

RESTITUYENTES Y RESTITUIBLES LA RESTITUCION TEXTUAL EN EL BDFSN*

Jesus-Luis Cunchillos - José-Angel Zamora

En «Tratamiento informático de las lenguas Siro-palestinas del II y I milenio a. C. Desarrollo actual y estrategia de futuro», escribíamos:

«¿Qué filólogo no se ha encontrado alguna vez con un fragmento de palabra que necesitaría restituir para poder leer el texto, objeto de su estudio? ¿Qué filólogo o comentarista de textos no ha sentido la necesidad de disponer de una concordancia para dilucidar los contextos en los que aparece cada una de las palabras o cadenas grafemáticas incompletas de las que se compone su texto objeto de estudio? ¿Alguien ha catalogado las reglas que hay que seguir o los pasos que hay que dar para realizar alguna de las operaciones antes mencionadas? Y si alguien ha establecido reglas o indicado los pasos ¿quién ha jerarquizado los pasos a dar o las reglas a ejecutar? No podemos extendernos aquí en la clarificación del punto de partida, los hechos epigráfico-filológicos, ni en las reglas que relacionan los datos... en restituciones y en concordancias. Lo haremos más detalladamente en otro lugar». Nos proponemos mostrar aquí las posibilidades de restitución de palabras incompletas mediante el uso de la informática.

LA HERMENEUMÁTICA. EL PROYECTO BDFSN II

En el proyecto «Banco de datos filológicos semíticos noroccidentales II: ugarítico, fenicio y púnico» se han llevado a cabo durante los últimos años diferentes realizaciones informáticas. Con ellas se pretende formalizar primero y automatizar después el proceso de análisis crítico de textos. Se intentan seguir los diferentes pasos que un investigador, consciente o inconscientemente, realiza en su camino hacia la comprensión del texto, desde su primer enfrentamiento con el objeto de estudio hasta la interpretación crítica final. Esta labor se ha denominado *Hermeneumática* (acrónimo de «interpretación automática»)². Exige un esfuerzo de introspección por parte del especialista que haga evidentes todos y cada uno de los pasos del proceso de investigación. Una vez analizados y formalizados los pasos se reproduce el trabajo informáticamente. Se parte de la informatización racionalizada de la documentación,

* Trabajo realizado en el marco del Proyecto de Investigación «Banco de datos filológicos semíticos noroccidentales II: ugarítico, fenicio y púnico», PB 93-0107, financiado por la DGCYT.

¹ Conferencia pronunciada por J.-L. Cunchillos en el *1st International Symposium. A Decade in Ancient Near Eastern Studies (1986-1996)*. Universidad Autónoma de Madrid, 2-4 de Diciembre de 1996 (en prensa).

² Véase J.-L. Cunchillos, *Des outils pour le développement de l'herméneumatique*, en *IOSOT, Congress Volume* (Cambridge 1995), Leiden 1997, 97-124.

creando bancos de datos gestionados por bases relacionales. El sistema de gestión de la Base de Datos se amplía y adapta a las necesidades del filólogo. Se relacionan los datos de acuerdo con alguno de los pasos que el especialista recorre a lo largo de la investigación. Diferentes aplicaciones se unen a los datos del banco (datos textuales, morfológicos, léxicos), formando diferentes módulos. Cada nuevo módulo, con más datos y herramientas, se añade e integra con los ya existentes.

El proyecto trabaja sobre varios *corpora* documentales, correspondientes a épocas y lugares diferentes, para los que se realizan bancos de datos y aplicaciones informáticas específicos. El prototipo de todos ellos fue el BDFSN ugarítico, cuyos módulos 1 y 2 fueron publicados en 1996³. Alrededor de los datos básicos (todos los textos conocidos en lengua ugarítica, con las diferentes lecturas de los especialistas⁴) se integran una amplia serie de herramientas, que permiten trabajar con la información primera y extraer de ella información nueva. El proyecto ultima ahora un tercer módulo.

