

Análisis lingüístico automático en distintos niveles

Rodolfo Bonino

Facultad de Humanidades y Artes
Universidad Nacional de Rosario
rodolfobonino@yahoo.com.ar

Resumen

En este trabajo se propone implantar en el programa XFST un modelo de análisis, basado en *5P*, a partir de rasgos distintivos de las categorías léxicas. El tratamiento de la sintaxis es la última instancia de un sistema que se inicia en morfemas lexemáticos y morfemas flexivos, etiquetados de un modo que hace posible calcular automáticamente el comportamiento sintáctico de las entradas léxicas. Las propiedades de *5P* permiten integrar a los morfemas en categorías léxicas y a estas en construcciones sintácticas mayores.

Palabras claves: Lingüística Computacional – Morfología – Sintaxis – análisis automático – generación automática – *5P* – XFST

Abstract

This work aims to introduce a *5P*-based analysis model into the XSFT program as from lexical categories' distinctive features. The way syntax is handled is the last step of a system which begins with an analysis of lexemes and inflectional morphemes. These are labeled in such a way that it is possible to autocalculate the syntactic behavior of lexical entries. *5P* properties enable morphemes to be integrated into lexical categories, and these categories, then, to be integrated into even larger syntactic structures.

Key terms: Computational linguistics – Morphology- Syntax – Automatic analysis – Automatic generation 5 – XFST

1. INTRODUCCIÓN

En [3] se sitúa el modelo *5P* [2] en el contexto de la Lingüística Computacional y se explican sus postulados generales y la compatibilidad de este formalismo lingüístico con el formalismo computacional *Xerox Finite-State Tool* (XFST) [3].

En [2], Bès se formula el modelo *5P* a partir de una aplicación al estudio de la frase verbal núcleo del francés, que se define como una cadena de palabras que puede comenzar a la izquierda con la partícula de negación *ne*, por un pronombre clítico no nominativo, por un auxiliar o por un verbo y que termina a la derecha en un verbo, que es el núcleo de la frase, o por alguna de las formas que también se pueden encontrar a la derecha. El objeto de análisis incluye un solo elemento nuclear

con el que tienen valencia todos los demás y, consecuentemente, el análisis tiene un solo nivel; pero el formalismo propuesto tiene aplicaciones en objetos mucho más complejos.

En el presente trabajo, se formulan propiedades de *5P* tendientes a construir un objeto donde se dan relaciones en distintos niveles. El tratamiento de la sintaxis es la última instancia de un sistema que se inicia en morfemas lexemáticos y morfemas flexivos, etiquetados de un modo que hace posible calcular automáticamente el comportamiento sintáctico de las entradas léxicas. Las propiedades de *5P* permiten integrar a los morfemas en categorías léxicas, y a estas, en construcciones sintácticas mayores.

La implantación de las propiedades en XFST permite que el programa genere automáticamente secuencias gramaticales y produzca el análisis de los distintos niveles. La salida del análisis identifica los rasgos de los morfemas y las funciones sintácticas de las palabras y los sintagmas.

Dado que el interés está centrado en la formulación de propiedades y los procedimientos de implantación, el objeto lingüístico se limita a una morfología y una sintaxis rudimentarias: un conjunto de morfemas caracterizados por rasgos elementales que se integran en sustantivos y verbos que, en una segunda instancia, se integran en sintagmas que contienen a lo sumo artículo + sustantivo o verbo + objeto directo (sustantivo o artículos + sustantivo) y estos, a su vez, se integran en oraciones de sujeto – predicado.

2. MORFEMAS Y FUNCIONES SINTÁCTICAS

La gramática tradicional aborda el estudio de las categorías léxicas desde una perspectiva eminentemente semántica. A partir de [4], con el principio de proyección, se formaliza y perfecciona el análisis de las propiedades semánticas del léxico. Actualmente, distintas corrientes lingüísticas parten del presupuesto de que estas propiedades, conjuntamente con sus propiedades flexivas, son la base para la formalización de la sintaxis.

