

PRIMERA PARTE

UNAM/JURE EN EL INSTITUTO DE INVESTIGACIONES JURÍDICAS

A. Características de la información legislativa . . . . .	19
1. La información legislativa en general . . . . .	20
2. La información legislativa en México . . . . .	22
a. La legislación federal . . . . .	23
i) La frecuencia y cantidad de la producción legislativa . . . . .	23
ii) La vigencia de la legislación federal . . . . .	28
a') La reforma parcial . . . . .	28
b') La abrogación . . . . .	28
iii) La selección de los textos legislativos . . . . .	29
b. La legislación estatal . . . . .	29
i) La temporalidad . . . . .	29
ii) La vigencia de la legislación estatal . . . . .	30
iii) La selección de los textos legislativos . . . . .	31
B. Métodos de análisis de la información legislativa . . . . .	32
1. Los métodos "tradicionales" . . . . .	33
a. La <i>indexación</i> . . . . .	33
b. El <i>full-text</i> . . . . .	34
i) Ventajas . . . . .	36
ii) Inconvenientes . . . . .	36
2. El <i>abstract</i> legislativo . . . . .	38
a. Consideraciones generales . . . . .	38
b. Estructura y función de la ficha de análisis . . . . .	39
i) El <i>abstract</i> . . . . .	39
a') El párrafo y la frase . . . . .	40
b') Los niveles de interrogación . . . . .	41
ii) La zona de campos fijos y el resumen . . . . .	42
c. Los criterios de análisis . . . . .	45
d. Problemas de forma y de fondo . . . . .	49
i) De forma . . . . .	49
a') Redacción . . . . .	49
b') Signos de puntuación . . . . .	49
ii) De fondo . . . . .	50
a') Adaptación del <i>abstract</i> a la legislación . . . . .	50
b') Adaptación de la legislación al <i>abstract</i> . . . . .	50
C. Los instrumentos lingüísticos . . . . .	52
1. Los problemas lingüísticos y su solución . . . . .	52
a. Generalidades (sintaxis y morfología) . . . . .	52
b. Semánticos . . . . .	53
i) La sinonimia . . . . .	53
ii) La analogía . . . . .	53
iii) La polisemia . . . . .	54
2. El léxico . . . . .	54
3. El <i>thesaurus</i> . . . . .	60

## PRIMERA PARTE

# UNAM-JURE EN EL INSTITUTO DE INVESTIGACIONES JURÍDICAS

SUMARIO: A. *Características de la información legislativa*. 1. *La información legislativa en general*. 2. *La información legislativa en México*. a. *La legislación federal*. i) *La frecuencia y cantidad de la producción legislativa*. ii) *La vigencia de la legislación federal*. a') *La reforma parcial*. b') *La abrogación*. iii) *La selección de los textos legislativos*. b. *La legislación estatal*. i) *La temporalidad*. ii) *La vigencia de la legislación estatal*. iii) *La selección de los textos legislativos*. B. *Métodos de análisis de la información legislativa*. 1. *Los métodos "tradicionales"*. a. *La indexación*. b. *El full-text*. i) *Ventajas*. ii) *Inconvenientes*. 2: *El abstract legislativo*. a. *Consideraciones generales*. b. *Estructura y función de la ficha de análisis*. i) *El abstract*. a') *El párrafo y la frase*. b') *Los niveles de interrogación*. ii) *La zona de campos fijos y el resumen*. c. *Los criterios de análisis*. d. *Problemas de forma y de fondo*. i) *De forma*. a') *Redacción*. b') *Signos de puntuación*. ii) *De fondo*. a') *Adaptación del abstract a la legislación*. b') *Adaptación de la legislación al abstract*. C. *Los instrumentos lingüísticos*. 1. *Los problemas lingüísticos y su solución*. a. *Gramaticales (sintaxis y morfología)*. b. *Semánticos*. i) *La sinonimia*. ii) *La analogía*. iii) *La polisemia*. 2. *El léxico*. 3. *El thesaurus*. *Anexos*. *Bibliografía*.