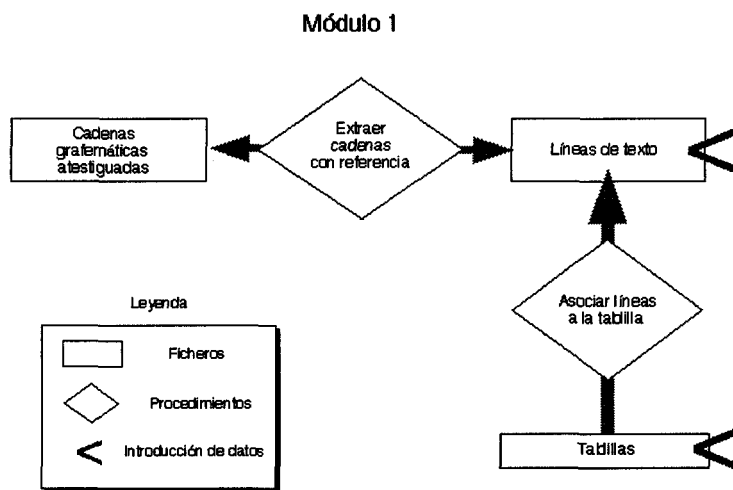


Fig. 1: Diagrama Entidad / Relación del módulo 1 del BDFSN

³ J.-L. Cunchillos, R. Cervigón, J.-P. Vita, J.-M. Galán y J.-Á. Zamora, *Generador de Segmentaciones, Restituciones y Concordancias (GSRC)*, CD-ROM, Madrid 1996.

⁴ Publicados también en soporte papel en TU: J.-L. Cunchillos y J.-P. Vita, *Textos Ugaríticos*, Madrid 1993.

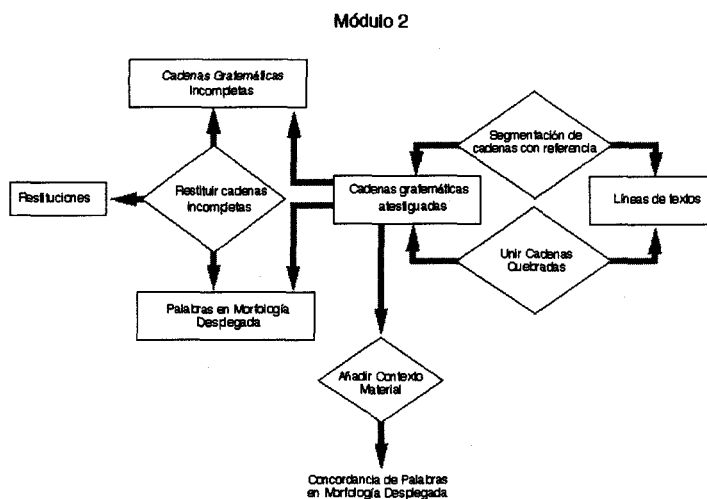


Fig. 2: Diagrama Entidad / Relación del módulo 2 del BDFSN

La programación que se ocupa de proponer restituciones a partes del texto dañadas en el documento original se ha llamado *Restituciones*. Se engloba en el segundo módulo del Banco de Datos. Es una de las tres aplicaciones básicas, aunque no únicas⁵, que actúan sobre el conjunto de la documentación.

La primera, *Segmentaciones*, está destinada a extraer del texto de partida Cadenas Grafemáticas (CG), sucesiones de caracteres que forman una unidad lingüística. El BDFSN ugarítico recoge todos los textos en lengua ugarítica publicados, así como las diferentes colaciones de las tablillas realizadas por los epigrafistas en diferentes momentos. La mayor parte de los textos, aunque no todos, provienen de las tablillas cuneiformes alfabéticas encontradas en Ras Shamra. El banco de datos almacena actualmente (mayo de 1998), en lo que se refiere a información textual, 1.361 tablillas, con un total de 26.944 líneas de texto. Las líneas de texto se hallan en transcripción, conservando todos los accidentes epigráficos connotados por el epigrafista. Sobre estas líneas de texto actúa *Segmentaciones*. Tras diferentes operaciones⁶, las cadenas resultantes son almacenadas. La mayor parte de estas

⁵ Entre las primeras herramientas con las que cuenta el BDFSN se encuentran los diferentes procesos de gestión, búsqueda, introducción y presentación de datos. Algunos procesos involucran a otros, como la limpieza de connotaciones epigráficas del texto, y se relacionan con los tres principales. Dos aplicaciones revisten también gran importancia. *División de cadenas* divide una cadena atestiguada cuando el especialista considera que contiene dos palabras diferenciables en morfología desplegada. *Unión de cadenas* une cadenas divididas por un cambio de línea. Ambos procedimientos se explicarán al detalle en otros artículos.

