



ARCHIVO DEL LABORATORIO DE DOCUMENTACIÓN GEOMÉTRICA DEL PATRIMONIO

ARCHIVE OF THE LABORATORY FOR THE GEOMETRIC
DOCUMENTATION OF HERITAGE

Sección de informes de investigación / **Research report section**

1

Información general / General information		
AUTOR:	José Manuel VALLE MELÓN Álvaro RODRIGUEZ MIRANDA	:AUTHOR
TITULO:	Representación 3D interactiva de relaciones de derivación e hiponimia en los paradigmas léxicos del inglés antiguo	:TITLE
FECHA:	diciembre 2011 / December 2011	:DATE
NUMERO:	LDGP_inv_001	:NUMBER
IDIOMA:	español / Spanish	:LANGUAGE

Resumen	
TITULO:	Representación 3D interactiva de relaciones de derivación e hiponimia en paradigmas léxicos del inglés antiguo
TITULO PROYECTO:	Nerthusv2: Base de datos léxica en 3D del inglés antiguo FFI08-04448/FILO
ENTIDAD FINANCIADORA:	Ministerio de Ciencia e Innovación (España)
INVESTIGADOR PRINCIPAL:	Javier MARTIN ARISTA (Universidad de La Rioja)
INVESTIGADORES:	13
PARTICIPANTES:	<ul style="list-style-type: none"> - [PRINCIPAL] Departamento de Filologías Modernas (Universidad de La Rioja). - Laboratorio de Documentación Geométrica del Patrimonio (Universidad del País Vasco UPV/EHU).
RESUMEN DEL PROYECTO:	Los objetivos del proyecto Nerthus son proporcionar un análisis general del léxico del inglés antiguo basado en las teorías actuales de la lingüística e incorporar los avances del análisis lexicológico en un producto lexicográfico que alcance los estándares del siglo XXI. El proyecto Nerthus ha optado por la base de datos léxica por sus múltiples ventajas respecto a un diccionario.
RESUMEN DE LA CONTRIBUCIÓN:	Representación tridimensional de la base de datos como estructuras topológicas manejables de forma interactiva por los usuarios.
DESCRIPTORES NATURALES:	lingüística, inglés antiguo, topología, visualización
DESCRIPTORES CONTROLADOS:	(Procedentes del Tesoro UNESCO [http://databases.unesco.org/thessp/]) Lingüística, Inglés, Topología, Visualización

Abstract	
TITLE:	Interactive 3D representation of relations of derivation and hyponymy holding in the lexical paradigms of Old English
TITLE:	Nerthusv2: Lexical database in 3D of Old English FFI08-04448/FILO
FUNDING AGENCY:	Spanish Ministry of Science and Innovation
MAIN RESEARCHER:	Javier MARTIN ARISTA (Universidad de La Rioja)
RESEARCHERS:	13
PARTNERS:	<ul style="list-style-type: none"> - [MAIN] Department of Modern Languages (Universidad de La Rioja) - Laboratory for the Geometric Documentation of the Heritage (University of the Basque Country UPV/EHU)
ABSTRACT OF THE PROJECT:	The aims of the Nerthus project are to provide an overall analysis of the Old English lexicon based on up-to-date linguistic theory and to incorporate the findings of lexicological analysis into a lexicographical product that meets 21 st century lexicographical standards. The Nerthus project has opted for the database format because it has several advantages over a dictionary.
ABSTRACT OF THE CONTRIBUTION:	3D representation of the database by means of topological structures that can be accessed interactively.
NATURAL KEYWORDS:	linguistics, Old English, topology, visualization
CONTROLLED KEYWORDS:	(From the UNESCO's thesaurus [http://databases.unesco.org/thesaurus/]) Linguistics, English, Topology, Visualization

Derechos / Rights		
DERECHOS:	<p>Está permitido citar y extraer el texto, siempre que la fuente sea claramente identificada (respecto a la consideración de "no comercial" ver el apartado "otros derechos"). / Permission is granted to quote and take excerpts from this text, provided that the source of such material is fully acknowledged (for the "non commercial" label see below in "others rights").</p> <div style="text-align: center;">  </div>	:RIGHTS
OTROS:		:OTHERS

Renuncia de responsabilidad / Disclaimer		
DESCARGO:	El uso de la información contenida en este documento se hará bajo la completa responsabilidad del usuario. / The use of the information contained in this document will be under the exclusive responsibility of the user.	:DISCLAIMER

Estructura / Framework		
ID PERMANENTE:	http://hdl.handle.net/10810/7027	:PERMANENT ID
ESTRUCTURA:	<ul style="list-style-type: none"> • ldgp_inf001_Nerthus.pdf: este documento / this document. 	:FRAMEWORK

Cita completa recomendada / Recommended full citation

CITA:	VALLE MELÓN, José Manuel. RODRIGUEZ MIRANDA, Álvaro. <i>Representación 3D interactiva de relaciones de derivación e hiponimia entre paradigmas léxicos del inglés antiguo</i> . Informe de investigación. Laboratorio de Documentación Geométrica del Patrimonio (Universidad del País Vasco UPV/EHU) –LDGP-. 2011.	:CITATION
-------	--	-----------

INDICE

	Página
<u>0.- Introducción</u>	4
<u>1.- Objetivos</u>	4
<u>2.- Desarrollo del proyecto</u>	5
2.1.- Esquema de trabajo.....	5
2.2.- Análisis de las necesidades de representación	6
2.3.- Estudio de la base de datos.....	7
2.4.- Análisis de estándares gráficos para la plasmación de los resultados	7
2.5.- Programación de alternativas para la visualización	8
2.6.- Análisis de programas gráficos de representación no propietarios.....	9
2.7.- Adaptación a los requerimientos.....	10
2.8.- Generación de paradigmas derivados	10
2.9.- Estructuración de los resultados.....	16
<u>3.- Resultados obtenidos</u>	17
3.1.- Directos.....	17
3.2.- Indirectos	17
<u>4.- Utilización de los modelos en plataformas diversas</u>	18
4.1. Visualización en XGLORE	18
4.2.- Visualización con Graphviz.....	24

0.- Introducción

La participación del Laboratorio de Documentación Geométrica del Patrimonio de la UPV/EHU dentro del proyecto “Nerthusv2: Base de datos léxica en 3D del inglés antiguo” ha supuesto todo un reto en la forma de abordar el objeto de investigación y también en la integración dentro de los procesos de gestión, producción y desarrollo del Laboratorio centrado en la medida, representación, gestión y difusión del patrimonio en todas sus vertientes.

El presente proyecto se enmarca en las actividades de representación, gestión y difusión de un tipo de elemento patrimonial concreto dentro de la categoría de *Patrimonio Inmaterial*: el Patrimonio Lingüístico.

El análisis de la información depositada en la base de datos léxica Nerthus, por el equipo de lingüistas encabezados por el profesor Martín Arista, ha desvelado un universo de posibilidades de interacción gráfica entre los elementos abstractos que incluyen las entradas léxicas las relaciones de derivación e hiponimia que se establecen entre éstas.

1.- Objetivos

El objetivo de la incorporación de Álvaro Rodríguez Miranda y José Manuel Valle Melón, de la Universidad del País Vasco (UPV/EHU), al proyecto “Nerthusv2: Base de datos léxica en 3D del inglés antiguo” (FFI08-04448/FILO) fue transformar los paradigmas léxicos del inglés antiguo, entendidos como el conjunto de lexemas que comparten una base morfológica y semántica así como las relaciones de derivación e hiponimia que se establecen entre éstos, en un conjunto de representaciones gráficas interactivas que facilitasen el acceso, la comprensión y la manipulación de la información depositada en una base de datos léxica alfanumérica.

Se propuso el desarrollo de un procedimiento que generara de manera automática el modelo gráfico a partir de la lectura de la base de datos léxica, y que representará además de las relaciones de derivación las características propias de cada predicado presentes en la base de datos.

Como reto adicional, nos propusimos garantizar la interactividad del usuario final mediante en el manejo tridimensional del modelo formado, permitiendo el movimiento espacial libre y disponiendo en primer plano de aquel o aquellos términos que se desean analizar en cada momento.

Finalmente, con el propósito de que los modelos generados sean accesibles, se optó por utilizar formatos de archivo gráfico estandarizados y de dominio público. Al mismo tiempo, para asegurar la trazabilidad del origen y disposición de la información, se decidió asignar los metadatos correspondientes de acuerdo con estándares internaciones como *Dublin Core*¹.

¹ <http://dublincore.org/>

En este documento se proponen estrategias para garantizar la trascendencia a lo largo del tiempo de los modelos generados, con la preparación de descripciones detalladas de los procesos de confección y de los resultados obtenidos.

2.- Desarrollo del proyecto

2.1.- Esquema de trabajo

El siguiente organigrama muestra las fases de desarrollo del trabajo. Como puede apreciarse, tras un análisis inicial se procedió a separar diferentes frentes que se centraban en cada uno de los componentes (base de datos léxica existente, posibilidades de representación que ofrecía el mercado y opciones de programación propia). Con los resultados de estos estudios se procedió a decidir la forma de actuar mediante la adaptación a los requerimientos y finalmente las fases de ejecución y preparación de los resultados.