### A. CARACTERÍSTICAS DE LA INFORMACIÓN LEGISLATIVA

Actualmente la mayoría de las experiencias en materia de informática jurídica se ubican en el campo de la jurisprudencia.<sup>1</sup> El nuestro es uno de los primeros proyectos que adopta el método del *abstract* para el análisis de la información legislativa.

Las características de la información legislativa se pueden apreciar en dos niveles: general y particular (el caso de México). Nos referiremos en un primer apartado a las características generales de la información legislativa, para enseguida tratar los rasgos específicos de la información legislativa en México.

<sup>1</sup> JURIDOC, SYDONI, LEXIS e ITALGIURE, por citar algunos.

## 1. La información legislativa en general

Se pueden apreciar las características generales de la información legislativa a través de su comparación con la jurisprudencia.

En el caso mexicano la obligatoriedad de la jurisprudencia se desprende de la Constitución política federal y de la Ley de amparo (Reglamentaria de los artículos 103 y 107 constitucionales). Sin entrar en detalles, podemos afirmar que el artículo 94 de la Constitución política federal y los artículos 192 a 197 de la Ley de Amparo establecen que la jurisprudencia de los tribunales federales superiores será obligatoria para los inferiores. Además de esto, para poder hablar de jurisprudencia obligatoria es necesario que existan cinco ejecutorias en el mismo sentido, sin existir una que interrumpa la acumulación de las anteriores con sentido diverso. La legislación, por otra parte, deriva su obligatoriedad de la observancia de un procedimiento legislativo y de su publicación en el *Diario Oficial* correspondiente. Las tesis relacionadas o ejecutorias aisladas de los tribunales sólo tienen efectos vinculatorios, de acuerdo con la Ley de Amparo, para las personas que fueron parte en el juicio de que se trate. La jurisprudencia, como hemos señalado, será obligatoria en los términos de los preceptos legales citados. A pesar de lo anterior, en la práctica la propia Suprema Corte ha sostenido el criterio, en diversos precedentes, que las tesis relacionadas o ejecutorias aisladas sí vinculan a los tribunales inferiores. El ámbito de aplicación personal de la legislación depende, una vez más, del documento base. Las leyes federales son obligatorias en toda la república, las estatales sólo dentro del estado del que emanan; existen circulares destinadas a ciertos funcionarios para su cumplimiento; la observancia de algunos reglamentos particulares se exige exclusivamente a miembros de una dependencia de gobierno; y así varios casos donde la obligatoriedad de la legislación se restringe o se amplía.

Ahora vamos a referirnos al problema de la temporalidad de la jurisprudencia y de la legislación. Mientras la jurisprudencia no está sujeta a reformas (aunque sí a cambios y modificaciones a cuyo análisis no entraremos), es éste, precisamente, el instrumento propio del cambio legislativo. Tanto las leyes como los reglamentos y algunos otros tipos de documentos legislativos (decretos o acuerdos) están sujetos a reformas que al publicarse en el *Diario Oficial* pasan a formar parte del texto original y que, por tanto, es necesario conservar y recuperar como tal. A este problema de temporalidad haremos referencia con más detenimiento en el próximo apartado.

Por otra parte, tratándose de la jurisprudencia mexicana, no siempre se han publicado las ejecutorias mismas de los ministros, sino resúmenes o tesis elaborados por empleados del *Semanario Judicial de la Federación*, órgano encargado de su publicación. Esto provocó que en ocasiones se

cambiara el sentido o se omitieran elementos importantes de la ejecutoria que les da origen. Por reformas a la Ley de Amparo publicadas en el *D. O.* de 16-I-84, los ministros están obligados a deliberar sobre el contenido de las tesis que se publican. Si a este problema agregamos el de la extemporaneidad en la publicación de las tesis y el de su gran extensión, podremos comprender las dificultades para conocer la jurisprudencia aplicable en un momento determinado.