⁶ Que incluyen decisiones del experto y serán también explicadas en otro artículo.

cadena son Palabras en Morfología Desplegada (PMD). Otras, en zonas de texto dañadas (frecuente en el caso de las tablillas, pero también de otros soportes de la escritura, como el papiro, pergamino, papel, etc.), no pasan de ser Cadenas Grafemáticas Incompletas (CGI). Las CGI son cadenas que quedan cortadas por lagunas textuales, espacios ilegibles. El epigrafista utiliza signos convencionales para señalar de que forma se presenta una laguna textual. Muchas veces el espacio epigráfico ilegible no es determinado por el epigrafista. Por ejemplo, el investigador connota cadenas que se cortan abruptamente, como *m[...]*, o cuyo comienzo se ha perdido, como *[...]kt*, o perdidas en su parte intermedia, como *m[...]t*. En tales casos, la cadena se almacena en el Banco de Datos tomando los caracteres legibles y representando el espacio epigráfico indeterminado mediante un guión largo: *m—*, *—kt*, *m—t*. Si el epigrafista ha connotado el número de caracteres ilegibles, éstos se representan mediante un guión corto, por ejemplo *ml-*, *--kt*, o *m--t*.

Sobre todas ellas actúan dos herramientas más, ya en el segundo módulo. La que ahora nos ocupa, *Restituciones*, trata de encontrar posibles soluciones (Restituyentes) a las Cadenas Grafemáticas Incompletas. *Concordancias* gestiona la información relativa a la ubicación de las cadenas (completas e incompletas). Proporciona listas con las ocurrencias contextualizadas de cada CG. Esto es, elabora concordancias de las palabras ugaríticas en morfología desplegada (PMD) y de todas las cadenas incompletas (CGI).

RESTITUCIONES: RESTITUYENTES Y RESTITUIBLES

Aunque las cadenas incompletas (CGI) quedan fuera de las obras de referencia y no reciben habitualmente tratamiento de conjunto, en general cada epigrafista o filólogo intenta llenar los huecos más llamativos o más significativos del documento que estudia. La herramienta *Restituciones* actúa como el especialista al restituir a nivel puramente epigráfico. La programación del BDFSN obtiene las posibles reconstrucciones siguiendo la labor que el especialista llevaría a cabo con cada cadena incompleta (CGI). BDFSN lo hace con los datos disponibles en el banco en ese momento del proceso, es decir, con el vocabulario atestiguado. Las restituciones publicadas hasta el momento (en CPU⁷ y en GSRC⁸, como veremos) se basan enteramente en el vocabulario atestiguado. Esto es, con las Palabras en Morfología Desplegada (PMD) extraídas por la programación *Segmentaciones* de los textos introducidos⁹. El ordenador compara cada Cadena Grafemática Incompleta (CGI) con el vocabulario existente. Si existen Palabras en Morfología Desplegada (PMD), que

⁷ J.-L. Cunchillos, J.-P. Vita, *Concordancia de Palabras Ugaríticas en Morfología Desplegada. Banco de datos filológicos semíticos noroccidentales (BDFSN). Datos ugaríticos, II*, Madrid, CSIC, 1995, 3 vols, 2.889 páginas.

⁸ J.-L. Cunchillos, R. Cervigón, J.-P. Vita, J.-M. Galán y J.-Á. Zamora, *Generador de Segmentaciones, Restituciones y Concordancias (GSRC)*, CD-ROM, Madrid 1996. En Internet, J.-L. Cunchillos, J.-P. Vita, J.-M. Galán y J.-Á. Zamora, *Banco de Datos, GSRC - Internet. Febrero 1997*, en *Šapānu. Publicaciones en Internet del Laboratorio de Hermeneumática* (<http://www.labherm.filol.csic.es/>), Madrid 1997-...

⁹ Se divide por tanto el conjunto de CG en PMD y CGI. Las PMD sirven para completar las CGI.

cumplen las condiciones exigidas para restituir la cadena incompleta, las PMD se almacenan como Restituyentes de la CGI. La CGI pasa a ser Restituible o Cadena Grafemática Restituible (CGR).