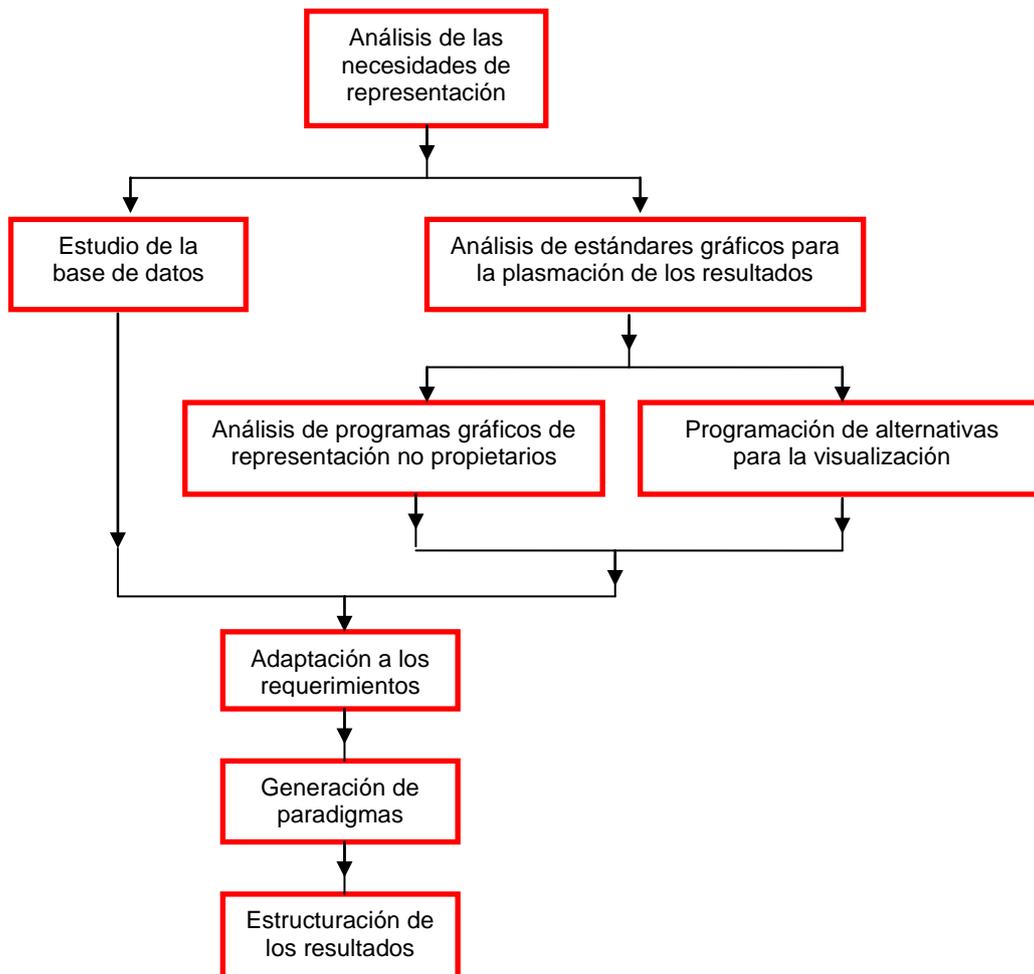


Fig.1 Organigrama del proyecto

2.2.- Análisis de las necesidades de representación

Un examen inicial de la estructura de la base de datos léxica Nerthus permite observar que se trata de una clasificación muy rica que cuenta con una desagregación muy detallada, lo que a priori permitiría plantear soluciones creativas sin introducir un esfuerzo adicional a la labor filológica de investigación; bien al contrario, se trataría más bien de representar de manera gráfica los resultados ya disponibles.

Las primeras reuniones de trabajo se centraron en el dibujo manual de algunos de los paradigmas derivativos que finalmente se han generado, y también de algunos modelos sintéticos de grafos, con el fin de concretar las expectativas del equipo de lingüistas y desarrollar las estrategias de generación de modelos por nuestra parte.

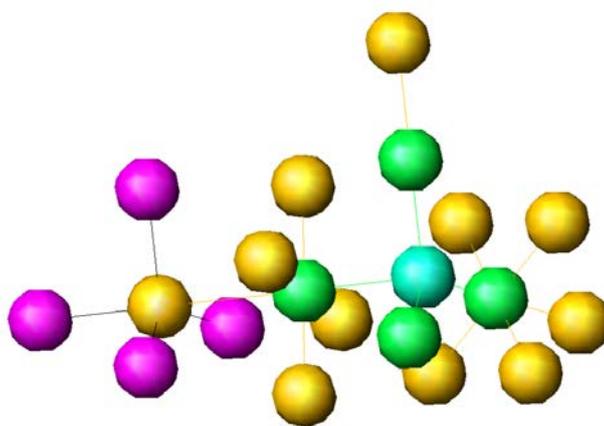


Fig.2 Modelo sintético preliminar con relación de tercer nivel

Como resultado de estos análisis se fijaron los siguientes objetivos:

- La representación debería permitir ver el nombre de cada entrada léxica relacionada con sus ascendientes y descendientes semánticos y morfológicos.
- Los nombres deberían estar siempre paralelos al plano de representación
- Cada nivel debería tener una característica cromática o morfológica que lo identificase.
- Los sucesivos procesos de derivación léxica y el resto de la información gramatical debía poder visualizarse para cada entrada, bien mediante elementos emergentes o mediante cualquier otro sistema de acceso.

Además, se acordó que los paradigmas derivativos se representaran de forma individualizada, con el fin de dar coherencia gráfica a la representación y que hubiera un grafo a modo de llave que sirviera de inicio y tránsito entre paradigmas.

2.3.- Estudio de la base de datos

En esta fase fue necesaria una descripción pormenorizada de los contenidos de cada uno de los campos que forman la base de datos léxica, y de las relaciones que se establecen entre ellos, ya que de la correcta lectura depende la interpretación acertada de las historias derivativas y la atribución de relaciones.

También en este estadio se establecieron condiciones de concatenación de campos en función de su contenido y de su tipo de derivación.

Se estudiaron las posibilidades de manipulación directa de la bases de datos y se decidió trabajar con una copia, lo que liberaba de la necesidad de implementar la solución informática programando directamente sobre el software Filemaker®, aplicación propietaria en la que encuentra la base de datos original.

Se analizaron diversas alternativas de exportación de datos, que conservaran la información completamente, optándose por la opción de exportar a un fichero de texto ASCII con los campos delimitados por comas, como se explica más adelante.

2.4.- Análisis de estándares gráficos para la plasmación de los resultados.

Como establecíamos en los objetivos, se pretende que el formato de los archivos gráficos generados responda a estándares abiertos y de estructura pública, por ello analizamos alternativas que permiten la representación de relaciones espaciales tridimensionales, pero con la condición de que su estructura sea abierta, y estén escritas en ASCII. Se evaluaron los formatos siguientes:

- DXF (*Drawing eXchange Format*): formato ASCII de intercambio gráfico de Autodesk®, cuya documentación para las diferentes versiones existentes puede consultarse en el siguiente enlace [<http://usa.autodesk.com/adsk/servlet/item?linkID=10809853&id=12272454&siteID=123112>].
- VRML (*Virtual Reality Modeling Language*): corresponde al estándar internacional ISO/IEC 14772:1997, mantenido por el *Web3D Consortium* para la distribución de modelos tridimensionales por Internet, sus especificaciones están disponibles en la página web [<http://www.web3d.org/x3d/specifications/#vrml97>].
- XML (*eXtensible Markup Language*): en realidad XML es una forma de codificación en la que los elementos se agrupan de forma jerárquica mediante parejas de etiquetas (una de apertura y otra de cierre) que engloban los elementos que están contenidos, esta estructura es repetible con el nivel de profundidad deseado. Desarrollado por el *World*

Wide Web Consortium (W3C), puede consultarse en su página web [<http://www.w3.org/standards/xml/>].

- DOT: formato desarrollado dentro del proyecto *open source* denominado *GraphViz* [<http://www.graphviz.org/content/dot-language>] aunque aceptado por otros software de visualización de grafos (*Xglore*, *Tulip*, etc). Los ficheros DOT definen una estructura topológica y contienen dos partes, en la primera se definen los nodos y en la segunda sus conexiones.

Existe una gran facilidad de transformación de unos a otros. Una diferencia fundamental es que los dos primeros (DXF y VRML) utilizan una geometría espacial basada en coordenadas mientras que los otros (XML y DOT) definen las conexiones y estructura jerárquica.

En nuestro caso, es preferible este segundo tipo, en el que lo que se definen son las conexiones (topología) y se da libertad al sistema para que sitúe los elementos en el espacio. Entre los dos formatos, se ha preferido el DOT debido a su simplicidad y porque permite relaciones múltiples a un mismo nodo que no son posibles en XML (que requiere estructuras estrictamente jerárquicas en las que un nodo no puede tener más de un antecedente directo).

2.5.- Programación de alternativas para la visualización

Paralelamente a la etapa anterior se desarrolló una programación en lenguaje JAVA de un visualizador que satisficiera los requisitos expresados anteriormente. Esta etapa fue desarrollada por Mikel Jiménez Eguiluz, como proyecto fin de carrera de la titulación de *Ingeniera Técnica en Informática de Gestión* (Escuela Universitaria de Vitoria-Gasteiz, UPV-EHU). A pesar de que la solución se desarrolló ad hoc para el proyecto, no se logró que resolviera todas las expectativas, sin duda por un error en el enfoque de la generación de la geometría espacial, centrada en exceso en la posición espacial absoluta.