Podemos afirmar que, tanto en la forma como en el fondo, las disposiciones que produce un juez y un legislador son muy distintas. En primer lugar debemos considerar que, mientras las leyes o reglamentos contienen invariablemente títulos, libros, capítulos, secciones u otras formas de división, que para efectos de análisis y recuperación es importante conservar, las tesis de jurisprudencia son un resumen del aspecto jurídico de una controversia, tomado a partir de los considerandos de la sentencia que la resolvió. (Como se sabe, las sentencias judiciales mismas se estructuran de la siguiente manera: Proemio, resultandos, considerandos y puntos resolutorios.) Por otro lado, la estructura de las decisiones de justicia es constante, el juez considera siempre los elementos de hecho, los analiza a la luz de las disposiciones jurídicas aplicables y produce una resolución. Esta estructura constante no existe en las disposiciones legislativas debido a la diversidad de disposiciones que puede generar el poder legislativo. Así, encontramos que existen leyes, reglamentos, decretos, acuerdos, circulares, oficios, etc., y todos ellos en mayor o menor medida gozan de fuerza obligatoria. En este orden de ideas resulta difícil encontrar una estructura constante en la construcción legislativa. En relación con la extensión de los textos, no son menos evidentes las diferencias, pues las ejecutorias, en promedio, nunca rebasan un máximo y la extensión de los textos legislativos depende en gran parte de la naturaleza del documento en cuestión.

En relación con el fondo, encontramos que las diferencias son aún más graves. El lenguaje del legislador prescinde de los elementos de hecho a que tiene que referirse un juez y, por definición es más abstracto. Esto significa que, si bien ambos tipos de documento contienen consideraciones de derecho, la aparición de cuestiones fácticas es propia de la jurisprudencia.

Si consideramos que la labor del juez es primordialmente interpretativa, entenderemos por qué los conceptos jurídicos son más explícitos en la jurisprudencia que en la legislación. Tomemos un ejemplo típico: el artículo 1910 del Código Civil para el Distrito Federal contiene el concepto de responsabilidad civil y dice lo siguiente: "El que obrando ilícitamente o contra las buenas costumbres cause daño a otro, está obligado a repararlo, a menos que demuestre que el daño se produjo como consecuencia de culpa o negligencia inexcusable de la víctima". Si revisamos

cuidadosamente el texto del Código Civil, observamos que de él no se desprende expresamente el concepto de la responsabilidad civil. Sin embargo, el juez que califica la conducta de una persona como ilícita o contra las buenas costumbres y que establece el nexo causal entre tal conducta y el daño causado a otra persona sin que medie culpa o negligencia de ésta, necesariamente tendrá que fincar la responsabilidad civil del sujeto haciendo referencia expresa al citado concepto.

Admitido el carácter interpretativo de la jurisprudencia, no es difícil establecer la diferencia primordial entre ésta y la legislación. A las abstracciones de ley se opone la concreción de la jurisprudencia. Mientras ésta se ocupa de un problema jurídico en particular, aquélla tiene un carácter heterogéneo. Esto es de gran importancia para efecto de la obligatoriedad de ambas.

## 2. *La información legislativa en México*

Dos son las características de nuestro sistema político constitucional que influyen en la información legislativa. En primer lugar, el régimen que para el gobierno de la república adoptan los artículos 40 y 115 de nuestra Constitución política federal, y segundo la fuerte predominancia del poder ejecutivo federal sobre el legislativo y el judicial, así como sobre los gobiernos de las entidades federativas.

La adopción de un sistema federal implica la existencia de dos niveles de competencia legislativa. A esta situación tenía que responder nuestro sistema de recuperación automática de información legislativa. De ahí la distinción entre la legislación federal y la estatal (aquí se publica la municipal) y la del Distrito Federal que orienta a UNAM-JURE.