Dado que el objetivo del presente trabajo no es realizar un estudio semántico y morfológico sino comprobar la viabilidad del procedimiento utilizado; se toman las clasificaciones semánticas más básicas de verbos y sustantivos limitadas a variaciones morfológicas elementales. Sin perder de vista que estas clasificaciones serían insuficientes para un análisis más amplio, lo que se intenta mostrar es que mediante el recurso de determinar las propiedades de los morfemas es posible generar construcciones sintácticas.

En el marco de *5P*, las categorías se definen como un conjunto de rasgos. Los rasgos son etiquetas utilizadas para identificar relativa y opositivamente las categorías y se definen por el conjunto de valores del rasgo. Las categorías máximas son aquellas a las cuales no se les puede agregar ningún otro rasgo. Las categorías no máximas pueden ser cualesquiera de los rasgos que caracterizan a las categorías máximas, siempre que se hayan definido previamente. Si se establece una propiedad para una categoría no máxima, esa propiedad será válida para todas las categorías máximas que aquella subsume.

En lo que respecta al tema de este trabajo, se utilizan categorías máximas para etiquetar los morfemas y las funciones sintácticas que pueden cumplir las palabras generadas a partir de estos morfemas en el sintagma. Siempre que no tengan valor cero (como es, por ejemplo, el caso del morfema flexivo nominal singular), las etiquetas que caracterizan a los morfemas se asocian a cadenas del español, que se utilizan a modo de ejemplo.

Los morfemas lexemáticos se definen como la concatenación de un conjunto de rasgos que identifica sus propiedades semánticas con uno que identifica sus propiedades morfológicas:

Morfemas lexemáticos verbales:

Propiedades semánticas:

Valores de rasgo: transitivo (*cant-*), intransitivo (*trabaj-*) y unipersonal (*h-*)

Propiedades morfológicas:

Valores de rasgo: raíz de conjugación regular (*cant-*, *trabaj-*), raíz de conjugación irregular (*h-*).

La selección de las cadenas determina que solo sean posibles las concatenaciones –que se especifican mediante propiedades, como se explicará más adelante– de los valores: transitivo-raíz de conjugación regular, intransitivo-raíz de conjugación regular e impersonal-raíz de conjugación irregular; si se ampliara el número de entradas, se podrían dar otras combinaciones como, por ejemplo, transitivo-raíz de conjugación irregular (*dar*).

Morfemas lexemáticos nominales:

Propiedades semánticas:

Valores de rasgo: nombre animado (*hombre, mujer, muchach-*), nombre inanimado (*tango, zamba*)

Propiedades morfológicas:

Valores de rasgo: raíz de flexión 1: moción de género - plural en *-s* (*muchach-*), raíz de flexión 2: masculino sin moción de género - plural en *-s* (*hombre, tango*), raíz de flexión 3: femenino sin moción de género - plural *-es* (*mujer*), raíz de flexión 4: femenino sin moción de género - plural *-s* (*zamba*).

Los morfemas flexivos verbales se definen como la concatenación de dos conjuntos de rasgos: los que permiten calcular sus propiedades de concordancia y los que permiten calcular el morfema lexemático con el que se asocian (propiedades de integración):

Morfemas flexivos verbales:

Propiedades de concordancia:

Valores de rasgo: singular, plural (se utiliza únicamente la tercera persona)

Propiedades de integración:

Valores de rasgo: desinencia de conjugación regular (*-a -an*) y desinencia de conjugación irregular (*-ay*).

Los morfemas flexivos verbales se definen como la concatenación de un conjunto de rasgos de género con uno de número:

Morfemas flexivos nominales:

Género: masculino, femenino.

Número: singular, plural.

Género y número son categorías no máximas definidas por un rasgo referido a la concordancia y a las propiedades de integración:

Masculino:

Valores de rasgo: masculino 1 (*-o*), masculino 2 (*0*)

Femenino:

Valores de rasgo: femenino 1 (-a), femenino 2 (0)

Singular:

Valor de rasgo: singular (0)

Plural:

Valor de rasgo: plural 1 (-s), plural 2 (-es)

Los artículos se definen directamente como categorías máximas, es decir, a cada uno de ellos se le asigna un valor de rasgo de cada uno de los conjuntos utilizados para definirlos (nombre de la categoría, rasgo definido / indefinido, rasgo de género y rasgo de número). A partir de estos, se definen categorías no máximas necesarias para la formulación de propiedades.