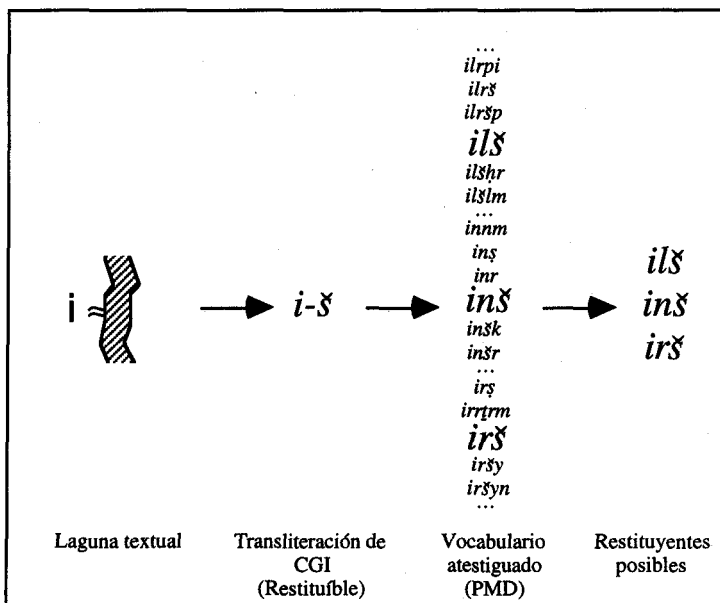


Fig. 4: Esquema simplificado de funcionamiento de *Restituciones*

Restituciones busca encajar una o más Palabras en Morfología Desplegada en el patrón dado por el epigrafista. Si el epigrafista connota de forma general una laguna, sin indicar su longitud (p. ej.: *ml—*), *Restituciones* actúa con el mayor margen posible. Busca palabras completas que posean los grafemas legibles en la posición y lugar adecuados, más cualquier número de grafemas en el lugar de la laguna (desde ninguno al mayor número existente).

En el caso del ejemplo *ml—*, GSRC propone 58 Restituyentes:

<i>ml</i>	<i>mlu</i>	<i>mlhś</i>	<i>mlkn</i>	<i>ml'tn</i>
<i>m^la</i>	<i>mlun</i>	<i>mly</i>	<i>mlkn'm</i>	<i>mlgt</i>
<i>mlak</i>	<i>mlbr</i>	<i>mlk</i>	<i>mlkrpi</i>	<i>mlš</i>
<i>mlakk</i>	<i>mlbš</i>	<i>mlki</i>	<i>mlkršp</i>	<i>mlrm</i>
<i>mlakm</i>	<i>mlbšh</i>	<i>mlkbn</i>	<i>mlkt</i>	<i>mlt</i>
<i>mlakt</i>	<i>mlghy</i>	<i>mlkh</i>	<i>mlktn</i>	<i>mlth</i>
<i>mlakth</i>	<i>mld</i>	<i>mlky</i>	<i>ml</i>	<i>mlthm</i>
<i>mlakty</i>	<i>mldy</i>	<i>mlkyy</i>	<i>mlm</i>	<i>mlth</i>
<i>mlaktk</i>	<i>mlh</i>	<i>mlkym</i>	<i>mln</i>	<i>mltn</i>
<i>mlat</i>	<i>mlhmy</i>	<i>mlkytn</i>	<i>mls</i>	<i>mltn</i>
<i>mlⁱ</i>	<i>mlhmt</i>	<i>mlkk</i>	<i>mlsm</i>	
<i>mlit</i>	<i>mlht</i>	<i>mlkm</i>	<i>ml'n</i>	

Si el epigrafista connota con exactitud la falta de algún signo mediante uno o varios guiones (p. ej. *ml-t*, *-lk-*, etc.), cada guión se restituye por un único grafema. Es el primer nivel de restitución y el considerado de probabilidad más alta con este método.

En el caso del ejemplo *ml-t*, GSRC propone 5 Restituyentes:

mlat
mlit
mlht
mlkt
mlgt

Puede darse el caso de que, procediendo así, no se obtengan resultados. En un segundo nivel de restitución, la programación busca Restituyentes dando un margen de un grafema a lo connotado por el epigrafista. Cada guión puede así ser completado por dos grafemas, o por ninguno. Por ejemplo, una laguna de dos guiones puede ser restituida por uno o tres grafemas. Las soluciones se proporcionan siempre después de las soluciones del primer nivel de restitución, y se consideran de menor probabilidad.

En el caso del ejemplo anterior *ml-t*, para el que GSRC proponía 5 Restituyentes en el primer nivel, las soluciones de segundo nivel son 3:

mlakt
mlhmt
mlt

Como se aprecia en los dos primeros Restituyentes, se ha considerado la posibilidad de que entre *l* y *t* hubiera espacio suficiente para dos grafemas. Como se aprecia en el tercer Restituyente, *mlt*, se ha considerado también la posibilidad de que el espacio disponible no diera cabida a signo alguno. Son soluciones con un margen de un signo sobre lo connotado por el epigrafista (*ml-t*, un único grafema perdido).

Si tampoco de este modo se consiguen Restituyentes, la programación otorga mayor margen en el número de grafemas de la laguna. Los resultados son consi-

derados como de baja probabilidad, pero arrojan a veces cadenas a considerar. Si al final de este proceso todavía no se obtienen resultados válidos, el usuario puede considerar divisiones de la cadena y usar *Restituciones* para comprobar sus hipótesis. Por ejemplo, una cadena *ml--rt* puede, después de todo, ocultar dos palabras en morfología desplegada. Puede dividirse en *ml-* y *--rt*, ambas con múltiples restituyentes. Entre ellos, *mlk* y *ugrt*, restituyentes ambos muy probables en combinación.

