

Integración de proposiciones de infinitivo al sistema ClaTex

Aiala Rosá
Grupo de Procesamiento de Lenguaje Natural (PLN),
Instituto de Computación (InCo),
Facultad de Ingeniería, Universidad de la República,
Montevideo, Uruguay
aialar@fing.edu.uy

Julio de 2007

Resumen

El sistema ClaTex, desarrollado por el grupo de Procesamiento de Lenguaje Natural, segmenta oraciones en proposiciones, según una definición de proposición centrada en la presencia de un verbo conjugado. Este documento presenta una nueva versión del sistema que incluye proposiciones nucleadas por verbos en infinitivo. La extensión del sistema se realizó en base a un estudio teórico de los infinitivos como posibles núcleos de proposición y se efectivizó mediante la escritura de reglas contextuales que se incorporaron al conjunto de reglas original.

Palabras clave

Procesamiento del lenguaje natural, proposiciones, reglas contextuales.

1. INTRODUCCIÓN.....	3
2. USOS NOMINALES Y USOS VERBALES DE LOS INFINITIVOS.....	3
3. PERÍFRASIS Y LOCUCIONES CON INFINITIVOS	4
3.1 Pérfasis verbales de infinitivo.....	5
3.2 Locuciones con infinitivos.....	6
4. INCORPORACIÓN DE LAS PROPOSICIONES DE INFINITIVO AL SISTEMA	7
4.1 Descripción de las reglas contextuales.....	7
4.2 Reglas para infinitivos nominales	8
4.3 Reglas para perífrasis y locuciones	9
4.4 Reglas para proposiciones de infinitivo.....	12
4.5 Evaluación.....	13
5. CONCLUSIONES	15
BIBLIOGRAFÍA.....	17
ANEXO: LOCUCIONES ADVERBIALES, CONJUNTIVAS Y PREPOSITIVAS CON INFINITIVOS, RECONOCIDAS POR FREELING	18

1. Introducción

El sistema ClaTex (Caviglia 2003, Caviglia 2006) se aplica a textos en español para segmentar sus oraciones en proposiciones. La definición de proposición en la que se basa está centrada en la presencia de un verbo conjugado. Esta restricción permitió desarrollar una primera versión del sistema, quedando pendiente el reconocimiento de otras unidades predicativas posibles, nucleadas por: infinitivos, participios, gerundios, adjetivos, nominalizaciones.

El trabajo que se describe en este documento consiste en la extensión de ClaTex, mediante el reconocimiento de proposiciones de infinitivo. Como el sistema opera en base a un conjunto de reglas, su alcance se extiende mediante la incorporación de nuevas reglas.

Como primer paso en el reconocimiento de proposiciones de infinitivo, es necesario identificar algunos contextos en los cuales los infinitivos no constituyen núcleos de proposición, ya sea porque son utilizados con valor nominal, es decir, funcionando como nombres, o porque aparecen conformando perífrasis o locuciones. Para la identificación de estos infinitivos especiales se escribió un conjunto de reglas.

Los infinitivos restantes son los que nuclean las proposiciones que se pretende reconocer. Para armar estas proposiciones se escribió otro conjunto de reglas, basadas en las reglas que generan las proposiciones finitas.

En el capítulo 2 de este documento se presenta el análisis de los infinitivos nominales. En el capítulo 3 se estudian las perífrasis y locuciones con infinitivos. En el capítulo 4 se da una introducción al formalismo de reglas contextuales y se presentan las reglas definidas en el contexto de este trabajo. Finalmente, el capítulo 5 contiene las conclusiones.

En todas las etapas del trabajo se utilizaron los textos periodísticos del corpus CorIn (Grassi 2001) y fragmentos del corpus para consultas en línea Corpus del Español [CDE].

2. Usos nominales y usos verbales de los infinitivos.

Existe una gran discusión en torno a la categoría gramatical que les corresponde a los infinitivos. Éstos presentan muchas de las características de los nombres y también de los verbos. En este trabajo, se toma como referencia el artículo de Hernanz (Hernanz 1999), donde se sostiene que un infinitivo puede tener, o bien valor verbal, o bien valor nominal, dependiendo del contexto en el cual es utilizado.

En el diccionario de la Real Academia Española [DRAE] sólo algunos infinitivos tienen, además de la categoría verbo, una segunda entrada con la categoría nombre (como *poder*, *ser*, *atardecer*). Estos elementos, que en sus usos nominales Hernanz llama "falsos infinitivos", están lexicalizados como sustantivos. Su significado actual se aleja bastante del significado del infinitivo del que derivan y además perdieron el sentido de acción característico de los verbos.

Los infinitivos restantes, si bien tienen en el diccionario una única entrada como verbos, aparecen en ciertos contextos con valor nominal. Hernanz los llama "infinitivos nominales", para ellos es posible diferenciar las estructuras sintácticas en donde funcionan como verbos de las estructuras en donde funcionan como nombres. Este es el caso de *bromear* en el ejemplo 2.1¹, que presenta, claramente, características de grupo nominal, al ir acompañado por un determinante y un adjetivo.

¹ Todos los ejemplos de esta sección fueron tomados del Corpus del Español [CDE].

Ejemplo 2.1 Similar inoportunidad lució el Presidente en los predios vaticanos, con un **bromear** confianzudo, como de recién arribado al jet set europeo.

Para el trabajo que se expone en este documento, es indispensable identificar los infinitivos con valor verbal, ya que son los únicos que serán considerados núcleos de las proposiciones de infinitivo. *bromear*, en el ejemplo anterior, no debe ser tomado en cuenta a la hora de identificar las proposiciones.