De igual modo se procede con las funciones sintácticas: *sujeto*, *predicado*, *determinante del sujeto*, *determinante del objeto*, *núcleo nominal del sujeto*, *núcleo nominal del objeto*, *núcleo verbal*, *objeto directo*. El procedimiento de generación va desde un nivel inferior (morfemas) a niveles superiores (sintagmas y oración), para que el análisis represente de algún modo estos niveles se emplea el signo \: los rasgos que identifican a los morfemas constituyen el punto de partida de la generación y no llevan ninguna marca; las funciones *determinante del sujeto*, *determinante del objeto*, *núcleo nominal del sujeto*, *núcleo nominal del objeto*, y *núcleo verbal* son funciones que corresponden a palabras, entidades que están en el primer nivel de la generación, esto se indica con \. La función *objeto directo* corresponde a sintagmas conformados por *determinante* y *núcleo nominal*, esto se indica con \|. Las funciones *sujeto* y *predicado* corresponden a un tercer nivel que se marca con \|\|.

Una vez definidos los morfemas, se declaran las propiedades que determinan su combinatoria:

El verbo se define como la unión del conjunto de morfemas lexemáticos verbales y el de morfemas flexivos verbales a los que se adiciona la intersección de las siguientes propiedades:

Propiedad de unicidad: hay al menos un morfema lexemático verbal.

Propiedad de exigencia: el morfema lexemático verbal exige morfema flexivo verbal.

Propiedades de exclusión:

- 1) Las raíces de conjugación regular excluyen desinencias de conjugación irregular.
- 2) Las raíces de conjugación irregular excluyen desinencias de conjugación regular.

Propiedades de precedencia: el morfema lexemático verbal precede al morfema flexivo verbal.

El nombre se define como la unión del conjunto de morfemas lexemáticos nominales y el de morfemas flexivos nominales a los que se adiciona la intersección de las siguientes propiedades:

Propiedades de unicidad:

- 1) Hay al menos un morfema lexemático nominal.
- 2) No hay más que un morfema de género
- 3) No hay más que un morfema de número

Propiedades de exigencia:

- 1) El morfema lexemático nominal exige un morfema de género concatenado con un morfema de número. (*mujer*: está formado por el morfema lexemático *mujer*, el morfema flexivo de género 0 y el morfema flexivo de número 0)
- 2) Las raíces de flexión 1 exigen género 1. (*muchach-* exige *-a, -o*)
- 3) El género 1 exige raíces de flexión 1. (*-a, -o* exige *muchach-*)
- 4) El plural 2 exige raíces de flexión 3 (*-es* exige *mujer*).

Dado que las propiedades de exigencia no son recíprocas, es necesario formular 2) y 3)

Propiedades de exclusión:

- 1) Las raíces de flexión 2 excluyen femenino. (*hombre – tango* excluye morfema 0 con valor femenino)
- 2) Las raíces de flexión 3 y 4 excluyen masculino. (*mujer – zamba* excluye morfema 0 con valor masculino)
- 3) Las raíces de flexión 3 excluyen plural 1. (*mujer* excluye plural *-s*)

La exigencia 4) y la exclusión 3) tornan innecesaria la formulación de propiedades que establezcan que el morfema plural *-s* exige las raíces 1, 2 o 4 y que estas raíces excluyen el morfema plural *-es*.

Propiedades de precedencia: el morfema lexemático verbal precede al morfema flexivo verbal.