Por otra parte, el predominio del poder ejecutivo federal en nuestro país produce una curiosa coincidencia entre el lenguaje que emplea en disposiciones material y formalmente ejecutivas, y el utilizado por el legislador federal en instrumentos propios de su competencia (en virtud de que casi la totalidad de las leyes provienen de iniciativa del ejecutivo). Asimismo, el ejecutivo federal elabora numerosos convenios,<sup>2</sup> a los que las entidades federativas únicamente se adhieren, sin posibilidad alguna de negociar sus términos. De esta manera las competencias federal y estatal, a que hicimos referencia anteriormente, se diluyen en favor de la federación para las materias de estos convenios.

Debido a estas consideraciones se procederá al análisis de la legislación

<sup>2</sup> Convenio único de desarrollo (21 de junio de 1984), de colaboración administrativa en materia fiscal (10 de diciembre de 1983), de coordinación fiscal federal (21 de junio de 1984), de colaboración electoral (19 de octubre de 1982), para la impartición de justicia celebrados con la Procuraduría General de la República y las entidades federativas (13 de septiembre de 1984).

federal en primer término, para concluir con algunas observaciones sobre la legislación estatal. -

### a. La legislación federal

En materia de informática jurídica tres son los principales problemas que nos plantea la legislación federal: su frecuencia y cantidad de producción, su vigencia y los criterios a seguir para su selección.

#### i) La frecuencia y cantidad de la producción legislativa

La frecuencia y cantidad de la producción legislativa puede apreciarse en las siguientes gráficas (véase páginas 24-27).

De las gráficas 1 y 2 se desprende la importancia que en materia de producción legislativa tiene el cambio de sexenio. En efecto, se observa que el número de disposiciones relevantes aumenta considerablemente en 1982, y especialmente en los meses de noviembre (último mes del sexenio de López Portillo) y diciembre (primer mes del sexenio de Miguel de la Madrid).

Asimismo, el cambio del poder ejecutivo federal influye en el lenguaje empleado en los textos respectivos, esto se observa en el caso de los *fideicomisos*, institución promovida bajo el régimen de Luis Echeverría, figura que cambia durante el de López Portillo por la de *fondos*. Lo mismo sucede con conceptos como los de *planeación democrática*, *renovación moral de la sociedad*, *modernización administrativa*, empleados por el actual régimen.

Por otra parte, ciertas instituciones propias de un régimen, suelen desaparecer por disposición del presidente entrante. Los ejemplos son numerosos y sólo citamos los casos de FONAPAS (*Diario Oficial* del 31 de enero de 1983), INPI, IMAN e INJUVE.

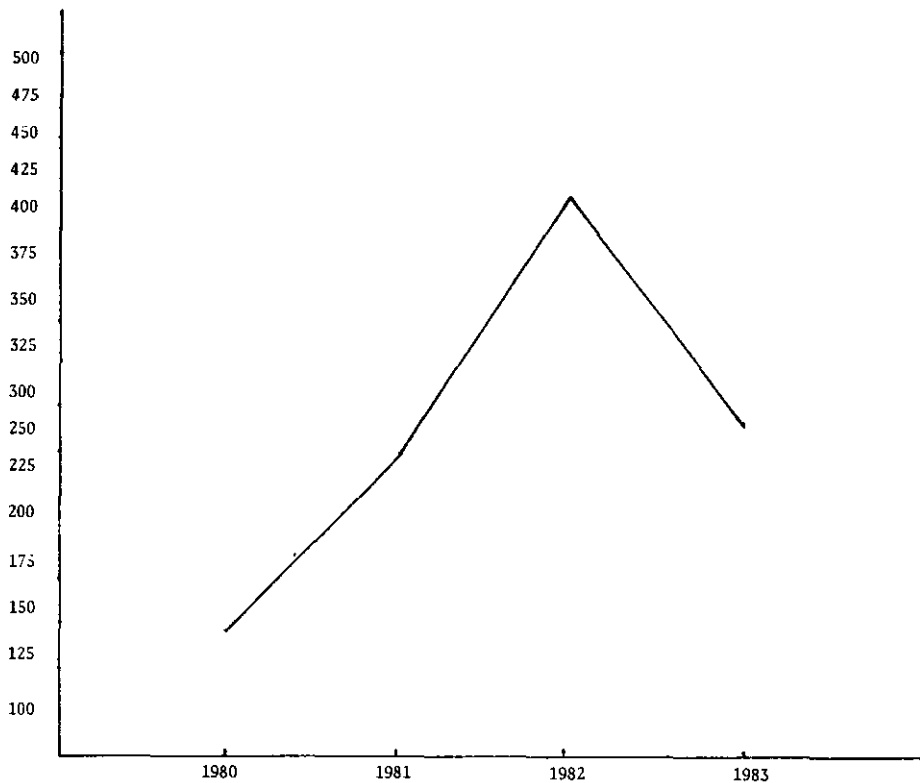
Las gráficas 3 y 4 nos muestran los periodos críticos de la tarea legislativa. Curiosamente no es durante el periodo ordinario de sesiones cuando aumenta la producción legislativa, sino que ésta alcanza sus puntos más altos en diciembre y enero. A esta situación se opone la baja actividad durante febrero, mayo y septiembre (apertura del periodo ordinario de sesiones).