Todas las opciones de restitución están basadas en el vocabulario existente, y se proporcionan al especialista como una ayuda. El especialista tiene siempre la última palabra a la hora de aceptar decisiones o elegir entre las diversas propuestas de la programación. Puede tener en cuenta el contexto u otros elementos, como el análisis morfológico de los Restituyentes, que se le proporcionará con el tercer módulo.

RESULTADOS: CPU Y GSRC

Todo este trabajo dio lugar a dos publicaciones principales: la *Concordancia de Palabras Ugaríticas* (CPU¹⁰, en soporte papel) y el *Gestor de Segmentaciones, Restituciones y Concordancias* (GSRC¹¹, en soporte electrónico).

La *Concordancia de Palabras Ugaríticas* no atañe directamente a *Restituciones*, sino a *Concordancias*. Pero, al ser *Concordancias* un procedimiento posterior a *Segmentaciones* y *Restituciones*, actúa ya sobre los resultados de ambas aplicaciones. Con los resultados de *Concordancias* se elaboró CPU. Presentaba la novedad de catalogar no sólo las Palabras en Morfología Desplegada, PMD, sino también las Cadenas Grafemáticas Incompletas, CGI. Normalmente los fragmentos de palabras no se contemplan al elaborar una concordancia. Junto a 6521 PMD, la CPU proporcionaba 1981 Cadenas Grafemáticas Incompletas. Y relacionaba unas con otras. Cada Palabra en Morfología Desplegada aparecía con posibles Restituibles, obtenidas por el procedimiento de *Restituciones*. De estas cadenas incompletas (CGI), 1428 resultaron Cadenas Grafemáticas Restituibles (CGR). Son todas aquellas cadenas para las que existen una o más restituciones posibles. Los signos aislados en contexto fragmentario tienen un gran número de Restituyentes posibles, por lo que en la publicación se separaban del resto por su menor utilidad. Eran las 380 CGRU (Cadenas Grafemáticas Restituibles Unilíteras), recogidas también al final de la CPU. La aplicación *Restituciones* dejaba un pequeño grupo de cadenas, 173, sin Restituibles. Las cadenas para las que no se obtienen Restituyentes se denominan CGSR (Cadenas Grafemáticas Sin Restitución). En total, los datos sobre los que se elaboró la CPU, eran los siguientes:

CG	= PMD	+ CGI		
8502	6521	1981		
		CGR	CGRU	CGSR
		1428	380	173

¹⁰ Citado en nota 7.

¹¹ Citado en nota 8.

De un total de 8502 Cadenas Grafemáticas, 1981 se hallan incompletas. Este material supone un 23'3%, es decir, casi una cuarta parte del volumen total de cadenas grafemáticas atestiguadas. El material fragmentario era antes desestimado total o parcialmente. Ahora, las 173 cadenas sin restitución posible representan apenas el 2% del total.

Los Restituyentes proporcionados por la CPU eran el resultado de aplicar sistemáticamente *Restituciones* al conjunto de las Cadenas Grafemáticas Incompletas, CGI. Es decir, CPU refleja una parte de los datos del BDFSN tal como se encontraban en el momento de publicarse CPU. La programación de *Introducción de textos* del BDFSN conecta con *Restituciones* cada vez que se introduce en el banco un nuevo texto. Al introducir nuevas tablillas la programación propone automáticamente soluciones a las nuevas cadenas incompletas. Si las tablillas contienen nuevo vocabulario, *Segmentaciones* incorporará las nuevas cadenas al vocabulario atestiguado y *Restituciones* se basará en ellas para nuevas soluciones. Todas quedan almacenadas. La información está permanentemente actualizada.

Actualmente el BDFSN Ugarítico recoge 30.952 datos relativos a restituciones. Si proporcionáramos aquí la lista de los Restituibles con sus Restituyentes y la lista de Restituyentes con sus Restituibles harían falta más de 200 páginas dispuestas a tres columnas. Tal hecho da idea del volumen de información generada y de la dificultad de manejarla o publicarla a la manera tradicional. Además, la posibilidad de introducir nuevos datos supone, en un sistema interrelacionado, el enriquecimiento de todos los niveles de información. Con ello, envejecen con rapidez las publicaciones en soporte papel. CPU continúa siendo la única publicación hoy existente en la que se proporciona el tipo de información de la que hablamos, y los datos anteriores son en esencia todavía válidos. Pero la forma más útil de poner a disposición de los especialistas las herramientas informáticas son los soportes electrónicos. Los sistemas de almacenamiento magnético, el CD-ROM, Internet, han sido formas elegidas para publicar los resultados de esta y otras realizaciones del proyecto, y lo seguirán siendo en el futuro¹². Tales soportes proporcionan un instrumento activo y abierto. Por ello fue publicado el *GSRC* (*Generador de Segmentaciones, Restituciones y Concordancias*), primero en soporte CD-ROM y después en Internet¹³.

