 1999]; en [Bernardos, 1997] y [Nieto, 1999] se muestra cuál era el corpus en distintos momentos del proyecto.

Para finalizar con esta sección, se incluye un ejemplo de una entrada y algunos de sus correspondientes textos de salida.

Entrada: Consulta sobre la propiedad densidad a 20 ºC.

*Salida1:*La densidad a 20 ºC se mide en gr/cm^3 y toma valores entre 0 y 25 gr/cm^3. Esta propiedad depende del peso atómico y del volumen atómico a 20 ºC según la fórmula: Densidad a 20 ºC = Peso atómico / Volumen atómico a 20 ºC. Además, sólo tiene un único valor para cada elemento.

Salida 2: La densidad a 20 ºC puede tener un solo valor para cada elemento químico y toma valores entre 0 y 25 gr/cm^3. Se mide en gr/cm^3. Esta propiedad depende del peso atómico y del volumen atómico a 20 ºC . La fórmula Densidad a 20 ºC = Peso atómico / Volumen atómico a 20ºC sirve para calcular la densidad a 20 ºC de los elementos químicos.

Salida 3: La densidad a 20 ºC se mide en gr/cm^3 y toma valores entre 0 y 25 gr/cm^3. Además, puede tener un único valor para cada elemento. Esta propiedad depende del peso atómico y del volumen atómico a 20 ºC . La fórmula Densidad a 20 ºC = Peso atómico / Volumen atómico a 20 ºC sirve para calcular la densidad a 20 ºC de los elementos químicos.

Salida 4: La densidad a 20 ºC puede tener un solo valor para cada elemento químico y se mide en gr/cm^3. Toma valores entre 0 y 25 gr/cm^3. . Esta propiedad depende del peso atómico y del volumen atómico a 20 ºC . La fórmula Densidad a 20 ºC = Peso atómico / Volumen atómico a 20ºC sirve para calcular la densidad a 20 ºC de los elementos químicos.

Como ya se explicó, la entrada en este caso sólo tiene uno de los cuatro componentes identificados en la sección 2.2. Si, por ejemplo se consideraran distintos tipos de usuario, los textos del ejemplo podrían corresponder a un usuario general, mientras que para un estudiante podrían incluirse ejemplos del valor de la densidad a 20ºC de distintos elementos, e incluso comparaciones entre esos valores.

4 Conclusiones y trabajo futuro.

Se ha presentado un método para construir un corpus dentro del área de la GLN. Este método sistematiza la tarea de análisis de requisitos relacionada con los datos de entrada y los textos de salida de un sistema de GLN. Cabe destacar, además, que el resultado obtenido (el corpus) no sólo sirve como requisito del sistema, sino que es muy útil para su diseño.

El método aquí descrito se distingue de otros en dos aspectos fundamentalmente:

- Se dedica una fase a averiguar las entradas del sistema. Estos datos de entrada se asocian a los textos de salida correspondientes en una fase posterior. De este modo el corpus no contendrá textos que el sistema nunca generará.
- Pone un especial énfasis en la reutilización. La utilización de patrones puede servir para que el producto obtenido pueda utilizarse en otros proyectos con características similares.

Se ha explicado cada fase y se han dado algunas pautas para realizarlo. Sin embargo, conviene señalar que las ideas aquí expuestas son de carácter general, habrá que adaptarlas a cada aplicación en concreto, como se ha mostrado para el proyecto ONTOGENERATION.

Quedaría por estudiar hasta qué punto se puede aplicar este método a sistemas de GLN diferentes del proyecto ONTOGENERATION, para ver en qué medida es necesario cambiar, eliminar o añadir algún paso. Con este fin se están planteando otros proyectos de GLN en los que se aplique el método propuesto aquí. Se pretende, además, que esos proyectos aborden textos del mismo tipo que los de ONTOGENERATION para comprobar cuál es la verdadera utilidad de los patrones extraídos.

5 Agradecimientos.

El nombre completo del proyecto ONTOGENERATION es "Sistema de consulta y recuperación de información mediante la generación de textos en español en el dominio de las sustancias químicas" y fue financiado por la UPM dentro del plan "Ayudas para la realización de proyectos de investigación y desarrollo, dirigidas a grupos potencialmente competitivos" (Referencia A9706).

En este proyecto participaron Guadalupe Aguado, Alberto Bañón, Socorro Bernardos, Asunción Gómez, Esperanza Nieto, Alberto Olalla y Álvaro Sánchez.

6 Referencias.

[Aguado *et al.*, 1998]: G. Aguado, A. Bañón, J. Bateman, S. Bernardos, M. Fernández, A. Gómez, E. Nieto, A. Olalla, R. Plaza, A. Sánchez. "Ontogeneration: Reusing Domain and Linguistic Ontologies for Spanish Text Generation". Presentado en *Workshop on Applications of Ontologies and Problem Solving Methods, ECAI'98*, Brighton (Reino Unido), 1998.

[Bañón, 1999]: A. Bañón. *Modelo de generación multisentencial EPRS*. Trabajo Fin de Carrera, Facultad de Informática, Universidad Politécnica de Madrid, Madrid (España), 1999.

[Bateman, 1997]: J.A. Bateman. "Enabling technology for multilingual natural language generation: the KPML development environment". *Natural Language Engineering*, 1: 1-42. Cambridge University Press, Cambridge (Reino Unido), 1997.

[Bernardos, 1997]: S. Bernardos. "GUME: Extensión de la Ontología GUM para el Español". Trabajo Fin de Carrera, Facultad de Informática, Universidad Politécnica de Madrid. Madrid (España), 1997.

[Dale y Reiter, 1997]: R. Dale y E. Reiter. "Tutorial on Building Applied Natural Language Generation Systems". *ANLP-97*, 1997.[Pressman, 1993]: R. Pressman. *Ingeniería del Software: Un Enfoque Práctico*. MCGraw-Hill, 3ª edición, 1993.

[DRAE 1992]: *Diccionario de la Lengua Española, RAE*. Espasa-Calpe, 1992.

[Fernández, 1996]: Fernández, M.. *Chemicals: Una Ontología de Elementos Químicos*, Trabajo Fin de Carrera, Facultad de Informática, Universidad Politécnica de Madrid (España), 1996.

[Halliday, 1985]: M. A. K. Halliday. *An Introduction to Functional Grammar*. Edward Arnold, Londres (Reino Unido), 1985.

[Nieto, 1999]: Esperanza Nieto. *Metodología para adaptar una gramática sistémica-funcional para la generación en castellano*. TFC, Facultad de Informática, Universidad Politécnica de Madrid (España), 1999.

[Olalla, 1999]: A. Olalla. *Sistema de GLN basado en ontologías: Ontogeneration*. Trabajo Fin de Carrera, Facultad de Informática, Universidad Politécnica de Madrid (España), 1999.

[Pressman, 1993]: R. Pressman. *Ingeniería del Software: Un Enfoque Práctico*. MCGraw-Hill, 3ª edición, 1993.

[Reiter y Dale, 1997]: E. Reiter, R. Dale. "Building Applied Natural Language Generation", *Journal of Natural Language Engineering*, 3(1): 57-87, 1997.