Todo lo anterior exige una programación detallada de las distintas actividades de actualización del sistema UNAM-JURE. El análisis y la captura de los documentos que deben ingresar a la base de datos debe hacerse tomando en consideración los periodos críticos antes mencionados. Igualmente, el diseño y actualización de los instrumentos lingüísticos tienen que dar cuenta de los cambios en el lenguaje e instituciones, propios de nuestro sistema político.

**Gráfica 1.- Disposiciones federales relevantes publicadas en el D.O. de 1980-1983.**

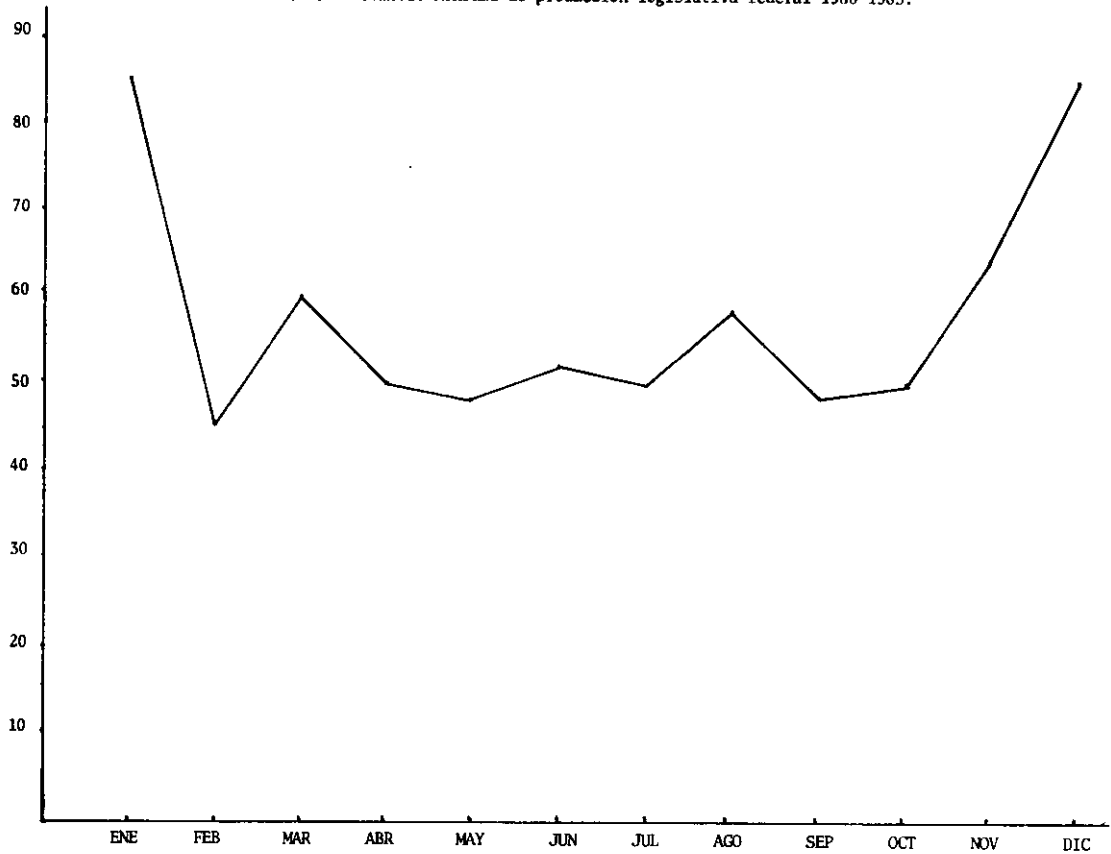