El *Generador de Segmentaciones, Restituciones y Concordancias* incluye, como su propio nombre indica, la herramienta *Restituciones*, junto con el resultado de su aplicación sistemática al conjunto de los textos. Tener almacenadas las relaciones entre Restituibles y Restituyentes permite al usuario consultar con rapidez los posibles Restituyentes de determinadas CGI. En el ejemplo anterior *ml*— hemos aprovechado esta circunstancia: *ml*— es una cadena fragmentaria realmente atestiguada.

El *GSRC* es especialmente útil para el epigrafista que lee por primera vez una tablilla. Al incluir también la programación generadora, el *GSRC* permite al usuario

¹² Durante el año 1998 se ha publicado la nueva actualización en Internet del *GSRC* y próximamente se publicarán la versión del *GSRC* para Windows en CD-ROM y los módulos 1 y 2 del BDFSN fenicio (denominado *Melqart, Sistema Gestor de Datos Fenicios y Púnicos*). Gran parte de su funcionamiento interno será explicado mediante artículos como el presente.

¹³ Citados en n. 8.

introducir su propia CGI, incluso si no está atestiguada en los textos, y dejar que *Restituciones* genere posibles soluciones para ella. En el caso del ejemplo *ml-t* hemos hecho uso de esta opción y buscado una cadena, *ml-t*, no atestiguada en los textos ugaríticos. La programación lo advierte al usuario y le pregunta si desea, a pesar de ello, buscar Restituyentes a la cadena.

BÚSQUEDA DE PALABRAS A TRAVÉS DE GRAFEMAS

Poder introducir cualquier cadena para restitución, aunque no esté atestiguada, tiene aplicaciones secundarias de gran utilidad. No sólo da gran libertad al usuario a la hora de lanzar hipótesis sobre textos y cadenas. También permite pedir, por ejemplo, Restituyentes de cadenas que comiencen por determinados caracteres, uno o varios. Se pueden pedir restituciones de cadenas como *m—*, *ml—* o *mlk—*. El resultado es, en la práctica, el hallazgo de todas las palabras atestiguadas en el *corpus* ugarítico que empiezan por *m*, *ml* o *mlk*. Para este último caso, GSRC encuentra 16 soluciones. Es decir, 16 Restituyentes de *mlk—* o, lo que es lo mismo, 16 palabras ugaríticas que comienzan por *mlk*:

<i>mlk</i>	<i>mlky</i>	<i>mlkk</i>	<i>mlkrpi</i>
<i>mlki</i>	<i>mlkyy</i>	<i>mlkm</i>	<i>mlkršp</i>
<i>mlkbn</i>	<i>mlkym</i>	<i>mlkn</i>	<i>mlkt</i>
<i>mlkh</i>	<i>mlkytn</i>	<i>mlkn‘m</i>	<i>mlktn</i>

Lo mismo puede hacerse con cadenas que terminen en *k*, o en *lk* (*—k*, *—lk*), o que contengan dichos grafemas (*—k—*, *—lk—*). O, incluso, que los dispongan de una determinada manera en su interior. Es decir, se posee un instrumento de búsqueda de cadenas a través de grafemas. Por ejemplo: los Restituyentes de *-l-t*, son todas las PMD cuyo segundo grafema es *l* y el cuarto *t*. Las soluciones a este último ejemplo son 42:

<i>alit</i>	<i>hlbt</i>	<i>lit</i>	<i>mlat</i>	<i>pltt</i>
<i>altt</i>	<i>hlmt</i>	<i>lbt</i>	<i>mlit</i>	<i>šl‘t</i>
<i>ilgt</i>	<i>hlqt</i>	<i>lgt</i>	<i>mlht</i>	<i>qlht</i>
<i>ilht</i>	<i>zlm̄t</i>	<i>lht</i>	<i>mlkt</i>	<i>šlht</i>
<i>glbt</i>	<i>kl̄at</i>	<i>lh̄t</i>	<i>ml̄gt</i>	<i>šl̄mt</i>
<i>glyt</i>	<i>kl̄bt</i>	<i>lyt</i>	<i>‘lyt</i>	<i>šl̄št</i>
<i>dlht</i>		<i>lkt</i>	<i>‘lmt</i>	<i>tlht</i>
<i>hlkt</i>		<i>llit</i>	<i>gl̄mt</i>	<i>tl̄t</i>
		<i>llt</i>		<i>tl̄t</i>
		<i>lmt</i>		
		<i>lqt</i>		