En este sentido, fue necesario analizar el comportamiento del *tagger* FreeLing [FL], que se utiliza para pre-procesar los textos que toma como entrada el sistema ClaTex. En particular, importa saber cómo son etiquetados los infinitivos, pues sería deseable que solamente llevaran la categoría verbo aquellos que efectivamente se usan con valor verbal en el texto que se procesa.

El diccionario de FreeLing ofrece las dos etiquetas, verbo y nombre, para muchos de los infinitivos incluidos en su diccionario. Se realizó un pequeño estudio en una porción del Corin que contiene 462 ocurrencias de infinitivos con las 2 etiquetas. El 95.45% de estos infinitivos fue bien etiquetado. Este valor está dentro de lo esperado, si se tiene en cuenta que los *taggers* para el español suelen tener, para la asignación de etiquetas en general, una tasa de error del 5%.

Otros infinitivos siempre son etiquetados como verbos, por ser ésta la única opción contenida en el diccionario del *tagger*. Para estos casos se escribieron reglas contextuales que permiten reconocer los usos nominales. Las reglas se basan en la presencia de determinantes (ejemplo 2.2), adjetivos (ejemplo 2.3) y complementos genitivos (ejemplo 2.4); elementos que, generalmente, acompañan al nombre.

Ejemplo 2.2 Sois mi comer, mi beber, mi **vestir**.²

Ejemplo 2.3 .. su fuerza dramática y el fácil **fluir** de un idioma coloquial ..

Ejemplo 2.4 El **resurgir** del tapiz como arte en el siglo XIX ...

A diferencia de otros determinantes, el artículo definido *el* puede preceder tanto a un infinitivo nominal (ejemplos 2.3 y 2.4) como a un infinitivo verbal (ejemplos 2.5 y 2.6). Por lo tanto, no es evidencia de uso nominal de los infinitivos. En los ejemplos 2.3 y 2.4, el adjetivo y el genitivo son prueba de que *fluir* y *resurgir* están siendo usados con valor nominal, el artículo definido no alcanzaría por sí solo para desambiguar.

Ejemplo 2.5 .. lo que dificulta el **tener** ahora mismo una serie completa ..

Ejemplo 2.6 .. tuvieron como principal característica el **estar** dirigidos por grupos bereberes ..

En definitiva, para la elaboración de las reglas para el reconocimiento de infinitivos nominales, los elementos a tener en cuenta serán los determinantes, el determinante *el* (como caso particular), los adjetivos y los complementos de genitivo.

3. Perífrasis y locuciones con infinitivos

Los infinitivos que forman perífrasis o integran locuciones no constituyen núcleos de proposición, por lo que deberán ser descartados por las reglas que construyen las proposiciones de infinitivo.

² Sólo importa en este ejemplo el verbo *vestir*, que es el único que no tiene la opción nombre en el diccionario de FreeLing, ni en el DRAE. *comer* y *beber* sí tienen la categoría nombre y además son etiquetados correctamente por el *tagger*.

Sin embargo, una expresión que contiene un infinitivo y que se comporta “normalmente” como perífrasis o locución, en algún contexto particular, puede requerir una interpretación en donde los elementos que la componen funcionan en forma independiente.

La expresión <ir + a + infinitivo>, por ejemplo, constituye una perífrasis en el ejemplo 3.1, pero no en el ejemplo 3.2³.

Ejemplo 3.1 Yo **voy a votar** por la reforma, pero con poco entusiasmo, ...

Ejemplo 3.2 Escuchame, cuando yo **voy a coordinar** a Argentina, a hacer coordinación en Argentina,...

En los corpus estudiados, la proporción de usos no perifrásticos de estas expresiones es poco representativa. Por esta razón, no se aborda la tarea de identificar estos contextos especiales.

En las dos secciones que siguen se describe el trabajo de identificación de perífrasis y locuciones. En los dos casos, el reconocimiento se lleva a cabo mediante la aplicación de reglas. Queda abierta, por lo tanto, la lista de expresiones a reconocer: dada una nueva perífrasis o locución, basta generar una regla similar a las que se presentan aquí que la reconozca.

3.1 Pérfasis verbales de infinitivo

En la tabla 3.1 se muestra una lista de perífrasis verbales de infinitivo, extraída de (Gómez Torrego 1999).

Tabla 3.1

Perífrasis modales	deber + infinitivo
	deber de + infinitivo
	tener que + infinitivo
	haber de + infinitivo
	haber que + infinitivo
	poder + infinitivo
Perífrasis aspectuales	ir a + infinitivo
	empezar a + infinitivo
	comenzar a + infinitivo
	ponerse a + infinitivo
	echar a + infinitivo
	echarse a + infinitivo
	romper a + infinitivo
	estar a punto de + infinitivo
	volver a + infinitivo
	soler + infinitivo
	acostumbrar + infinitivo
	acabar de + infinitivo
	terminar de + infinitivo
	dejar de + infinitivo
	parar de + infinitivo
	cesar de + infinitivo
Otras perífrasis	venir a + infinitivo
	llegar a + infinitivo
	acertar a + infinitivo
	alcanzar a + infinitivo
	acabar por + infinitivo
	terminar por + infinitivo
	tardar en + infinitivo
	no tener por qué + infinitivo
	no haber por qué + infinitivo

Para todas las expresiones de la tabla se definieron reglas que las reconocen y etiquetan como unidades verbales. Como consecuencia, en la etapa de reconocimiento de proposiciones, estas expresiones darán lugar a una única proposición, con la perífrasis como

³ Ejemplos tomados del CorIn.

núcleo. Por lo tanto, para la oración del ejemplo 3.3, no se marcarán dos proposiciones como se ve en el ejemplo 3.4, sino una sola, como muestra el ejemplo 3.5.