Gráfica 2.- Disposiciones federales relevantes publicadas en el D.O. durante el período de sesiones del Congreso 1980-1983.

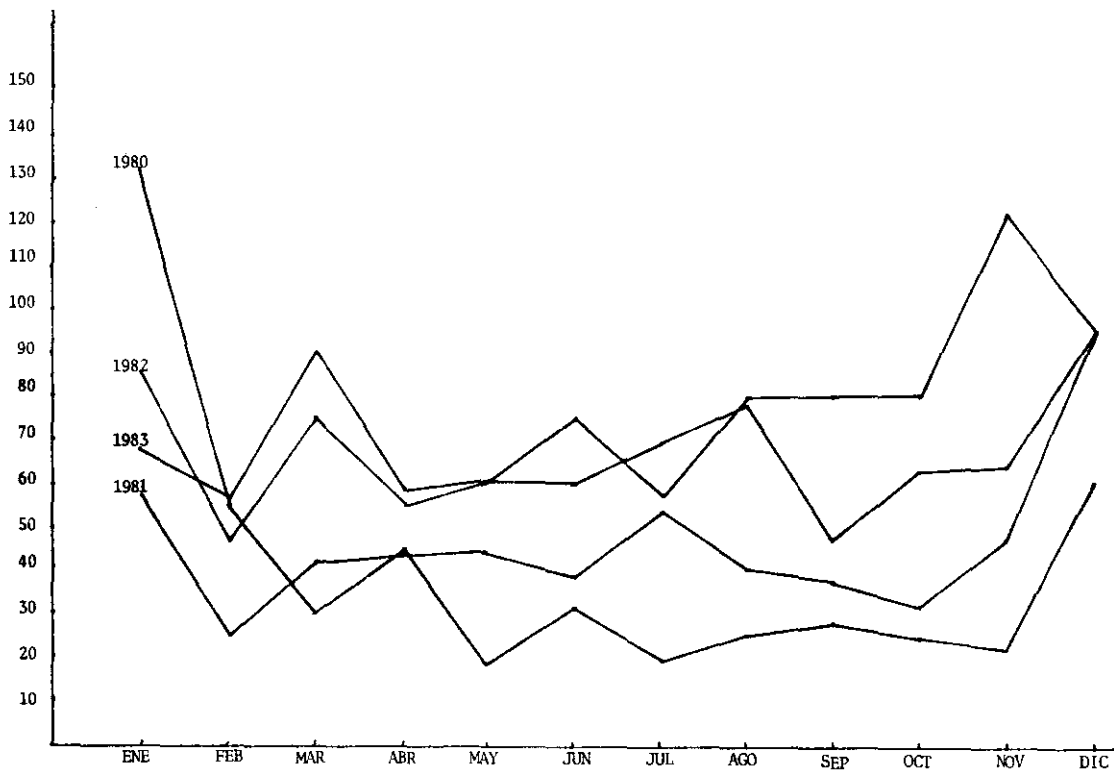




Gráfica 3.- Promedio mensual de producción legislativa federal 1980-1983.



Gráfica 4.- Producción legislativa anual 1980-1983.



## ii) La vigencia de la legislación federal

Como señalamos anteriormente, el instrumento propio del cambio legislativo es la *reforma*. Esta figura puede afectar parte del texto en cuestión, o afectarlo en su totalidad:

### a') La reforma parcial

Ésta puede consistir en una modificación-adición o derogación.