En el cuadro 1 se muestran las salidas generadas automáticamente a partir de esta implantación con el comando *print lower-words* de XFST, este comando genera cadenas lingüísticas sin analizar:

hay
trabajan
trabaja
cantan
canta
muchacha
muchachas
muchacho
muchachos
hombre
hombres
mujer
mujeres
tango
tangos
zamba
zambas
unas
unos
una
un
las
los
la
el

Cuadro1: Salidas producidas por el comando *print lower-words*

En el cuadro 2 se muestran parcialmente las salidas producidas automáticamente por el comando *print words*. Este comando genera cadenas lingüísticas con el análisis correspondiente:

```

< : h><Vbunip/:0><rconjirreg/:0><dconjirreg/:ay><vsg/:0>
< : trabaj><Vbintr/:0><rconjreg/:0><dconjreg/:an><vpl/:0>
< : trabaj><Vbintr/:0><rconjreg/:0><dconjreg/:a><vsg/:0>
< : cant><Vbtr/:0><rconjreg/:0><dconjreg/:an><vpl/:0>
< : cant><Vbtr/:0><rconjreg/:0><dconjreg/:a><vsg/:0>
< : muchach><Nmanim/:0><rfl1/:0><fem1/:a><sg/:0>
< : muchach><Nmanim/:0><rfl1/:0><fem1/:a><pl1/:s>
< : muchach><Nmanim/:0><rfl1/:0><mascl/:o><sg/:0>
< : muchach><Nmanim/:0><rfl1/:0><mascl/:o><pl1/:s>
< : hombre><Nmanim/:0><rfl2/:0><mascl2/:0><sg/:0>
< : hombre><Nmanim/:0><rfl2/:0><mascl2/:0><pl1/:s>
< : mujer><Nmanim/:0><rfl3/:0><fem2/:0><sg/:0>
< : mujer><Nmanim/:0><rfl3/:0><fem2/:0><pl2/:es>
< : tango><Nmin/:0><rfl2/:0><mascl2/:0><sg/:0>
< : tango><Nmin/:0><rfl2/:0><mascl2/:0><pl1/:s>
< : zamba><Nmin/:0><rfl4/:0><fem2/:0><sg/:0>
< : zamba><Nmin/:0><rfl4/:0><fem2/:0><pl1/:s>

```

Cuadro 2. Selección de salidas producidas por el comando print words.

Como se puede observar en la primera línea, la representación del análisis es: *h* verbo unipersonal, raíz de conjugación irregular; desinencia de conjugación irregular: *ay*, verbo singular.

Las salidas de este análisis morfológico se pueden utilizar para formalizar la sintaxis, para ello será necesario declarar nuevas propiedades que definan estos objetos lingüísticos, como se verá en los siguientes apartados.

3. SINTAGMAS NOMINALES Y SINTAGMAS VERBALES

En esta etapa se trabaja sobre sintagmas nominales constituidos por determinante y núcleo y con sintagmas verbales constituidos por verbo y objeto directo (sintagma nominal). Por esta razón, es necesario definir en primer término al SN.

Todas las propiedades de los sintagmas toman como fuente el análisis morfológico, y le adicionan nuevas propiedades.

Los elementos que integran los SSNN son el artículo concatenado con la función de determinante del objeto o determinante del sujeto. Este conjunto de categorías y funciones, a su vez se concatena con el conjunto formado por nombre concatenado con las funciones de núcleo nominal del sujeto o del objeto. Las relaciones entre estos elementos se especifican con la intersección las siguientes propiedades:

Propiedades de unicidad:

- 1) En SN hay al menos un N.
- 2) En SN no hay más que un artículo.

En este nivel, el sustantivo sin artículo recibe un análisis morfológico como sustantivo y un análisis sintáctico como sintagma nominal.

Propiedades de exclusión:

- 1) El morfema flexivo nominal masculino singular excluye artículo femenino singular, masculino plural y femenino plural.
- 2) El morfema flexivo nominal femenino singular excluye artículo masculino singular, masculino plural y femenino plural.
- 3) El morfema flexivo nominal masculino plural excluye el artículo masculino singular, femenino singular y femenino plural.
- 4) El morfema flexivo nominal femenino plural excluye el artículo masculino singular, femenino singular y masculino plural.

Las propiedades de exclusión son simplemente propiedades que regulan la concordancia.

Propiedades de exigencia:

- 1) El determinante del objeto exige núcleo nominal del objeto.
- 2) El determinante del sujeto exige núcleo nominal del sujeto.