En la tercera columna aparecen Restituyentes de *-l-t* aparentemente erróneos. El grafema *l* está en primer lugar en la cadena, y no en segundo. Este tipo de soluciones se produce por otra opción contemplada por el procedimiento *Restituciones*. Se considera la posibilidad de que el carácter connotado por el guión (*-*) sea un

separador o un espacio en blanco, sin grafema. *GSRC* actúa así siempre que con ello no esté proponiendo una división. Es decir, siempre que el grafema connotado por guión no sea intermedio en la cadena. Por ello, los grafemas connotados mediante guión al inicio o final de palabra se desechan también para hallar algunos Restituyentes posibles más. En este caso, los agrupados en la tercera columna.

En un segundo nivel de probabilidad (con margen de un grafema para la laguna connotada), *Restituciones* propone 51 restituciones de *-l-t*. En un último nivel, con mayor margen, propone 33 más. Estos dos niveles de probabilidad, al igual que la consideración especial de los grafemas iniciales o finales, son útiles al buscar Restituyentes alternativos. Para usar *Restituciones* como instrumento de búsqueda de cadenas a través de grafemas, son más útiles los resultados habituales del primer nivel.

Con ello *Restituciones* viene a aumentar la potencia de gestión de la Base de Datos, ampliando las posibilidades de búsqueda. Algo muy útil cuando se investigan problemas específicos de la documentación (usos de determinados grafemas, presencia de determinadas consonantes, combinaciones de las mismas, presencia de morfemas, prefijos, terminaciones, etc.).

POSIBLES OCURRENCIAS DE PALABRAS EN CONTEXTOS FRAGMENTARIOS

El *GSRC* dispone también de la opción de hacer funcionar *Restituciones* a la inversa. No solamente busca los Restituyentes de cadenas incompletas (CGI). Puede buscar también los Restituibles de una cadena completa, de una PMD. Por ejemplo, el usuario puede pedir a la herramienta *Restituciones* que encuentre los Restituibles de la cadena *inš*. La cadena no está incompleta, se trata de una PMD. *Restituciones* encuentra siete Cadenas Grafemáticas Incompletas que pueden ser restituidas por *inš*:

—*nš*
 —*nš*—
 -*nš*—
i-š
in—
in-----
in-----

Estas cadenas incompletas son los Restituibles del Restituyente *inš*. Lo que equivale a decir que los lugares donde aparecen tales cadenas son posibles ocurrencias de *inš*. Con las búsquedas de posibles ocurrencias, se completa la concordancia de la palabra. *Concordancia* puede hallar los lugares exactos donde aparece *inš* y los lugares donde podría aparecer, generando la concordancia de cada Restituible. Los estudios sistemáticos de palabras ugaríticas disponen así, con la combinación de varias herramientas del *GSRC*, de una forma de trabajo rápida, exhaustiva y novedosa.

A la interrelación de datos y aplicaciones responde también la programación de gestión y presentación de datos del *GSRC*. Está diseñada para navegar constantemente de unas peticiones a otras. Las diferentes presentaciones de resultados

permiten volver sobre datos anteriores o requerir nueva información sobre la ya hallada. A modo de ejemplo, en el caso antes citado, al solicitar Restituibles de la PMD *inš* se nos proporcionan siete CGI como respuesta. Podemos requerir más información sobre una de ellas, por ejemplo *i-š*. Entre las diversas opciones que se nos ofrecen para obtener más información sobre *i-š*, se encuentra nuevamente la búsqueda de Restituyentes mediante *Restituciones*. GSRC halla entonces 3 restitu-yentes posibles: *ilš*, *inš*, *irš*. Es el ejemplo propuesto en la figura 4, con el que ilustrá-bamos el proceder de *Restituciones*.

Con las soluciones se nos proporciona información sobre el nivel de probabilidad de la restitución y el número exacto de Restituyentes obtenidos. El número de alter-nativas ya es significativo. A veces, un gran número de Restituyentes proporciona un número de posibilidades demasiado elevado para resultar útil. A la inversa, un número demasiado reducido indica la dificultad para encontrar vocabulario adecuado. Pero en ambos casos la información suministrada permite llegar más lejos. Las respuestas se despliegan en diferentes ventanas y menús. Podemos consultar el contexto de cada ocurrencia, incluso el texto entero en el que se enclavan, o continuar buscando información sobre las palabras adyacentes, otras ocurrencias, etc.