Ejemplo 3.3 ... la búsqueda errática **suele conducir** al fracaso ...

Ejemplo 3.4 [... la búsqueda errática
[**suele**
[**conducir** al fracaso ...
]
]
]

Ejemplo 3.5 [... la búsqueda errática
[**suele conducir** al fracaso ...
]
]

3.2 Locuciones con infinitivos

Las locuciones con infinitivos, al igual que las perífrasis, deben ser detectadas por el sistema para dejar inhabilitados como núcleos de proposición a los infinitivos que ellas contienen. Diferentes tipos de locuciones pueden contener infinitivos entre sus componentes: locuciones verbales (*dar que hablar*), locuciones adverbiales (*es decir*), locuciones prepositivas (*al decir de*) y locuciones conjuntivas (*a no ser que*).

Por un lado, las locuciones verbales con infinitivos deben ser consideradas como unidades, de otro modo, podrían dar origen a dos proposiciones: una nucleada por el verbo conjugado y otra nucleada por el infinitivo. El ejemplo 3.6, extraído del CorIn, contiene una de estas locuciones: *querer decir*.

Ejemplo 3.6 Estoy buscando a alguien que me pueda explicar qué **quiere decir**.

Un análisis de la oración anterior donde no se reconoce la locución como unidad, como se muestra en el ejemplo 3.7, no es correcto.

Ejemplo 3.7 [Estoy buscando a alguien
[que me pueda explicar
[qué **quiere**
[**decir**
]
]
].

Escribiendo reglas para marcar las locuciones con infinitivos, se obtiene el siguiente análisis, que sí es correcto:

Ejemplo 3.8 [Estoy buscando a alguien
[que me pueda explicar
[qué **quiere decir**
]
]
].

En la sección 51.1.6 de (Gómez Torrego 1999), se presentan varias locuciones verbales con infinitivos. La tabla 3.2 contiene las locuciones que se listan en dicho artículo. Para marcarlas, se escribieron reglas que las agrupan y etiquetan como unidades verbales.

Tabla 3.2

querer decir
dar que pensar
dar que hablar
dar a conocer
dar a entender
dar de mamar
dar de beber
dar de comer
hacer saber
hacer llegar
echar a perder
dejar caer

Por otro lado, las locuciones adverbiales, conjuntivas y prepositivas que contienen infinitivos también deben ser identificadas. A diferencia de las locuciones verbales, muchas de estas locuciones son reconocidas y etiquetadas como unidades por el *tagger* FreeLing⁴. Sin embargo, se relevaron de los corpus varias locuciones adverbiales con infinitivos que no son reconocidas por el *tagger*. Para estas locuciones, que se muestran en la tabla 3.3, se escribieron reglas. Las reglas las marcan como unidades y les asignan la etiqueta adverbio.

Tabla 3.3

es decir
es un decir
vale decir
a saber
como ser
de nunca acabar
por así decir
(muy) a mí/tu/su/... pesar
ni que decir

4. Incorporación de las proposiciones de infinitivo al sistema

4.1 Descripción de las reglas contextuales

El sistema ClaTeX funciona en base a la aplicación secuencial de reglas que funcionan a nivel de la oración y asignan etiquetas a segmentos de texto que cumplen determinadas condiciones. Las reglas se agrupan en módulos. El texto es recorrido varias veces en su totalidad. En cada recorrida se aplican todas las reglas definidas dentro de un mismo módulo. La agrupación de reglas en módulos y el orden en que éstos se ejecutan dependen de qué etiquetas involucra cada regla. Así, una regla que se basa en la presencia de un elemento con capacidad de introducir una proposición (por ejemplo, un *que* etiquetado como *iniProp*), debe pertenecer necesariamente a un módulo posterior al que contiene las reglas que deciden cuáles son los elementos con capacidad de introducir proposiciones (por ejemplo, las reglas que asignan la etiqueta *iniProp* a determinadas ocurrencias de *que*⁵).

Las etiquetas, además de asignar una categoría al segmento que engloban, contienen atributos que permiten manejar información sobre ese segmento. Por ejemplo, para la etiqueta *uniVerb*, que se utiliza para marcar segmentos de texto que conforman una unidad verbal, se manejan los atributos que se muestran en la tabla 4.1.

⁴ La lista completa de locuciones con infinitivos que reconoce el *tagger* se encuentra en el anexo de este documento.

⁵ Sobre este punto concreto, puede verse (Caviglia 2003), en donde se describen en detalle todas las reglas para la desambiguación de *que* como *iniProp*.