Propiedad de precedencia: el artículo precede al nombre.

Los elementos que integran el sintagma verbal son el verbo y el sintagma nominal en función de objeto directo.

Propiedades de unicidad:

- 1) En SV hay al menos un V.
- 2) En SV no hay más que un objeto directo.

Propiedades de exclusión:

- 1) El verbo intransitivo excluye al objeto directo.
- 2) El verbo unipersonal excluye al artículo determinado (**hay la zamba*)
- 3) El verbo transitivo excluye al nombre animado en función de núcleo nominal del objeto directo. Esta propiedad es válida en relación con el recorte del objeto.

Propiedades de exigencia:

- 1) El objeto directo exige núcleo nominal del objeto.
- 2) El verbo unipersonal exige artículo indeterminado concatenado con nombre animado en singular o nombre animado en plural o nombre inanimado. Con esta propiedad se evitan construcciones dudosas como *??hay muchacha* y se generan construcciones indudablemente gramaticales como *hay tango, hay una muchacha*.

Propiedad de precedencia: el verbo precede al objeto directo.

En el cuadro 3, se muestra una selección de las 116 salidas generadas por el comando *print-words* y en el cuadro 4, una selección de las salidas analizadas que genera el comando *print*. Como puede observarse, esta implantación produce un nivel de análisis morfológico y un nivel de análisis sintáctico. Los verbos y sustantivos aparecen aislados porque coinciden con sintagmas verbales y nominales, respectivamente. Dado que los artículos aislados no forman sintagmas, no se generan ni analizan en este nivel.

hay una zamba

```

trabaja
cantan
cantan tango
canta una zamba
canta el tango
una muchacha
una mujer

```

Cuadro 3: selección de salidas del comando print-words

```

< : h><Vbunip/:0><rconjirreg/:0><dconjirreg/:ay><vsg/:0><\NV:0>< ARTINDfemsg:
una><\DETO:0>< : zamba><Nmin/:0><rfl4/:0><fem2/:0><sg/:0><\NNO:0><\OD:0>
< : trabaj><Vbintr/:0><rconjreg/:0><dconjreg/:an><vpl/:0><\NV:0>
< : cant><Vbtr/:0><rconjreg/:0><dconjreg/:an><vpl/:0><\NV:0>
< : cant><Vbtr/:0><rconjreg/:0><dconjreg/:an><vpl/:0><\NV:0>< :
tango><Nmin/:0><rfl2/:0><masc2/:0><sg/:0><\NNO:0><\OD:0>
< : cant><Vbtr/:0><rconjreg/:0><dconjreg/:a><vsg/:0><\NV:0>< ARTINDfemsg: una><\DETO:0><
: zamba><Nmin/:0><rfl4/:0><fem2/:0><sg/:0><\NNO:0><\OD:0>
< : cant><Vbtr/:0><rconjreg/:0><dconjreg/:a><vsg/:0><\NV:0>< ARTDETmascsg: el><\DETO:0><
: tango><Nmin/:0><rfl2/:0><masc2/:0><sg/:0><\NNO:0><\OD:0>
< : cant><Vbtr/:0><rconjreg/:0><dconjreg/:a><vsg/:0><\NV:0>< ARTINDmascsg: un><\DETO:0><
: tango><Nmin/:0><rfl2/:0><masc2/:0><sg/:0><\NNO:0><\OD:0>
< : muchach><Nmanim/:0><rfl1/:0><fem1/:a><sg/:0><\NNO:0>
< ARTINDfemsg: una><\DETS:0>< : muchach><Nmanim/:0><rfl1/:0><fem1/:a><sg/:0><\NNS:0>
< ARTINDfemsg: una><\DETS:0>< : mujer><Nmanim/:0><rfl3/:0><fem2/:0><sg/:0><\NNS:0>

```

Cuadro 4. Selección de salidas del comando print de XFST (ejemplos en negrita agregados manualmente)

4. ORACIÓN

En esta etapa se trabaja para obtener la generación y el análisis automáticos de oraciones intransitivas con sujeto animado y oraciones transitivas con sujeto animado y objeto inanimado. Las propiedades de la oración se formulan a partir de las salidas de las dos implantaciones anteriores.