La solución que se alcanzó permite disponer de la presentación gráfica de las entradas léxicas en forma de esferas que llevan asociado el nombre del predicado y establecen las relaciones con los predicados con los que se relacionan mediante líneas continuas. La imagen presentada permite la interactividad del usuario de manera que es posible desplazar, girar y escalar el conjunto acercando la imagen al predicado o predicados objeto de interés. También se ha conseguido mantener los textos de los predicados en posición paralela a la observación durante toda la manipulación.

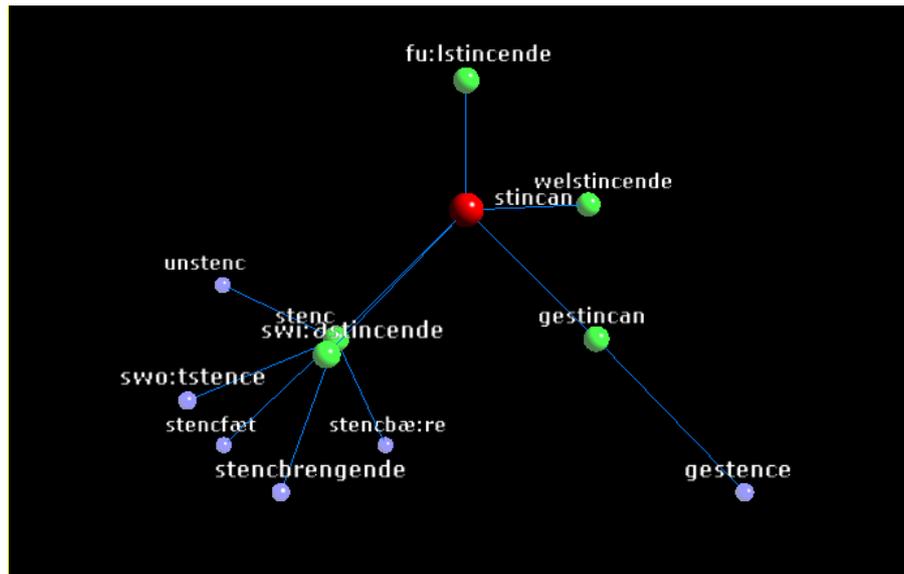


Fig.3 Ejemplo de visualización de la aplicación JAVA desarrollada

2.6.- Análisis de programas gráficos de representación no propietarios

Al mismo tiempo que se desarrollaba la programación de la solución anterior, se analizaron programas informáticos específicos para la representación de estructuras de grafos como las que se pretenden en este proyecto.

El conjunto de programas analizados ha sido:

- Base X [<http://www.informatik.uni-konstanz.de/en/arbeitsgruppen/dbis/research/basex>]
- Coatie-3D [<http://coatie3d.sourceforge.net>]
- Cortona [<http://www.cortona3d.com>]
- Graphviz [<http://www.graphviz.org>]
- Sketchup [<http://sketchup.google.com>]
- Skyrails [<http://www.visualcomplexity.com/vc/project.cfm?id=511>]
- Tulip [<http://tulip.labri.fr/TulipDrupal>]
- Xglore [<http://sourceforge.net/projects/xglore>]
- VRMLgraph [<http://vrmlgraph.i-scream.org.uk>]
- Willma [<http://wilma.sourceforge.net>]

En todos los casos se procuró que dispusieran de una versión con licencia abierta o libre de carga, acorde con el objetivo de difusión de los modelos a generar.

Se buscaron igualmente programas que pudieran leer archivos topológicos (como los antes comentados DOT o XML), abiertos y documentados.

Con estas premisas, los programas seleccionados fueron *Xglore* para la visualización tridimensional y *GraphViz* para la representación bidimensional.

2.7.- Adaptación a los requerimientos

A la vista de las necesidades expresadas por el equipo de lingüistas, las características de la base de datos léxica disponible y el análisis de alternativas realizado sobre las posibilidades de representación tanto disponibles de forma libre como programables por nosotros, se decidió proceder según los criterios siguientes:

- Se constató que las opciones de software libre ya disponibles ofrecían una variedad de opciones suficiente para las necesidades del proyecto por lo que se no se consideró conveniente continuar con el desarrollo propio.
- Se llegó a la conclusión de que los formatos topológicos (basados en conexiones) eran preferibles a los geométricos (basados en coordenadas).
- Dentro de las opciones, se optó por explotar el potencial del formato DOT que cuenta con las características favorables de estar documentado, ser fácilmente generable y ser operable por varios programas. El uso de este formato para representar datos y relaciones lingüísticas nos era desconocido pero pensábamos que factible.
- La solución adoptada debería incluir metadatos para garantizar su trazabilidad y facilitar su mantenimiento a lo largo del tiempo.

2.8.- Generación de paradigmas derivativos

Partiendo de la base de datos léxica Nerthus, se procede a exportar las columnas que interesa representar como atributos de los diferentes nodos, en este caso:

- predicate.
- status.
- alternativespellings.
- category_of_predicate.
- inflectional_morphology.
- inflectional_paradigm.
- predicate_translation.

A estos siete hay que añadir también “predicate_number” y “source_number” que, si bien no aparecerán en la lista de propiedades, son necesarios para identificar cada nodo y establecer sus conexiones.

La siguiente figura muestra la pantalla de exportación. Como puede verse, se utiliza el juego de caracteres UTF-8 para conservar los caracteres especiales.

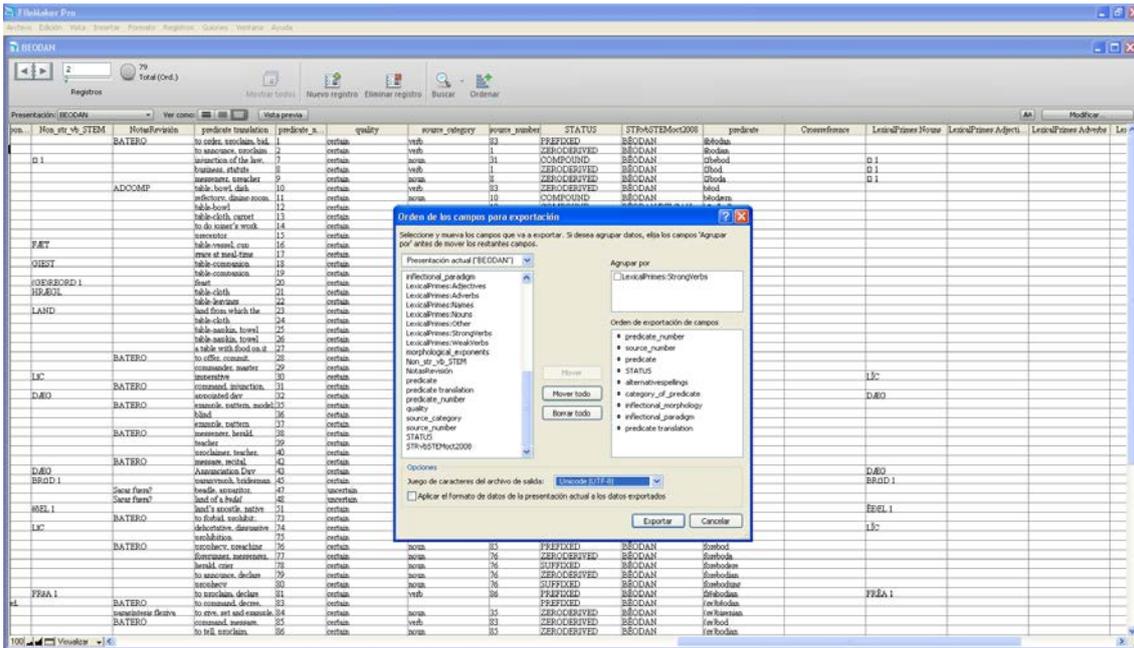


Fig.4 Ventana de exportación de la base de datos léxica

Se genera así un fichero de texto con la información de cada columna separada por comas como el siguiente extracto:

```
"2","1","ābodian","ZERODERIVED","","verb","weak 2","","to announce, proclaim"
"7","31","ǣbebod","COMPOUND","","noun","n.,","injunction of the law, command"
"8","1","ǣbod","ZERODERIVED","","noun","m.,","business, statute"
"9","8","ǣboda","ZERODERIVED","","noun","m.,","messenger, preacher"
"10","83","bēod","ZERODERIVED","bēad, biad","noun","m.,","table; bowl, dish"
"12","10","bēodbolle","COMPOUND","","noun","f.,","table-bowl"
...
```

El visor que se ha seleccionado para la representación tridimensional (Xglore) no permite los caracteres especiales, como las runas del inglés antiguo, por lo que se deben sustituir por combinaciones de caracteres ASCII. Las sustituciones a realizar fueron propuestas por el equipo de lingüistas para que la representación siguiese siendo comprensible y correcta. Las secuencias de caracteres para representar runas se basan en la solución propuesta por el Corpus de Helsinki.² A continuación se muestran los cambios realizados:

- ā → a:
- ē → e:
- ī → i:
- ō → o:
- ū → u:
- ŷ → y:
- æ → a+
- ǣ → a+:
- ð → d+
- þ → t+
- ø- → () detrás del texto.

² The Helsinki Corpus of English Texts: <http://icame.uib.no/hc/>.

Estas sustituciones se pueden realizar con la opción de “Reemplazar” del bloc de notas, excepto la del conjunto vacío “ø” que requiere localizar cada texto y modificarlo manualmente. También es necesario cambiar las diferentes versiones del término “ZERODERIVED” que se presenta en el campo “status” con el fin de simplificar las relaciones a representar.