Tabla 4.1

Etiqueta uniVerb	
Atributo	Descripción
lema	lema del verbo (infinitivo)
forma	indica si el verbo es finito, infinitivo, participio pasado o gerundio
modo	para verbos finitos, indica el modo
tiempo	para verbos finitos, indica el tiempo
negación	indica si la unidad verbal contiene adverbio de negación
sujeto	indica si la unidad verbal contiene un pronombre sujeto
persona	para verbos finitos, indica la persona
número	indica el número
discurso	indica si el verbo es de discurso (decir, explicar, ...)
interrogación	indica si el verbo puede introducir interrogación indirecta (preguntar, saber, ...)
clítico directo	indica si el verbo contiene un clítico directo (enclítico o proclítico)
clítico indirecto	indica si el verbo contiene un clítico indirecto (enclítico o proclítico)

Uno de los módulos del sistema contiene reglas que definen las unidades verbales, es decir, agrupan al verbo con determinadas partículas que van siempre asociadas a él: pronombres clíticos, pronombre sujeto y adverbio de negación. Estas reglas asignan la etiqueta *uniVerb* con diferentes valores en sus atributos. En el ejemplo 4.1 se muestra un ejemplo de regla para unidad verbal.

Ejemplo 4.1 regla v1: uniVerb --> \neg uniVerb/

Esta regla indica que al segmento conformado por un adverbio de negación y un verbo (el cual fue etiquetado como *uniVerb* en una etapa anterior) se le asigna la etiqueta *uniVerb*, es decir, se extiende una unidad verbal existente para incorporarle el adverbio de negación. Como se ve en el ejemplo, se indica con barras el principio (\) y el fin (/) del segmento que debe ser etiquetado con la categoría que aparece a la izquierda de la flecha. Esta nueva etiqueta *uniVerb* tendrá, a diferencia de la original, el valor **sí** para el atributo **negación**, indicando de este modo que se trata de una unidad verbal que contiene el adverbio de negación.

Una descripción detallada del formalismo de reglas contextuales puede encontrarse en (Wonsever 2004).

4.2 Reglas para infinitivos nominales

En el capítulo 2, se presentaron los elementos que, cuando modifican un infinitivo, dan evidencia de su valor nominal: determinantes, adjetivos, complementos con *de*. Desde un punto de vista lingüístico, esta evidencia alcanzaría para identificar usos nominales de infinitivos.

Sin embargo, cuando se pretende realizar un reconocimiento automático de estos elementos, se plantean algunos problemas de ambigüedad que no están presentes para un análisis humano. Un adjetivo que se encuentra inmediatamente antes o después de un infinitivo no siempre modifica al infinitivo. En el ejemplo 4.2, el adjetivo *solo* está predicando sobre el sujeto (no explícito) de *estar* y no sobre el propio infinitivo.

Ejemplo 4.2 Era mejor estar solo, después de todo.

Del mismo modo, un grupo introducido por la preposición *de* puede no estar funcionando como complemento del infinitivo nominal sino como complemento preposicional de un infinitivo verbal, como es el caso del ejemplo 4.3.

Ejemplo 4.3 la intervención del gobierno para salir de una depresión fuerte

En los dos casos, a partir del análisis de los dos corpus de estudio, se concluyó que las ocurrencias de elementos de este tipo modificando a los infinitivos nominales no es más frecuente que la de estos mismos elementos cumpliendo otras funciones. Por esta razón, la simple presencia de adjetivos o complementos con *de* no es suficiente para marcar infinitivos nominales.

Por otro lado, la gran mayoría de los infinitivos que, además de presentar adjetivos o complementos con *de*, están precedidos por el artículo *el*, tienen valor nominal. Por lo tanto, por razones estadísticas, se van a marcar infinitivos que ocurren en esos contextos como nominales.

Los infinitivos precedidos por un determinante, excluyendo el determinante *el*, serán marcados siempre como infinitivos nominales.

En la tabla 4.2, se muestran las reglas para el reconocimiento de infinitivos nominales, finalmente implementadas.

Tabla 4.2

Esquema reconocido por la regla	Descripción
infNom: detNoEl + inf	un infinitivo precedido por un determinante distinto de <i>el</i>
infNom: <i>el</i> + adj + inf	un infinitivo modificado por un adjetivo y precedido por <i>el</i>
infNom: <i>el</i> + inf + adj	
infNom: <i>el</i> + inf + <i>de</i>	un infinitivo seguido por <i>de</i> y precedido por <i>el</i>

4.3 Reglas para perífrasis y locuciones

Las reglas que reconocen las perífrasis asignan la etiqueta *uniVerb* (unidad verbal) al segmento que conforma la perífrasis completa. Dado el fragmento de texto de 4.4, que ya tiene algunas etiquetas asignadas por el sistema, la aplicación de las reglas para perífrasis de infinitivo agrega la etiqueta *uniVerb* que extiende las unidades verbales reconocidas anteriormente, como muestra el ejemplo 4.5.

Ejemplo 4.4 Algunas teorías [uniVerb/ podrán /uniVerb] [uniVerb/ elaborarse /uniVerb] una vez [uniVerb/ indagada /uniVerb] la antigüedad de las rocas, pero poco [uniVerb/ irá /uniVerb] a [uniVerb/ añadirse /uniVerb] con certeza.

Ejemplo 4.5 Algunas teorías **[uniVerb/ [uniVerb/ podrán /uniVerb] [uniVerb/ elaborarse /uniVerb] /uniVerb]** una vez [uniVerb indagada /uniVerb] la antigüedad de las rocas, pero poco **[uniVerb/ [uniVerb/ irá /uniVerb] a [uniVerb/ añadirse /uniVerb] /uniVerb]** con certeza.

En las etapas posteriores del procesamiento, la etiqueta *uniVerb* que señala la perífrasis (resaltada en negritas) oculta a la etiqueta *uniVerb* original.

Como se dijo anteriormente, la etiqueta *uniVerb*, además de indicar que el texto que encierra constituye una unidad verbal, aporta otro tipo de información sobre esa unidad.