Los elementos que la integran son un SN en función de sujeto y un SV en función de predicado. El primer constituyente puede estar constituido por un artículo en función de determinante y un sustantivo núcleo, y el segundo por un verbo núcleo y un SN objeto directo; pero no es necesario definir estas relaciones porque son las salidas de la etapa anterior. En esta etapa se definen las siguientes propiedades:

Propiedades de unicidad:

- 1) En O hay al menos un SV.
- 2) En O no hay más que un sujeto.

Propiedades de exclusión:

- 1) El sujeto excluye nombre inanimado (**la zamba canta*).
- 2) El verbo en singular excluye al núcleo nominal del sujeto en plural.
- 3) El verbo en plural excluye al núcleo nominal del sujeto en singular.
- 4) El verbo unipersonal excluye al sujeto.

Las propiedades 2) y 3) garantizan la concordancia.

Propiedades de exigencia:

- 1) El sujeto exige núcleo nominal del sujeto.

2) El sujeto exige determinante del sujeto.

Propiedad de precedencia: sujeto precede al predicado.

En el cuadro 5 se muestra una selección de las 268 salidas producidas por los comandos *print-words* y *print*, respectivamente.

```

hay una zamba
hay un tango
hay muchachas
trabajan
cantan tango
cantan
un muchacho trabaja
un muchacho canta tango
el muchacho trabaja
el muchacho canta tango
el muchacho canta el tango
el muchacho canta un tango
el hombre trabaja
unos muchachos cantan tango
unos muchachos cantan tangos
unos hombres trabajan
los muchachos trabajan
los muchachos cantan las zambas
los hombres cantan la zamba
los hombres cantan una zamba
los hombres cantan el tango
los hombres cantan un tango
unas muchachas trabajan
    
```

Cuadro 5. Selección de salidas del comando print-words

```

< : h><Vbunip/:0><rconjirreg/:0><dconjirreg/:ay><vsg/:0><\NV:0>< ARTINDfemsg:
una><\DETO:0>< : zamba><Nmin/:0><rfl4/:0><fem2/:0><sg/:0><\NNO:0><\OD:0><\PRED:0>
< : h><Vbunip/:0><rconjirreg/:0><dconjirreg/:ay><vsg/:0><\NV:0>< ARTINDmascsg:
un><\DETO:0>< : tango><Nmin/:0><rfl2/:0><masc2/:0><sg/:0><\NNO:0><\OD:0><\PRED:0>
< : h><Vbunip/:0><rconjirreg/:0><dconjirreg/:ay><vsg/:0><\NV:0>< :
muchach><Nmanim/:0><rfl1/:0><fem1/:a><pl1/:s><\NNO:0><\OD:0><\PRED:0>
< : trabaj><Vbintr/:0><rconjreg/:0><dconjreg/:an><vpl/:0><\NV:0><\PRED:0>
< ARTDETFempl: las><\DETS:0>< :
mujer><Nmanim/:0><rfl3/:0><fem2/:0><pl2/:es><\NNS:0><\SUI:0>< :
cant><Vbtr/:0><rconjreg/:0><dconjreg/:an><vpl/:0><\NV:0>< ARTDETFemsg: la><\DETO:0>< :
zamba><Nmin/:0><rfl4/:0><fem2/:0><sg/:0><\NNO:0><\OD:0><\PRED:0>
< ARTDETFempl: las><\DETS:0>< :
mujer><Nmanim/:0><rfl3/:0><fem2/:0><pl2/:es><\NNS:0><\SUI:0>< :
cant><Vbtr/:0><rconjreg/:0><dconjreg/:an><vpl/:0><\NV:0>< ARTINDfemsg: una><\DETO:0>< :
zamba><Nmin/:0><rfl4/:0><fem2/:0><sg/:0><\NNO:0><\OD:0><\PRED:0>
< ARTDETFempl: las><\DETS:0>< :
mujer><Nmanim/:0><rfl3/:0><fem2/:0><pl2/:es><\NNS:0><\SUI:0>< :
cant><Vbtr/:0><rconjreg/:0><dconjreg/:an><vpl/:0><\NV:0>< ARTDETMascsg: el><\DETO:0>< :
tango><Nmin/:0><rfl2/:0><masc2/:0><sg/:0><\NNO:0><\OD:0><\PRED:0>
< ARTDETFempl: las><\DETS:0>< :
mujer><Nmanim/:0><rfl3/:0><fem2/:0><pl2/:es><\NNS:0><\SUI:0>< :
cant><Vbtr/:0><rconjreg/:0><dconjreg/:an><vpl/:0><\NV:0>< ARTINDmascsg: un><\DETO:0>< :
tango><Nmin/:0><rfl2/:0><masc2/:0><sg/:0><\NNO:0><\OD:0><\PRED:0>
    
```