El siguiente paso consiste en ejecutar un programa que convierte este fichero de texto a la estructura del formato DOT que, como puede verse en el siguiente extracto, incluye dos partes dentro de un conjunto “digraph {}”: en la primera se definen los nodos comenzando por su identificador (que es el campo “predicate_number”) y después los valores de las diferentes propiedades seleccionadas, después de definir los nodos se pasa a indicar las relaciones entre ellos, cada elemento se relaciona con su antecesor (el cual viene indicado en el campo “source_number”) mediante una flecha (“->”).

```
digraph {
6  [label="a:stencan" Predicate="a:stencan" Status="PREFIXED" "Alternative
spelling"="" "Lexical category"="verb" "Morphological class"="weak 1"
"Inflectional paradigm"="" Translation="to scatter" ];
82  [label="fu:lstincende" Predicate="fu:lstincende" Status="COMPOUND"
"Alternative spelling"="" "Lexical category"="adjective" "Morphological
class"="weak and strong" "Inflectional paradigm"="" Translation="foul-stinking"
];
89  [label="(ge)stencan" Predicate="(ge)stencan" Status="ZERODERIVED" "Alternative
spelling"="" "Lexical category"="verb" "Morphological class"="weak 1"
"Inflectional paradigm"="" Translation="to scatter; afflict?; to stink; to pant"
];
90  [label="(ge)stencnes" Predicate="(ge)stencnes" Status="SUFFIXED" "Alternative
spelling"="" "Lexical category"="noun" "Morphological class"="f." "Inflectional
paradigm"="" Translation="odour" ];

(...)

89 -> 6 ;
133 -> 82 ;
96 -> 89 ;
89 -> 90 ;
97 -> 96 ;
133 -> 97 ;

(...)

}
```

El programa para hacer esta transformación ha sido desarrollado para este proyecto y está programado en lenguaje Tcl por lo que puede ejecutarse tras descargar de internet una aplicación gratuita denominada *ActiveTcl* [<http://www.activestate.com/activetcl>]. Los ficheros Tcl son ASCII y de una estructura similar a cualquier otro lenguaje de programación, por lo que pueden ser editados fácilmente para adaptarlos a nuevas necesidades (por ejemplo, si se desea variar el número de atributos).

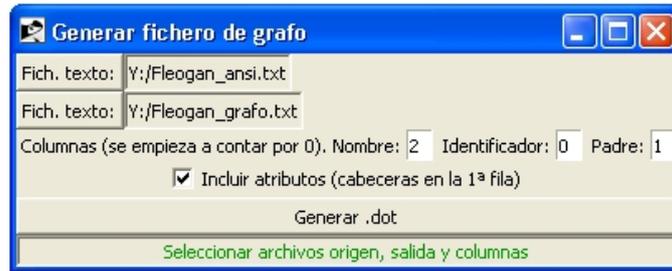


Fig.5 Software para convertir los ficheros de texto a grafos

Esta posibilidad de modificar el programa de forma ágil ha sido muy conveniente para la fase de desarrollo en la que se han ido haciendo cambios según se iban presentando los resultados. Actualmente el programa requiere que se hagan algunos cambios en el fichero ASCII con las entradas léxicas del paradigma organizadas por columnas, en concreto:

- Hay que reemplazar el separador de campos ya que originalmente utiliza la coma, pero este símbolo también aparece dentro de algunos campos; por ejemplo, en “predicate_translation”. Para sustituirlo se ha elegido el símbolo de asterisco, así que es necesario hacer dos sustituciones en los ficheros utilizando el comando de “Reemplazar” del bloc de notas:
 - o “,” → * : para eliminar la separación entre campos.
 - o Eliminar las comillas que queden reemplazándolas por una casilla en blanco.
- Añadir una primera fila que servirá de cabecera y que contendrá el nombre de cada una de las columnas.

De manera, se dispondrá de un archivo ASCII con la siguiente estructura:

```
"2", "1", "a:bodian", "ZERODERIVED", "", "verb", "weak 2", "", "to announce, proclaim"
"7", "31", "a+:bebod", "COMPOUND", "", "noun", "n.", "", "injunction of the law, command"
"8", "1", "a+:bod", "ZERODERIVED", "", "noun", "m.", "", "business, statute"
"9", "8", "a+:boda", "ZERODERIVED", "", "noun", "m.", "", "messenger, preacher"
"10", "83", "be:od", "ZERODERIVED", "be:ad, bi:ad", "noun", "m.", "", "table; bowl, dish"
"12", "10", "be:odbolle", "COMPOUND", "", "noun", "f.", "", "table-bowl"
"13", "10", "be:odcla:d+", "COMPOUND", "", "noun", "m.", "", "table-cloth, carpet"
"14", "10", "be:oddian", "ZERODERIVED", "", "verb", "weak", "", "to do joiner's work"
"15", "10", "be:odend", "SUFFIXED", "", "noun", "m.", "", "preceptor"
```

El encabezamiento aparece en varias líneas aunque sea único:

```
"Predicate          number" "Source          number" *Predicate*Status*"Alternative
spelling" *Lexical          category" *Morphological          class" *Inflectional
paradigm" *Translation
2*1*a:bodian*ZERODERIVED**verb*weak 2**to announce, proclaim
7*31*a+:bebod*COMPOUND**noun*n.**injunction of the law, command
8*1*a+:bod*ZERODERIVED**noun*m.**business, statute
9*8*a+:boda*ZERODERIVED**noun*m.**messenger, preacher
10*83*be:od*ZERODERIVED*be:ad, bi:ad*noun*m.**table; bowl, dish
12*10*be:odbolle*COMPOUND**noun*f.**table-bowl
13*10*be:odcla:d+*COMPOUND**noun*m.**table-cloth, carpet
14*10*be:oddian*ZERODERIVED**verb*weak**to do joiner's work
15*10*be:odend*SUFFIXED**noun*m.**preceptor
```

El fichero DOT generado ya se puede visualizar utilizando tanto el visor 3D (*Xglore*) como el visor 2D (*GraphViz*). En la siguiente figura se muestra la primera opción:

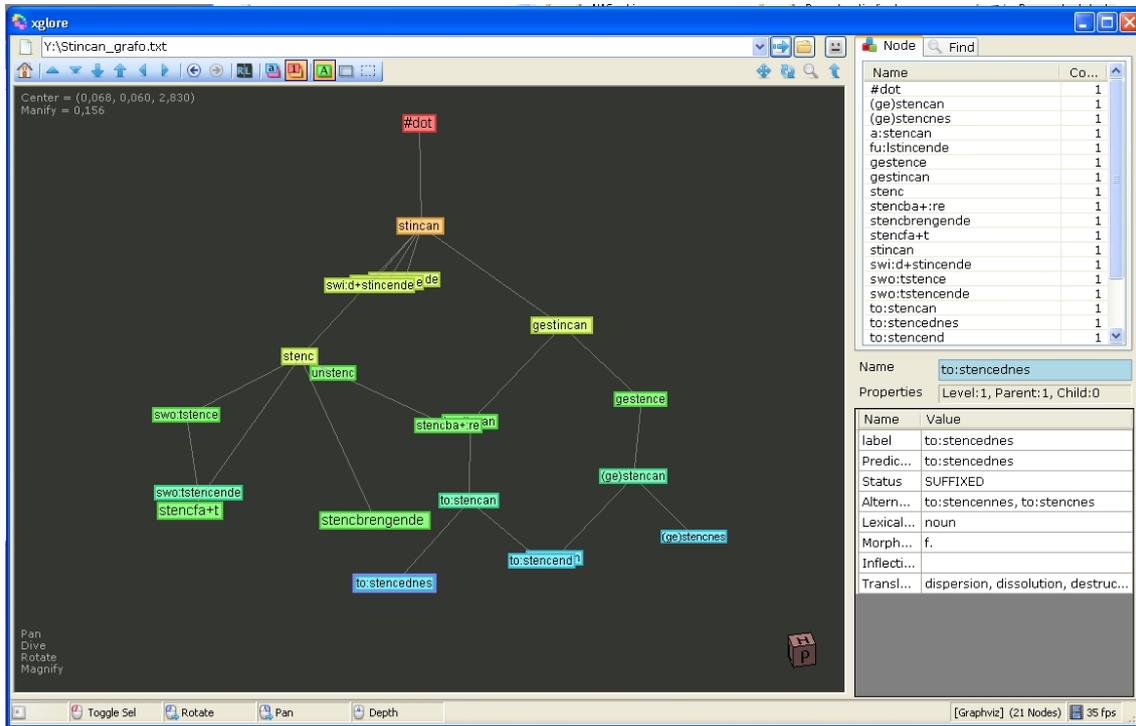


Fig.6 Vista interactiva tridimensional del grafo en el programa *Xglore*

Por cuestiones de organización, este grafo se edita para que el nodo principal (en la figura, “stincan”) no dependa directamente del nodo raíz (denominado #dot en *xglore*) sino que aparezca un nodo intermedio denominado “Nerthus”. La edición se podría hacer incorporando el código en el programa que genera el grafo pero es más sencillo editar directamente el fichero dot incorporando una línea para definir el nuevo nodo y otra para hacer la relación (señaladas en rojo en el siguiente extracto):

```

digraph {
0 [label="Nerthus" url="http://www.nerthusproject.com/3dmodel/Nerthus_grafo.txt"]
6 [label="a:stencan" Predicate="a:stencan" Status="PREFIXED" "Alternative
spelling"="" "Lexical category"="verb" "Morphological class"="weak 1"
"Inflectional paradigm"="" Translation="to scatter" ];
82 [label="fu:lstincende" Predicate="fu:lstincende" Status="COMPOUND"
"Alternative spelling"="" "Lexical category"="adjective" "Morphological
class"="weak and strong" "Inflectional paradigm"="" Translation="foul-stinking"
];
89 [label="(ge)stencan" Predicate="(ge)stencan" Status="ZERODERIVED" "Alternative
spelling"="" "Lexical category"="verb" "Morphological class"="weak 1"
"Inflectional paradigm"="" Translation="to scatter; afflict?; to stink; to pant"
];
90 [label="(ge)stencnes" Predicate="(ge)stencnes" Status="SUFFIXED" "Alternative
spelling"="" "Lexical category"="noun" "Morphological class"="f." "Inflectional
paradigm"="" Translation="odour" ];

(...)

0 -> 133 ;
89 -> 6 ;
133 -> 82 ;
96 -> 89 ;
89 -> 90 ;

```

```
97 -> 96 ;
133 -> 97 ;
133 -> 129 ;

(...)

}
```

Como puede apreciarse, se aprovecha el nuevo nodo para incorporarle un valor en el campo “url”, que se utiliza para enlazar cada paradigma derivativo con el grafo general mediante la dirección web en la que se encuentra alojado.