En el caso de las perífrasis, la misma regla que la "arma" obtiene la información necesaria de los dos verbos que forman la expresión. Algunos rasgos morfológicos se toman del verbo auxiliar y otros del infinitivo. En la tabla 4.3 se describe la forma en que se obtiene la información para cada atributo de la perífrasis.

Tabla 4.3

Atributo	Descripción
lema	lema del infinitivo
forma	se toman del auxiliar
modo	
tiempo	
negación	
sujeto	
persona	
número	se toman del infinitivo
discurso	
interrogación	si alguno de los dos verbos tiene algún clítico, la perífrasis también lo tiene
clítico directo	
clítico indirecto	

Para la expresión *ella tiene que preguntarlo*, la regla que reconoce las perífrasis <tener+que+inf> asignará la etiqueta *uniVerb* con los valores que se muestran en la tabla 4.4 para sus atributos.

Tabla 4.4

Atributo	Valor
lema	preguntar
forma	fin
modo	indicativo
tiempo	presente
negación	no
sujeto	sí
persona	3
número	s
discurso	no
interrogación	sí
clítico directo	sí
clítico indirecto	no

El hecho de tomar como lema para la perífrasis solamente el lema del infinitivo permite reconocer perífrasis anidadas como *empezó a poder comer*. En primer lugar, a partir del lema *empezar* y la presencia de la preposición *a*, se reconoce como perífrasis *empezó a poder*, que queda marcada con lema *poder*. Luego, a partir del lema *poder*, se reconoce como perífrasis *empezó a poder comer*.

Para evitar hacer una regla para cada una de las perífrasis de la tabla 3.1, éstas se agrupan por similitud de estructura, dando lugar a 9 clases de perífrasis, basadas en 8 grupos de verbos. Cada clase de perífrasis es reconocida por una regla, por lo que se confeccionaron 9 reglas para el reconocimiento de todas las perífrasis de la tabla de 4.5.

Tabla 4.5

Clase de perífrasis	Conjunto de verbos pertenecientes al grupo
verboGrupo1 + inf	verboGrupo1={poder, deber, soler, acostumbrar}
verboGrupo2 + de + inf	verboGrupo2={deber, haber, acabar, terminar, dejar, parar, cesar}
verboGrupo3 + a + inf	verboGrupo3={ir, empezar, comenzar, ponerse, echar, echarse, romper, volver, venir, llegar, acertar, alcanzar}
verboGrupo4 + a punto de + inf	verboGrupo4={estar}
verboGrupo5 + por + inf	verboGrupo5={acabar, terminar}
verboGrupo6 + en + inf	verboGrupo6={tardar}
verboGrupo7 + que + inf	verboGrupo7={haber, tener}
no + verboGrupo7 + por qué + inf	

El sistema ClaTex, en su versión original, cuenta con reglas para las perífrasis basadas en *haber* y *tener*, ya que éstas contienen la palabra *que* que debe ser descartada como introductor de proposición en estos casos. Las reglas originales, que no etiquetan las expresiones como perífrasis sino que simplemente descartan el *que* como *iniProp* se sustituyen, en esta nueva versión, por reglas que arman las perífrasis correspondientes y se ejecutan junto con las restantes reglas para perífrasis de infinitivo.

Las perífrasis pueden contener después del auxiliar un adverbio o alguna expresión: *tiene mucho que contar*, *tiene casi siempre que contar*. Las reglas que se implementaron permiten la ocurrencia de hasta 4 palabras entre el auxiliar y el resto de la perífrasis, exigiendo que estas palabras no sean verbos, puntuación ni iniciador de proposición. De todos modos, casos más complejos como el que muestra el ejemplo 4.6 no están contemplados por las reglas.

Ejemplo 4.6 aun cuando la información **podiera eventualmente**, como dijo el dirigente Fernández, "**rozar** el secreto"

El problema quedaría resuelto si el inciso *como dijo el dirigente Fernández* fuera reconocido en una etapa previa del procesamiento. De este modo, las reglas, tal como están escritas, reconocerían la perífrasis *poder rozar*, ya que sólo habría dos elementos entre el auxiliar y el infinitivo: un adverbio y un inciso. Igualmente, queda otro problema por resolver: cómo marcar la perífrasis *poder rozar* sin dejar dentro de ella el inciso que constituye también una proposición. El sistema no permite manejar etiquetas para rotular segmentos discontinuos.

Otro problema que queda sin resolver, también relacionado con la imposibilidad de asignar etiquetas a segmentos discontinuos, es la coordinación de perífrasis, cuando se omite el auxiliar de la segunda. En el ejemplo 4.7, el sistema podrá reconocer *hay que preservar* como perífrasis, pero no *hay que educar*.

Ejemplo 4.7 Mantenerlos en secreto no ayuda en nada a la preservación y valoración pero **hay que preservar** el acervo y **educar** la comunidad.

Para las locuciones verbales de infinitivo, se escribieron reglas similares a las que reconocen perífrasis, es decir, se extiende la etiqueta *uniVerb* correspondiente al verbo conjugado de modo de que comprenda el resto de la expresión (el infinitivo y otras partículas).

Para las locuciones adverbiales con infinitivos, se escribieron reglas que las etiquetan como adverbio, ocultando así al infinitivo que contienen.

4.4 Reglas para proposiciones de infinitivo

Análogamente a las reglas que arman las proposiciones nucleadas por verbos finitos, las reglas para proposiciones de infinitivo asignan la etiqueta *prop* a segmentos que contienen un verbo, esta vez en infinitivo, que va a constituirse en su núcleo.