EL DOCUMENTO ORIGINAL Y EL TRATAMIENTO DIGITAL DE IMÁGENES

Todos los tipos de restitución hasta ahora descritos se basan en la información con-notada por el epigrafista en la transcripción o transliteración del documento. Por lo tanto, todos los Restituyentes propuestos por la programación del BDFSN son posibili-dades pendientes de confirmación sobre el documento. Es decir, además de filológic-amente posibles, los Restituyentes deben ser epigráficamente posibles. Deben enca-jar físicamente en la laguna del documento original. En última instancia, toda restitui-ción debe ajustarse a las características materiales del original.

El tratamiento informático de las imágenes del documento abre nuevas posibilida-des, no desarrolladas en el GSRC. El epigrafista descarta aquellas restituciones que estima físicamente imposibles en el espacio disponible. Esta estimación se basa en el tamaño de los signos del documento, pero no está exenta de cierta subjetividad. Con la imagen digital pueden compararse grafemas del mismo documento, o del mismo es-criba, con el espacio exacto mostrado por la fotografía. Se puede llegar a copiar los signos de la restitución propuesta en el lugar dañado. Se obtiene una reconstrucción visual aproximada, pero más exacta que la simple apreciación personal.

DECISIONES DEL EXPERTO

Finalmente, el especialista es quien decide. La programación puede orientarle, se-gún la probabilidad asignada a la restitución. Además de los niveles descritos, la pro-gramación podría asignar diferentes niveles de probabilidad según otros criterios. Por ejemplo, si los Restituyentes están ya presentes en el mismo documento¹⁴, o en docu-

¹⁴ Habitualmente la presencia de una palabra en un documento aumenta la probabilidad de que tal palabra sea Restituyente en otra parte del mismo documento. Sólo en algunos casos, como

mentos del mismo género literario (o que presentan un mismo campo semántico), etc. Si fueran conocidos el escriba o la datación, se podría otorgar también mayor probabilidad a los Restituyentes con ocurrencias en documentos de un mismo escriba o cercanos en el tiempo. Pero la programación deja al experto las decisiones finales sobre el valor de los Restituyentes propuestos en cada contexto.

La decisión tomada por el especialista podrá almacenarse, ligando el Restituyente elegido a la ocurrencia de la CGI restituida. Puede decidirse, también, desechar algunas posibilidades, o jerarquizar otras según se adecúen más o menos al texto estudiado. Las decisiones deben ser siempre reversibles. Conforme el proceso de análisis avanza, el estudio morfológico, sintáctico o semántico podría refutar o confirmar las restituciones. En ese momento, el investigador podrá volver sobre sus antiguas decisiones y alterarlas de acuerdo con la nueva información disponible.

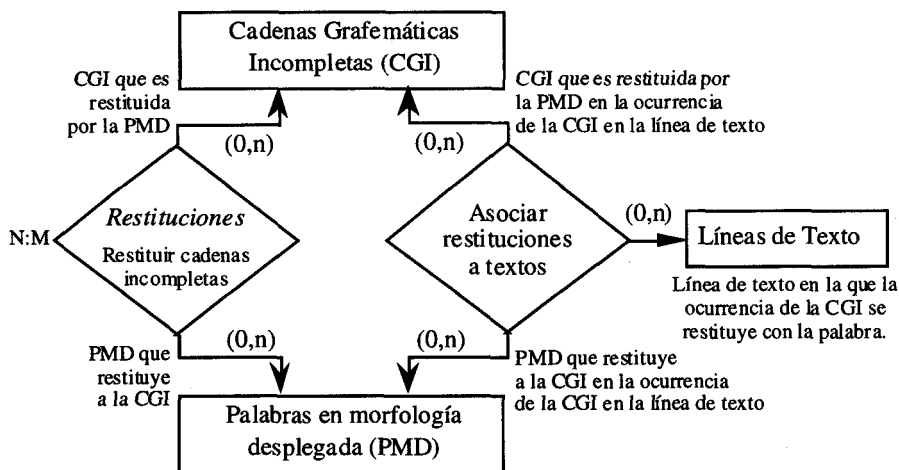


Fig. 5: Diagrama Entidad / Relación de la restitución de Cadenas Grafemáticas Incompletas

determinados listados (por ejemplo, de topónimos), el hecho se invierte, pues lo ya citado rara vez se repite.