Otro aspecto que el programa no contempla y que debe incluirse de forma manual incorporando las nuevas relaciones que sean necesarias es el hecho de que un nodo tenga más de un ascendiente directo.

La última modificación que se realiza consiste en incorporar los metadatos. Se ha elegido el esquema *Dublin Core* por su gran versatilidad y difusión. Como el formato DOT no considera la opción de incluir metadatos como tales se ha decidido incluirlos por duplicado, primero como comentarios (marcados en verde en el siguiente extracto) y también dentro del nodo “Nerthus” como atributos (marcados en azul). Por lo tanto, el fichero de grafo queda con el siguiente aspecto:

```
digraph {
  // Metadata, Dublin Core
  /* Dc.Title = Nerthus 3D: Lexical Database of Old English */
  /* Dc.Creator = Javier Martin Arista */
  /* Dc.Subject = Old English, lexicography, lexicology, morphology, 3D linguistic
  representation */
  /* Dc.Description = The lexical database of Old English Nerthus consists of ca.
  30,000 lexical entries provided with information on lexical category,
  inflectional morphology, spelling variants and derivational morphology. The
  database can be accessed online (www.nerthusproject.com/search-database) and
  searched by means of an engine that allows for a variety of queries. The 3D
  representation assures that lexical relations are displayed explicitly and
  exhaustively, and, moreover, qualifies as a semantically annotated resource
  devised for dissemination through the Web 3.0. */
  /* Dc.Publisher = Javier Martin Arista, Elisa Gonzalez Torres, Carmen Novo
  Urraca, Miguel Lacalle Palacios */
  /* Dc.Contributor = Laboratorio de Documentacion Geométrica del Patrimonio (LDGP)
  - Grupo de Investigacion en Patrimonio Construido (GPAC) - Universidad del Pais
  Vasco / Euskal Herriko Unibertsitatea (UPV/EHU) - SPAIN*/
  /* Dc.Contributor = Jose Manuel Valle Melon, Alvaro Rodriguez Miranda
  (Laboratorio de Documentacion Geometrica del Patrimonio UPV-EHU) */
  /* Dc.Date = 20111222 */
  /* Dc.Type = Lexical database, 3D */
  /* Dc.Format = dot (Graphviz) - to be represented also in 3D by xglore */
  /* Dc.Identifier = http://www.nerthusproject.com/3Ddatabase */
  /* Dc.Source = Nerthus Project, Universidad de La Rioja */

  0 [label="Nerthus" url="http://www.ehu.es/docarq/Nerthus/Nerthus_grafo.txt"
  Metadata_schema="Dublin Core" Dc_Title="Nerthus 3D: Lexical Database of Old
  English" Dc_Creator="Javier Martin Arista" Dc_Subject="Old English, lexicography,
  lexicology, morphology, 3D linguistic representation" Dc_Description="The lexical
  database of Old English Nerthus consists of ca. 30,000 lexical entries provided
  with information on lexical category, inflectional morphology, spelling variants
  and derivational morphology. The database can be accessed online
  (www.nerthusproject.com/search-database) and searched by means of an engine that
  allows for a variety of queries. The 3D representation assures that lexical
  relations are displayed explicitly and exhaustively, and, moreover, qualifies as
  a semantically annotated resource devised for dissemination through the Web 3.0."
  Dc_Publisher="Javier Martin Arista, Elisa Gonzalez Torres, Carmen Novo Urraca,
  Miguel Lacalle Palacios" Dc_Contributor1="Laboratorio de Documentacion Geométrica
  del Patrimonio (LDGP) - Grupo de Investigacion en Patrimonio Construido (GPAC) -
  Universidad del Pais Vasco / Euskal Herriko Unibertsitatea (UPV/EHU) - SPAIN"
  Dc_Contributor2="Jose Manuel Valle Melon, Alvaro Rodriguez Miranda (Laboratorio
  de Documentacion Geometrica del Patrimonio UPV-EHU)" Dc_Date="20111222"
```

```

Dc_Type="Lexical database, 3D" Dc_Format="dot(Graphviz) - to be represented also
in 3D by xglore" Dc_Identifier="http://www.nerthusproject.com/3Ddatabase"
Dc_Source="Nerthus Project, Universidad de La Rioja" 6 [label="a:stencan"
Predicate="a:stencan" Status="PREFIXED" "Alternative spelling"="" "Lexical
category"="verb" "Morphological class"="weak 1" "Inflectional paradigm"=""
Translation="to scatter" ];
82 [label="fu:lstincende" Predicate="fu:lstincende" Status="COMPOUND"
"Alternative spelling"="" "Lexical category"="adjective" "Morphological
class"="weak and strong" "Inflectional paradigm"="" Translation="foul-stinking"
];
89 [label="(ge)stencan" Predicate="(ge)stencan" Status="ZERODERIVED" "Alternative
spelling"="" "Lexical category"="verb" "Morphological class"="weak 1"
"Inflectional paradigm"="" Translation="to scatter; afflict?; to stink; to pant"
];
90 [label="(ge)stencnes" Predicate="(ge)stencnes" Status="SUFFIXED" "Alternative
spelling"="" "Lexical category"="noun" "Morphological class"="f." "Inflectional
paradigm"="" Translation="odour" ];
96 [label="gestence" Predicate="gestence" Status="ZERODERIVED" "Alternative
spelling"="gesta+nce" "Lexical category"="adjective" "Morphological class"="weak
and strong" "Inflectional paradigm"="" Translation="odoriferous"];

(...)

0 -> 133 ;
89 -> 6 ;
133 -> 82 ;
96 -> 89 ;
89 -> 90 ;
97 -> 96 ;
133 -> 97 ;
133 -> 129 ;

(...)

}

```

2.9.- Estructuración de los resultados

Además de representar los paradigmas derivativos individuales, se pretende enlazarlos de manera que el usuario pueda navegar y entre ellos. Para ello se ha creado un grafo llave que parte de un nodo denominado “Nerthus” y que enlaza los paradigmas existentes

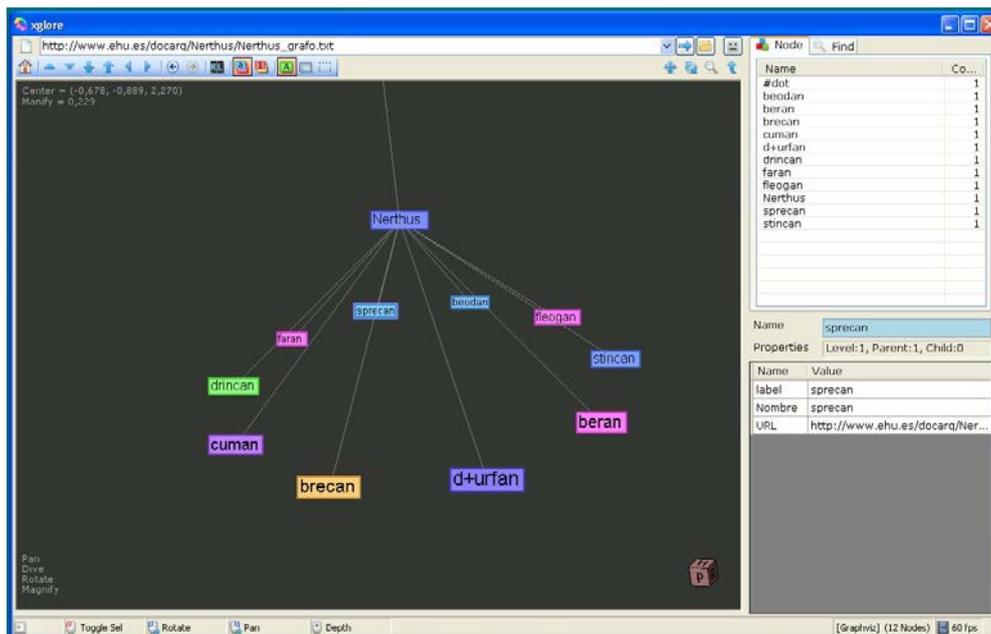


Fig.7 Grafo llave que permite la navegación entre los paradigmas existentes

Como puede verse en la figura anterior, se ha utilizado la posibilidad de almacenar atributos (zona inferior derecha) para incorporar el atributo "URL" que incluye la ubicación de los ficheros de cada paradigma.