El sistema original opera en base a diferentes módulos de reglas que generan proposiciones nucleadas por verbos finitos:

- Un módulo que arma las proposiciones locales, que van a ser insertadas y son introducidas siempre por algún elemento explícito de subordinación (una *iniProp*).
- Otro módulo que arma las proposiciones libres, que son segmentos que contienen un verbo conjugado, delimitado por ciertos elementos especiales, pero que no están introducidos por una *iniProp*.
- Un último módulo que se aplica sobre todas las proposiciones reconocidas y resuelve el relacionamiento entre éstas, a nivel de la oración. Este módulo genera nuevas proposiciones mediante los siguientes mecanismos: se extienden proposiciones incorporando segmentos que no tienen forma de proposición, se subordinan proposiciones locales a proposiciones libres, se coordinan diferentes tipos de proposiciones.

Las reglas que se definieron para infinitivos son la contraparte de las reglas que arman las proposiciones locales para verbos finitos. Estas nuevas reglas se agruparon en un módulo que se ejecuta antes que los tres módulos ya existentes para proposiciones. De este modo, una vez que se ejecutaron los módulos que conforman las unidades verbales, incluyendo perífrasis y locuciones de infinitivo, desambiguan posibles introductores de proposición y arman series e incisos, el primer módulo para proposiciones que se ejecuta es el que se generó en este trabajo, que contiene las reglas para reconocimiento de proposiciones de infinitivo. Luego de este módulo se ejecutarán los módulos que arman las proposiciones finitas y las reglas que resuelven el tipo de relacionamiento que se da, dentro de la oración, entre todas las proposiciones marcadas.

En este trabajo, se partió de la hipótesis de que las proposiciones de infinitivo están siempre subordinadas a otras, por esta razón, cada proposición de infinitivo que se reconoce es marcada como "insertada", atributo que será utilizado por las reglas que definen el relacionamiento de ellas con las demás. Estas últimas reglas se aplican por igual a proposiciones finitas e infinitivas.

Para dar forma a las reglas para proposiciones de infinitivo, se trabajó en la identificación de los elementos que definen el límite izquierdo y el límite derecho de cada proposición.

El límite derecho se establece aplicando los mismo criterios que se definieron para las proposiciones finitas. Más que el final de una proposición, lo que se marca con estas reglas es el comienzo de otra proposición. Así, dependiendo de si la proposición que allí empieza es insertada o no, y según si existen segmentos no proposicionales después de esta última (que eventualmente podrán extender la primera), el punto identificado como límite derecho será efectivamente el final de la proposición identificada o no. A grandes rasgos, los elementos que determinan el principio de una nueva proposición, son: un indicador de inicio de proposición (*que, si, donde*), un verbo con capacidad para nuclear una proposición (sin contar al núcleo de la proposición en cuestión), un elemento coordinador (comas, conjunciones).

Para definir el límite izquierdo se establecieron algunos criterios específicos para las proposiciones de infinitivo. La gran mayoría comienza con el propio infinitivo, en muchos casos precedido por una preposición. Si bien esta preposición no siempre es regida por el infinitivo sino por el verbo que lo subordina, se resolvió extender la decisión tomada respecto al resto de las proposiciones del sistema: toda preposición que precede a una proposición se incluye dentro de la proposición.

La proposición de infinitivo también puede estar introducida por un elemento de subordinación, el cual debe ser incluido dentro de la proposición. Los casos de proposiciones

de infinitivo introducidos por subordinadores explícitos son pocos, pero su reconocimiento no presenta dificultades ya que no hay posibilidad de cometer errores al incorporar los en todos los casos a la proposición.

Otro elemento que ocurre antes del infinitivo es el adverbio. En algunos casos, deben incluirse dentro de la proposición, en otros casos no. Se resolvió incluir dentro de la proposición los adverbios focalizadores, como *incluso*, *solamente*, *especialmente*, que generalmente están modificando a la proposición que les sigue.

Por último, como se dijo anteriormente, es posible encontrar infinitivos verbales precedidos por el artículo definido *el*.

Un caso especial son las proposiciones de infinitivo introducidas por verbos de percepción sensible (*ver*, *oir*) o verbos causativos (*hacer*). En estas oraciones, puede haber un segmento antepuesto al infinitivo, que funciona como sujeto de éste. En el ejemplo 4.8, el segmento *a Joaquín* funciona como sujeto del infinitivo *quedar*, por lo cual debería incorporarse a la proposición infinitiva.

Ejemplo 4.8 la que hizo a Joaquín **quedar** mal con sus amigos

No se encontraron en el CorIn infinitivos introducidos por verbos de percepción sensible o causativos con sujeto explícito antepuesto. Queda pendiente el análisis de un corpus mayor para estudiar este fenómeno con más precisión. En el contexto de este trabajo no se escribieron reglas que reconozcan e incorporen estos segmentos a la proposición de infinitivo.

En la tabla 4.6, se reunieron los criterios que finalmente se tuvieron en cuenta para la definición del límite izquierdo de las proposiciones de infinitivo.