Cuadro 6. Selección de salidas del comando print.

5. CONCLUSIONES

Las propiedades de *5P* que pueden ser implantadas en el programa XFST permiten generar y analizar cadenas lingüísticas en distintos niveles. En este trabajo se toman como punto de partida las propiedades de los morfemas lexemáticos y flexivos y, a partir de ellos, se formulan en una primera etapa propiedades de sintagmas nominales y sintagmas verbales; en una segunda etapa, que involucra a las dos anteriores, oraciones de sujeto - predicado.

Si se pretendiera elaborar un proyecto de mayor alcance, se debería incrementar el número de rasgos tanto de los morfemas flexivos como de los lexemáticos. La sistematización de la morfología flexiva en la lingüística no computacional hace que el análisis automático de la morfología resulte menos problemático para la investigación en lingüística computacional; pero el estudio del léxico es un tema sobre el cual la formalización que requiere el estudio computacional podría aportar una perspectiva de gran utilidad para el conocimiento de las lenguas naturales. Si se analiza el pequeño campo al que se circunscribe este trabajo, se puede observar, por ejemplo, que la transitividad es una condición necesaria, pero no suficiente, para que *tango* y *zamba*, con todas sus posibilidades morfológicas y de combinatoria con los artículos, cumplan la función de objeto directo; otro verbo transitivo como *comer* excluiría en la mayoría de los contextos a estas palabras como objeto directo. Para solucionar estas cuestiones, la lingüística, probablemente, debería recurrir a la estadística, que aportaría datos relevantes acerca de las ocurrencias en el habla real.

Otra cuestión insoslayable es la de la recursividad del lenguaje; en la reducida escala en la que se trabajó, se distinguieron solo tres niveles de la sintaxis: el de la oración, el del sintagma nominal y el del sintagma verbal, que incluye al sintagma nominal. La reducción del objeto permitió representar los niveles mediante el sistema de barras; pero si se pretende extender el análisis a estructuras más complejas, este procedimiento resultaría inaplicable. Este tema ha preocupado a los investigadores desde el inicio, pero hasta el presente no se ha podido resolver de manera satisfactoria.

En lo que respecta a la herramienta informática, el análisis generado automáticamente adolece de una complejidad y una falta de atractivo visual que, aunque no invalida su adecuación lingüística, lo hace de difícil comprensión para quien no está habituado al uso del programa.

En síntesis, el camino recorrido por este trabajo y los que lo preceden, a la vez que muestran que la lingüística computacional puede aportar conocimientos teóricos y metodológicos relevantes, plantean una serie de cuestiones que invitan a desarrollar la investigación tanto en el campo de la lingüística como en el de la informática.

Referencias

- [1] R. Bonino. "Presentación de la Lingüística Computacional" en Z. Solana (ed). La interlengua de los aprendientes del español como L2. Juglaría, 2009, págs. 7 – 18.
- [2] G. Bès. "La phrase verbale noyau en français" en Recherches sur le français parlé, N° 15, Provence, Publications de l'Université de Provence, 1999, págs. 273 – 358.
- [3] L. Karttunen, T. Gaál, and A. Kempe. Xerox Finite-State Tool. The Document Company - Xerox, 1997.
- [4] N. Chomsky. Lectures on Government and Binding. Mit Press, Cambridge (Massachusetts), 1981