Todos los ficheros se han almacenado en un servidor web, de tal forma que es posible acceder a ellos desde cualquier equipo con conexión a internet sin necesidad de disponer de una copia local. Esto permite centralizar el acceso y permite que los usuarios accedan siempre a la versión actualizada disponible en el servidor.

Para completar la navegación, los grafos de los paradigmas comienzan en un nodo "Nerthus" que permite volver al grafo llave y, desde aquí, presentar un nuevo paradigma.

3.- Resultados obtenidos

3.1.- Directos

La propuesta abordada y validada dentro del espacio 3D del proyecto, ha supuesto el establecimiento de un nuevo marco de trabajo y desarrollo para el proyecto NERTHUS, tanto actual como futuro.

Se ha representado un total de diez paradigmas, como ejemplo y validación de la propuesta desarrollada. El nivel de complejidad alcanzado ha permitido depurar problemas de programación y de concepto suscitados en el proceso de diseño y ejecución del proyecto.

Finalmente, se ha desarrollado un programa específico para la generación de nuevos paradigmas, partiendo de la información depositada en la base de datos léxica Nerthus.

3.2.- Indirectos

Un resultado indirecto y no previsto en el diseño inicial del proyecto ha venido dado por el hecho de que la generación de modelos tridimensionales supone una herramienta de extraordinaria utilidad para la detección de errores en la base de datos léxica, lo que permite su depuración y validación.

Como puede verse en las siguientes imágenes, en la izquierda existen nodos cuyas relaciones están mal establecidas y aparecen representados como dependiendo directamente del nodo "#dot". Tras la revisión, se corrige la base de datos y se obtiene la figura de la derecha.



Fig.8 Utilidad de la representación gráfica para detectar errores en la base de datos léxica

4.- Utilización de los modelos en plataformas diversas

4.1. Visualización en Xglore

Xglore es un explorador gráfico 3D para grafos tipo árbol acíclicos. Los nodos del grafo se colocan en el espacio 3D de manera circular, en base a su nivel desde el nodo raíz. La versión que proponemos utilizar es compatible con los formatos DOT y XML. Se encuentra disponible únicamente para plataforma Windows.

La utilización del programa Xglore parte de su descarga desde el enlace <http://sourceforge.net/projects/xglore/>. El programa puede ser utilizado libremente, ya que dispone de licencia "BSD license", cuyas características pueden leerse en <http://www.opensource.org/licenses/BSD-2-Clause>

Una vez descargado el programa, se dispone de una carpeta comprimida como la de la figura, en cuyo interior se encuentra el programa. Basta con descomprimir esta carpeta en una nueva y aparecen los siguientes archivos:

Nombre	Tamaño	Tipo	Fecha de modificación
GraphvizDot.dll	47 KB	Extensión de la apli...	09/05/2010 1:29
sample	2 KB	Documento XML	24/04/2010 18:17
xglore	192 KB	Aplicación	09/05/2010 1:29
xglore.exe	2 KB	Archivo CONFIG	24/04/2010 18:17
XML.dll	6 KB	Extensión de la apli...	09/05/2010 1:29

Fig.9 Ficheros de la carpeta en la que se presenta el programa Xglore

El programa Xglore no precisa instalación y se ejecuta automáticamente al pulsar sobre el icono .



Xglore.exe

Abriendo la aplicación se visualiza la siguiente pantalla:

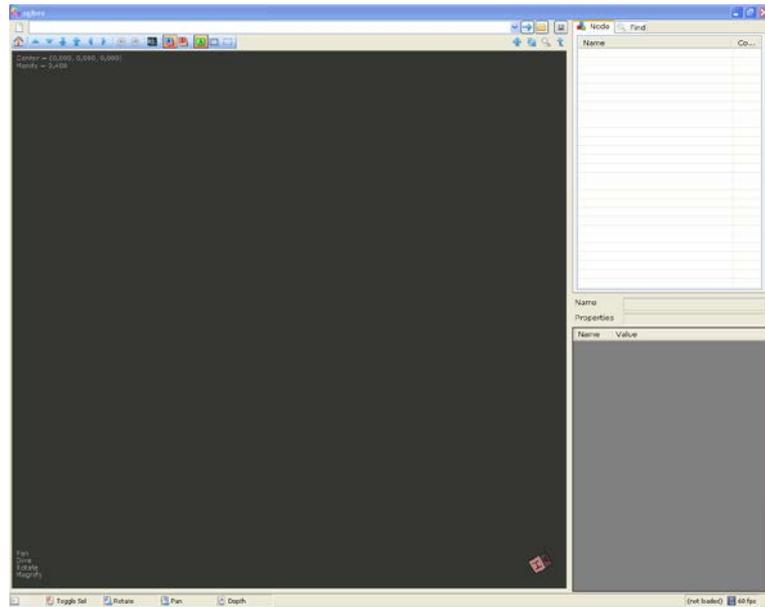


Fig.10 Aspecto del programa Xglore

Para comenzar la visualización del proyecto Nerthus, hay que acceder a la dirección URL en la que se encuentran las presentaciones de los paradigmas léxicos, http://www.nerthusproject.com/3dmodel/Nerthus_graph.txt, donde aparece el árbol generado en el proyecto, con los diez paradigmas completados como ejemplo.

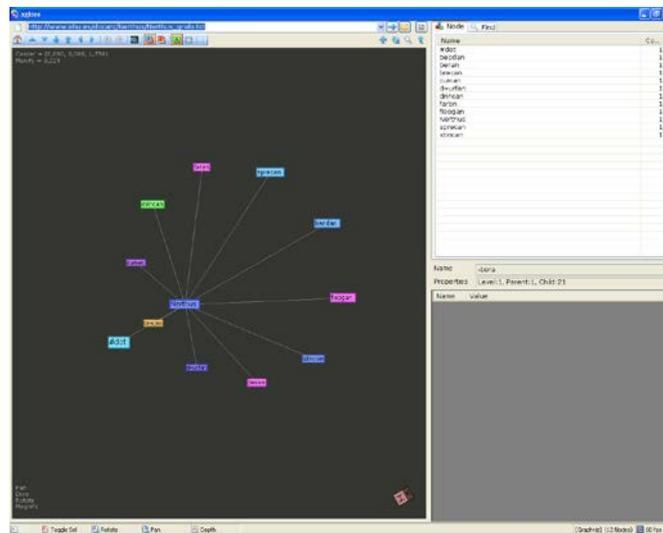


Fig.11 Visualización 3D interactiva en Xglore

El movimiento tridimensional se realiza moviendo el ratón pulsado sobre el espacio tridimensional (área en negro). Permite giros, desplazamientos y cambios de escala dependiendo del botón del ratón que se mantenga pulsado (izquierdo o derecho) y, adicionalmente, con teclas ("Shift" para los cambios de escala o "Ctrl" para movimientos continuos).

Seguidamente comentaremos las opciones de visualización.

- Deslizando el ratón sobre una de las entradas léxicas se produce la iluminación de ésta, sus ascendientes y descendientes. En el listado de entradas léxicas de la derecha aparece resaltado, al mismo tiempo que en la parte inferior se muestran las características propias de la entrada léxica en cuestión, así como el enlace URL para proceder a su visualización. Es suficiente con copiar esta información y pegarla en la línea situada en la cabecera para acceder al paradigma completo de una entrada léxica dada.

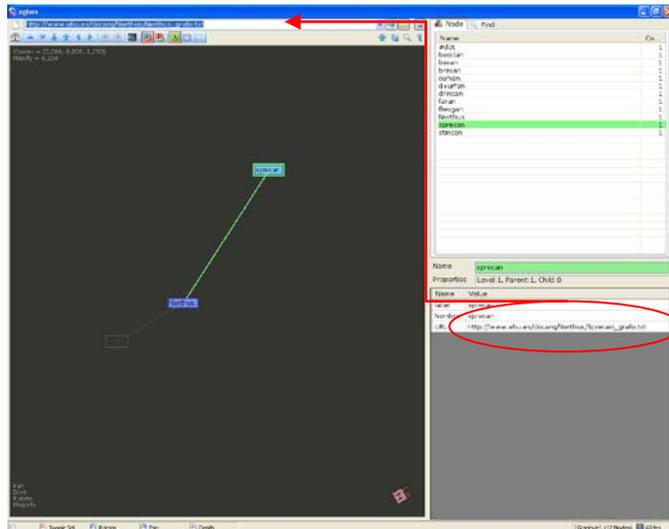


Fig.12 Zona donde se presentan los atributos de los nodos seleccionados en Xglore

- La inspección de una entrada léxica cualquiera puede realizarse de varias maneras, bien colocando el cursor del ratón sobre la entrada léxica en el gráfico (búsqueda gráfica), o bien localizándola en el listado de la derecha (palabra completa), o también con la opción "Find", situada en la parte superior derecha, de manera que la introducción progresiva de caracteres va acotando la búsqueda (búsqueda por aproximación).