Tabla 4.6

Esquema reconocido por la regla	Descripción
propInf: [infinitivo + ...]	se genera una proposición de infinitivo que comienza con el infinitivo y termina según los criterios para proposiciones finitas
propInf: [preposición + propInf]	se incorpora una preposición a una proposición de infinitivo ya reconocida
propInf: [adverbio de foco + propInf]	se incorpora un adverbio de foco a una proposición de infinitivo ya reconocida
propInf: [el + propInf]	se incorpora el artículo <i>el</i> a una proposición de infinitivo ya reconocida
propInf: [iniProp + propInf]	se incorpora un elemento iniciador de proposición a una proposición de infinitivo

4.5 Evaluación

Se evaluó el impacto de la incorporación de proposiciones de infinitivo al sistema Clatex, en base a su aplicación sobre algunos de los textos que fueron utilizados en la evaluación final de la primera versión del sistema. Los textos se describen en la tabla 4.7.

Tabla 4.7

Texto	Publicación	Palabras
t1	Brecha 5/7/2002	2000
t9	El Observador 6/1/2000	625
t22	El País Digital 10/7/2002	550
t27	La República 10/7/2002	1040

En los textos, se encontraron 140 infinitivos, 138 fueron etiquetados como verbos y 2 como nombres. Se chequearon uno a uno los 140 infinitivos a fin de verificar la corrección del etiquetado.

Se comprobó que los dos casos etiquetados como nombres tienen, efectivamente, valor nominal: *titular* en 4.9 y *poder* en 4.10. Se trata de dos "falsos infinitivos", según la definición que se introdujo en la sección 2 de este documento. Para los dos casos, el diccionario de FreeLing contiene la etiqueta nombre, además de la etiqueta verbo.

Ejemplo 4.9 *sustituir* *sustituir* VMN0000
 al al SPCMS
 titular **titular** NCCS000
 del del SPCMS
 ministerio ministerio NCMS000
 de de SPS00
 Asuntos asunto NCMP000
 Exteriores exterior AQOCP0

Ejemplo 4.10 desde desde SPS00
 que que PROCN000
 Aznar aznar NCMS000
 llegó llegar VMIS3S0
 al al SPCMS
 poder **poder** NCMS000
 . . Fp

Para los infinitivos etiquetados como verbos por FreeLing, se comprobó también que el etiquetado es correcto. Se trata en general, de verbos que sólo tienen en el diccionario de FreeLing la opción verbo. Algunos pocos tienen las dos opciones y fueron bien desambiguados, como *poder* en 4.11.

Ejemplo 4.11 Para para SPS00
 poder **poder** VMN0000
 adquirir adquirir VMN0000
 o o CC
 portar portar VMN0000
 un uno DIOMS0
 arma arma NCFS000

El *tagger* introduce dos errores al encontrarse frente a dos ocurrencias de una palabra desconocida: el nombre propio *Comcar*. Este nombre propio es etiquetado como infinitivo con valor verbal, seguramente a partir de un análisis de su terminación. Estos dos errores llevan a una segmentación errónea, ya que las reglas que arman las proposiciones de infinitivos generan proposiciones con esta palabra como núcleo.

No hay ocurrencias en los textos procesados de infinitivos nominales, en el sentido que se le da a esta denominación en la sección 2 del documento. Se trata de infinitivos que ocurren con valor nominal en algunos contextos, sin tener la etiqueta nombre entre sus opciones. Al no haber ocurrencias de estos infinitivos en los textos, las reglas para su detección no fueron evaluadas.

Por otro lado, se evaluó el reconocimiento de locuciones y perífrasis. Las locuciones con infinitivos presentes en los textos son *a partir de*, *a pesar de* y *a pesar de que*. Las tres son reconocidas por el *tagger* (están incluidas en la tabla del anexo), por lo cual no se aplicaron reglas para reconocerlas.

Para las perífrasis se encontraron tres casos en los cuales aparecen perífrasis coordinadas que comparten el auxiliar, por lo que sólo el primer infinitivo es reconocido como perteneciente a una perífrasis (ejemplo 4.12). El resto de las perífrasis de los textos fueron marcadas correctamente.

Ejemplo 4.12 ... [hay que equilibrar] las finanzas, privatizar las empresas públicas, reducir el Estado en todo ...

Finalmente, en cuanto a la segmentación de las proposiciones de infinitivo y su relacionamiento con otras proposiciones de la oración en la cual participan, los resultados son similares a los que se obtienen para proposiciones finitas. En general, las proposiciones de infinitivo quedan bien delimitadas y su forma de insertarse en otras proposiciones se resuelve según los criterios definidos para las demás proposiciones.

Dos ejemplos de oraciones bien segmentadas se ven en 4.13 y 4.14. La oración de 4.13 contiene 2 proposiciones de infinitivo, correctamente delimitadas, coordinadas entre sí e insertadas en la subordinada con *que*. En el ejemplo 4.14, vemos una proposición de infinitivo anidada en otra.

Ejemplo 4.13

```
[ Aun desde ese punto de vista resulta intrincado el camino
  [ que hizo la cultura política uruguaya
    [
      [ para entender toda protesta como sinónimo de violencia ]
      Y
      [ condenarla ]
    ]
  ]
].
```

Ejemplo 4.14

```
[
  [ Mientras el fmi reclama , entre otras cosas ,
    [ reducir el Estado
      [ para seguir enviando sus remesas ]
    ]
  ],
  el gobierno pone cara
  [ de aplicarse ]
].
```

Sin embargo, se dan algunos errores en la inserción de algunas subordinadas, errores que se heredan de los criterios para resolver el relacionamiento entre proposiciones finitas. En 4.15, la proposición introducida por *porque* no debería ir insertada en la infinitiva sino en la principal.

Ejemplo 4.15

```
[ Entre la gente educada no es necesario
  [ salir
    [ a protestar
      [ porque un buen patrón ,
        [ como es el ministro ] ,
        cuidará de los puestos de trabajo de sus subordinados ]
    ]
  ]
].
```

5. Conclusiones

Con el objetivo de extender el sistema LaTeX, en este trabajo se incorporó el reconocimiento de proposiciones de infinitivo.

El diseño modular del sistema y su funcionamiento basado en la aplicación de reglas hicieron que el trabajo de implementación resultara bastante simple:

- Se incorporaron las reglas que detectan los infinitivos que no van a conformar núcleos de proposición al módulo de las reglas para unidades verbales.
- Se definió un nuevo módulo que contiene las reglas que arman las proposiciones de infinitivo.

- Se incluyó en la agenda del sistema la aplicación de las reglas nuevas.

El trabajo incluyó un estudio del comportamiento del *tagger* FreeLing en cuanto al etiquetado de los infinitivos para los cuales cuenta en su diccionario con las opciones *verbo* y *nombre*.

Los resultados de la evaluación son buenos, se mantienen los niveles de reconocimiento que se obtuvieron con el sistema original, que sólo marca proposiciones finitas.

Algunos aspectos a trabajar en etapas futuras son:

- la incorporación de nuevas locuciones con infinitivos,
- el estudio de casos de ambigüedad en secuencias de palabras que pueden funcionar como perífrasis o locuciones en algunos contextos y como palabras independientes en otros,
- la identificación del sujeto antepuesto del infinitivo para incorporarlo a la proposición,
- la posibilidad de marcar perífrasis discontinuas, con elementos entre el auxiliar y el infinitivo,
- la posibilidad de reconocer perífrasis coordinadas, que comparten el auxiliar.

Bibliografía

Bosque, I. y V. Demonte (dir) (1999), *Gramática Descriptiva de la Lengua Española*, Espasa Calpe, Madrid.

Caviglia, S., J. Couto, A. Rosá, D. Wonsever (2003), *Reconocimiento de indicadores de inicio de proposición en español* en *Actas del VIII Simposio de Lingüística Aplicada*, Santiago de Cuba, Cuba.

Caviglia, S., J. Couto, A. Rosá, D. Wonsever (2006), *Un sistema para la segmentación en proposiciones de textos en español* en *Letras de hoje*, número 144, Volumen 41, ISSN 0101-3335, EDIPUCRS, Porto Alegre, Brasil.

Gómez Torrego, L. (1999), *Los verbos auxiliares. Las perífrasis verbales de infinitivo* en Bosque, I. y V. Demonte 1999, cap. 51, tomo 2.

Grassi, M., M. Malcuori, J. Couto, J.J. Prada, D. Wonsever (2001), *Corpus informatizado: textos del español del Uruguay (CORIN)*, en *SLPLT-2 - Second International Workshop on Spanish Language Processing and Language Technologies*, Jaén, España.

Hernanz, M. L. (1999), *El infinitivo* en Bosque, I. y V. Demonte 1999, cap. 36, tomo 2.

Wonsever, D., J.-L. Minel (2004), *Contextual Rules for Text Analysis* en *Lecture Notes in Computer Science*.

Referencias en la web

[CDE] Corpus del español: <http://www.corpusdelespanol.org> (último acceso en junio de 2007).

[DRAE] Diccionario de la Real Academia Española: <http://buscon.rae.es/draeI/> (último acceso en junio de 2007).

[FL] FreeLing: <http://garraf.epsevg.upc.es/freeling/> (último acceso en junio de 2007).

ANEXO: Locuciones adverbiales, conjuntivas y prepositivas con infinitivos, reconocidas por FreeLing

Significado de las etiquetas del *tagger*:

RG adverbio
 SP... preposición
 CS conjunción subordinante

a_juzgar_como	RG	como_quien_oye_llover	RG
a_juzgar_por	SPS00	dar_carpetazo	RG
a_mal_dar	RG	de_andar_por_casa	RG
a_medio_hacer	RG	de_armas_tomar	RG
a_más_andar	RG	de_mal_vivir	RG
a_más_correr	RG	de_nunca_acabar	RG
a_más_ganar	RG	de_poder_a_poder	RG
a_más_no_poder	RG	en_un_abrir_y_cerrar_de_ojos	RG
a_más_tardar	RG	en_un_decir_amén	RG
a_más_ver	RG	es_hablar_por_demás	RG
a_no_ser_que	CS	hacer_cimarra	RG
a_partir_de	SPS00	hacer_la_cusca	RG
a_partir_del	SPCMS	hasta_más_ver	RG
a_pesar_de	SPS00	ni_hablar_del_peluquín	RG
a_pesar_de_que	CS	ni_que_decir_tiene	RG
a_pesar_de_todo	RG	no_decir_chus_ni_mus	RG
a_pesar_del	SPCMS	no_ser_cosa_de	SPS00
a_seguir_bien	RG	para_dar_y_tomar	RG
a_todo_andar	RG	según_el_decir_de	SPS00
a_todo_correr	RG	sin_chistar_ni_mistar	RG
a_todo_meter	RG	sin_contar_con	SPS00
a_todo_moler	RG	sin_contar_con_que	CS
a_todo_tirar	RG	sin_ir_más_lejos	RG
al_clarear_el_día	RG	sin_venir_a_pelo	RG
al_correr_de_la_pluma	RG	tirar_de_cupitel	RG
al_correr_del_tiempo	RG	ya_es_mucho_decir	RG
al_decir_de	SPS00		