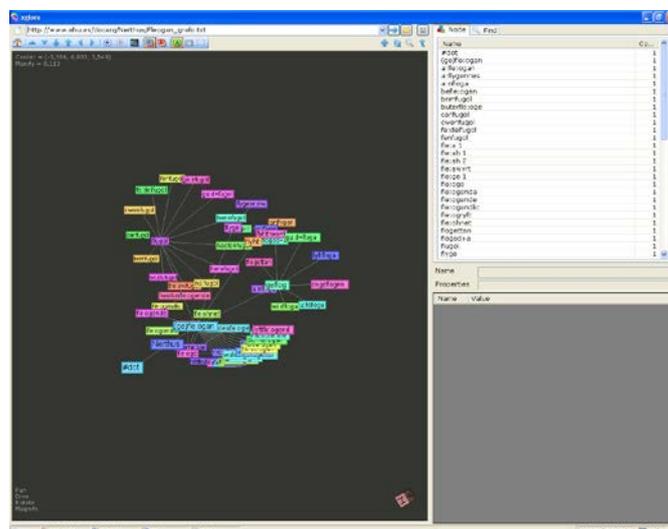


Fig.13 Búsqueda textual dentro de la información de cada nodo en Xglore

- Selección de una entrada léxica concreta, en el caso de la figura siguiente “geflog” del paradigma “Fleogan”, que muestra sus atributos en la parte inferior izquierda.

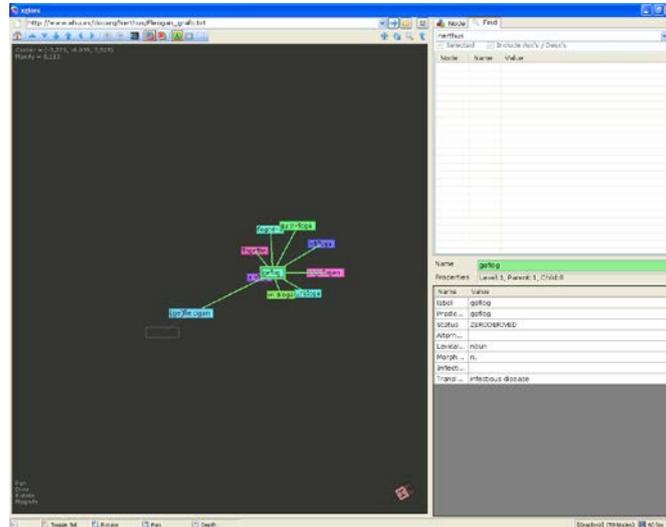


Fig.14 Búsqueda textual dentro de la información de cada nodo en Xglore

- Utilizando la tecla “Alt” sobre el espacio tridimensional se pueden seleccionar varios nodos al mismo tiempo.
- Movimientos:
 - o Si no se ha seleccionado ninguna entrada léxica o grupo de predicados, al pulsar sobre la zona de tapiz negro de la visualización, con el botón izquierdo del ratón, se produce un giro del conjunto entorno al centroide de los objetos. Si se pulsa sobre el botón derecho, se produce el desplazamiento del conjunto.
 - o Si se tiene seleccionado uno o varios nodos el giro y el desplazamiento se realizan centrados en la selección.
 - o Cuando se tiene seleccionada una entrada léxica, la acción sobre la rueda del ratón permite visualizar, de manera progresiva, el conjunto de los antecedentes de dicha entrada.
 - o Al pulsar la tecla shift, junto al botón izquierdo del ratón, se consigue una ampliación de las relaciones gráficas, posibilitando la separación de aquellos nodos que están próximos. Si la tecla shift se pulsa al mismo tiempo que el botón derecho, se obtiene la ampliación o reducción del conjunto, pero manteniendo las relaciones espaciales entre los nodos.

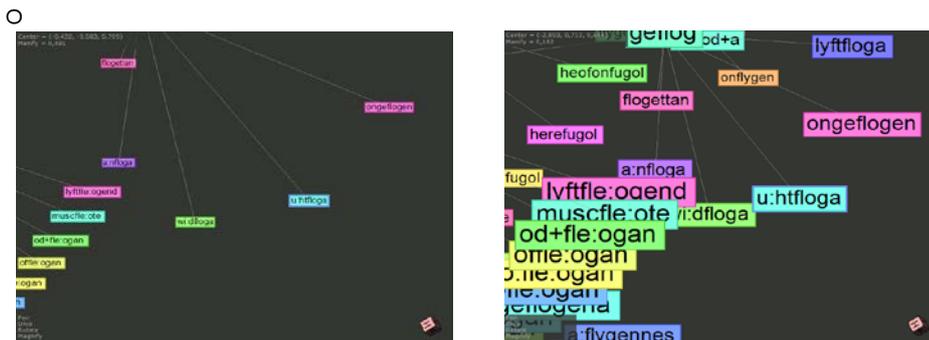


Fig.15 Efecto de la combinación de la tecla Shift, junto con el botón izquierdo y derecho del ratón

- Opciones del programa Xglore.



Fig.16 Barras de opciones en Xglore

- El icono  devuelve la visualización a la posición inicial, descargando todas las selecciones y manipulaciones realizadas.
- Las opciones de esta sección  muestran las vistas del conjunto, desde una posición superior, inferior, frontal, posterior, lateral izquierda y derecha.

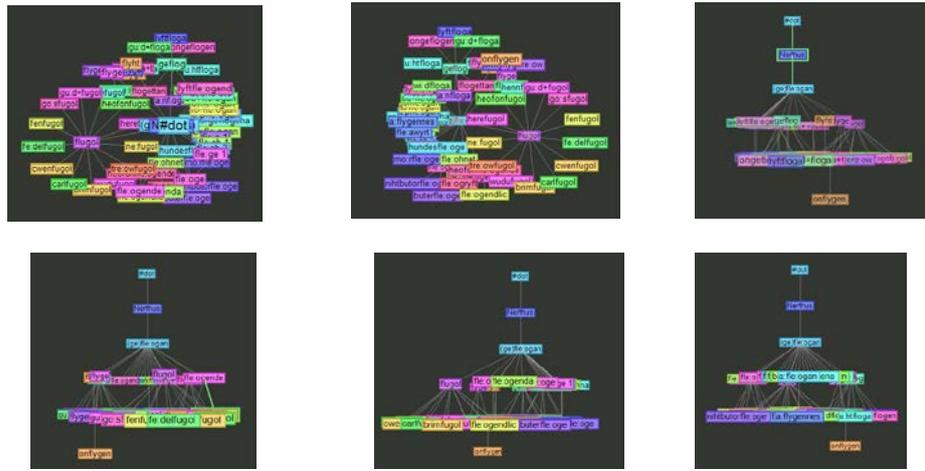


Fig.17 Ejemplos de vistas predefinidas en Xglore

- Es posible volver a una visualización anterior o siguiente, mediante los botones .
- Xglore dispone de la opción de visualización estereoscópica, mediante un estereóscopo. Para generar esta función basta con pulsar sobre el botón .
- También es posible cambiar la coloración de los nodos mediante los botones . La opción de la izquierda realiza la coloración de manera alfabética, mientras que la de la derecha colorea en función del nivel o categoría del nodo.

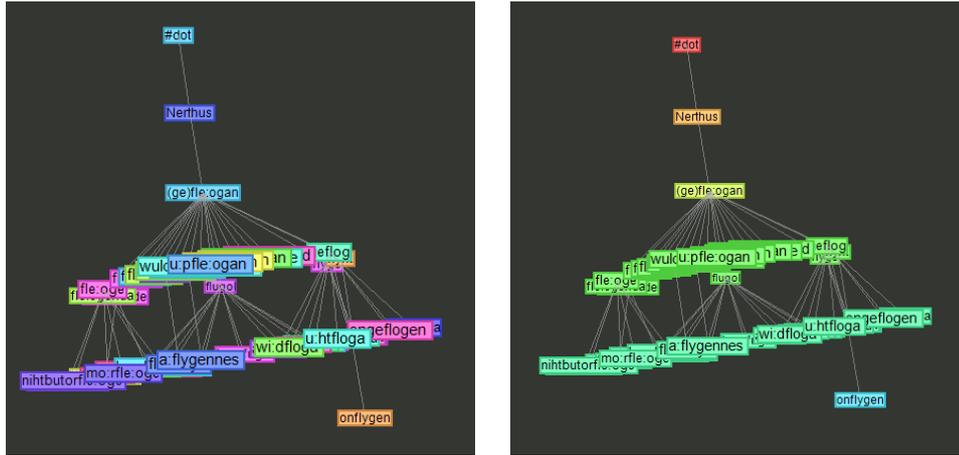


Fig.18 Opciones de color de los nodos

- Opciones de representación de los nodos no activos . La primera de las opciones muestra todos los nodos, tanto activos como no; la segunda representa los nodos no activos como bocetos; y la tercera no representa los nodos no activos.

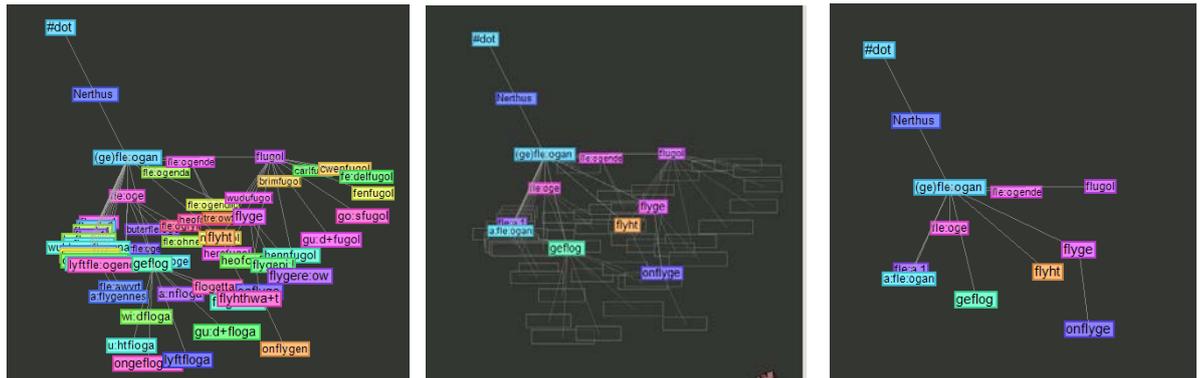


Fig.19 Opciones de visualización de nodos

Existe un video de demostración sobre el programa que facilita su uso, disponible en <http://www.youtube.com/watch?v=xEe5mHLvdIk>

4.2.- Visualización con *Graphviz*

El formato DOT para la representación de grafos puede utilizarse con diferentes programas. Uno de los que más rendimiento ofrece es el responsable de su definición: *GraphViz*. A diferencia de *Xglore*, la representación es bidimensional y no permite movimiento, aunque ofrece la posibilidad de edición interactiva de los grafos.

Como se ha indicado previamente, la descarga gratuita del software se puede realizar desde la web [http://www.graphviz.org/] y está disponible para múltiples plataformas (*Linux, Windows y Mac*).

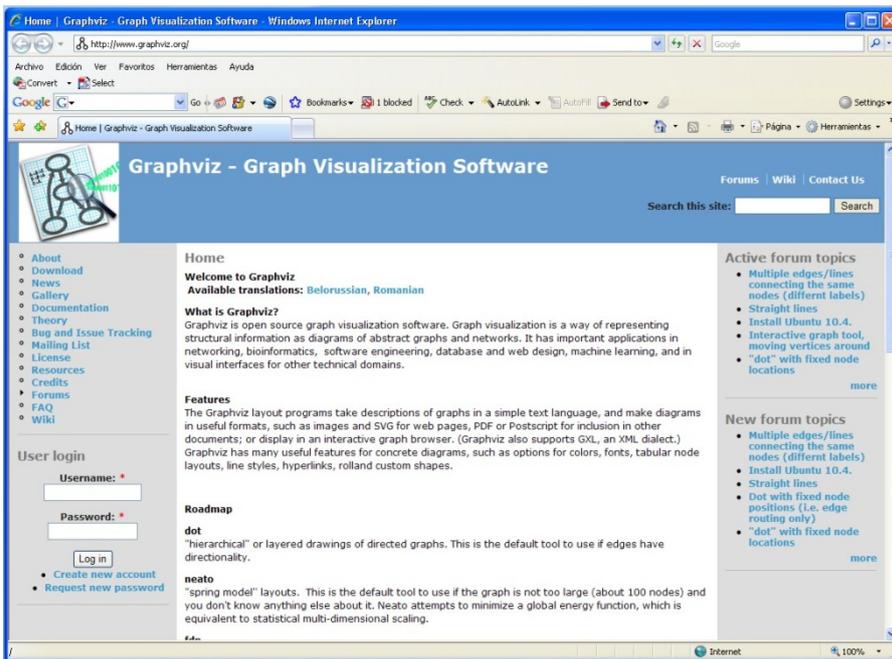


Fig.20 Página web del software *GraphViz*

Una vez cargado un fichero DOT con la estructura del grafo, se presentan dos ventanas. En una aparece el código del grafo en forma de texto y en la otra la representación gráfica. El grafo puede editarse en la ventana de texto y se comprueban las modificaciones directamente en la parte gráfica.

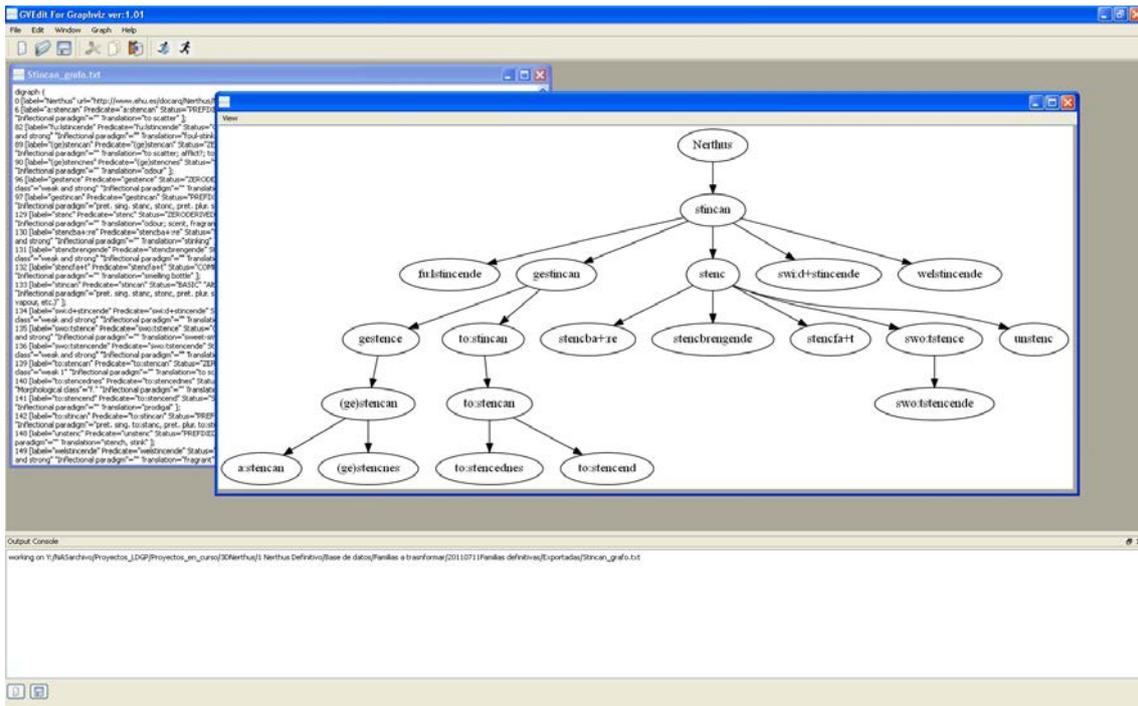


Fig.21 Ventana de texto y grafo en *GraphViz*

Posteriormente, existe la posibilidad de exportar el grafo a diferentes formatos gráficos (imágenes, pdf, VRML, etc...).

GraphViz permite explotar toda la funcionalidad del formato DOT, lo que incluye la posibilidad de modificar la simbología como se representan tanto los nodos como los enlaces (formas, colores, etc...) y además soporta un juego mayor de caracteres que el ASCII simple, permitiendo así nuevas opciones de representación gráfica de los resultados.

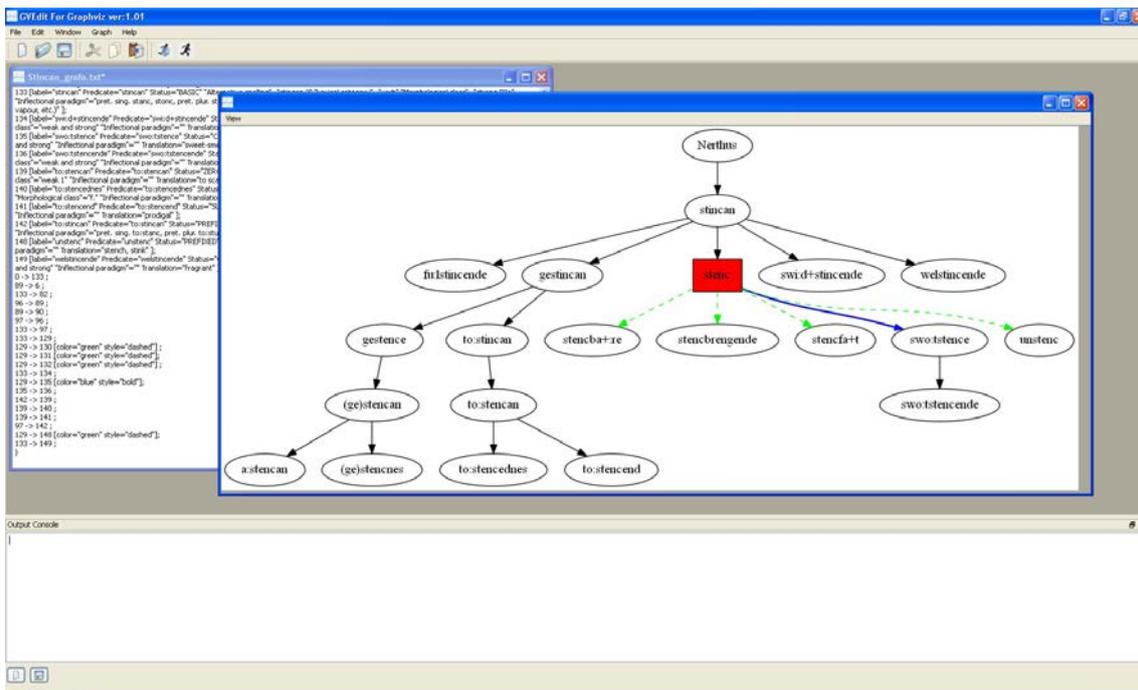


Fig.22 Edición de la simbología en *GraphViz*

Por otra parte, la capacidad de edición dinámica del grafo es muy útil para comprobar errores en los enlaces y realizar las correcciones oportunas.



LABORATORIO DE DOCUMENTACIÓN GEOMÉTRICA DEL PATRIMONIO
Grupo de Investigación en Patrimonio Construido -GPAC- (UPV-EHU)



UPV EHU

Aulario de las Nieves, edificio de Institutos Universitarios
C/ Nieves Cano 33, 01006 Vitoria-Gasteiz (España-Spain).
Tfno: +34 945 013222 / 013264
e-mail: ldgp@ehu.es web: <http://www.ldgp.es>