La *modificación* implica la sustitución de parte de un texto por otro, mientras que la *derogación* se encarga de la supresión de dicha parte, y a la *adición* le corresponde la inclusión de nuevas disposiciones al texto original.

Estas clases de reforma legislativa plantean un grave problema de vigencia, pues generalmente los usuarios de una base de datos se interesan por conocer los textos vigentes. Para lograr esto, UNAM-JURE debe ocuparse no solamente de las disposiciones originales, sino también de sus reformas.

Sin embargo, la detección del texto vigente puede resultar imposible, puesto que el legislador recurre con frecuencia a la derogación implícita,<sup>3</sup> utilizando para ello las disposiciones transitorias.

En este sentido, UNAM-JURE es muy flexible, pues permite la recuperación de todos los textos relacionados con los conceptos utilizados en la interrogación, desplegándolos cronológica y regresivamente sin prejuzgar sobre su vigencia. Corresponde al usuario el determinar, en función de las fechas de publicación, el texto vigente.

### b') La abrogación

Por medio de esta figura el legislador dispone la supresión total de un texto legislativo (una ley, por ejemplo), en ocasiones con el fin de expedir uno nuevo.

Con fundamento en el artículo noveno del Código civil para el Distrito Federal en materia del fuero común y para toda la república en materia del fuero federal, una ley podrá ser abrogada únicamente "...por otra posterior que así lo declare expresamente o que contenga disposiciones total o parcialmente incompatibles con la ley anterior". Este precepto hace referencia a dos tipos de abrogación: a la abrogación expresa de una ley anterior por una ley posterior (generalmente decretada en los artículos transitorios de ésta), y a la tácita: "...que resulta no de un texto legal

<sup>3</sup> "Se derogan todas las disposiciones que se opongan al presente ordenamiento".

expreso sino de la incompatibilidad total o parcial que existe entre los preceptos de (ambas leyes”).<sup>4</sup>

Una vez más afirmamos que nuestro sistema no puede ocuparse de la abrogación tácita, pues ello implica una labor interpretativa de fondo, ajena a los fines del análisis documental. Sin embargo, UNAM-JURE sí se ocupa del análisis de los artículos transitorios y, por lo tanto, acoge la abrogación expresa.

### iii) La selección de los textos legislativos

Los diarios y periódicos oficiales de la federación y las entidades federativas son el documento-fuente de nuestra base de datos. En estas publicaciones aparecen una gran diversidad de textos, no todos ellos jurídicamente relevantes para los usuarios de nuestro sistema.

Sin embargo, debido a la importancia de la legislación federal, el criterio para la selección de ésta es más amplio que tratándose de la producción legislativa estatal (que incluye la municipal).

A continuación enunciamos algunas de las disposiciones irrelevantes para efecto de la selección y que, consecuentemente, no son objeto de análisis: resoluciones agrarias, convocatorias para el otorgamiento de contratos con la administración pública, movimientos del personal diplomático acreditado en México, movimientos y licencias a notarios, creación de notarías, otorgamiento de premios y condecoraciones, declaratorias de nacionalización de templos, y avisos judiciales. Los efectos espaciales y personales, restringidos, y la temporalidad de este tipo de disposiciones motivan nuestra decisión.

## b. La legislación estatal

Al igual que en la legislación federal, la estatal nos plantea ciertas dificultades de temporalidad, vigencia y selección:

### i) La temporalidad

Además de la existencia de 32 órganos de información legislativa (Boletín Oficial, Gaceta de Gobierno, Periódico Oficial), se agregan su distribución deficiente, su diferente periodicidad y la carencia de técnica legislativa, dificultando, pues, su tratamiento automático.

<sup>4</sup> “Abrogación”, *Diccionario jurídico mexicano*, México, UNAM, 1982, tomo I, p. 31.

En cuanto a la administración de los periódicos oficiales estatales se puede afirmar, en primer lugar, que su envío a nuestro Centro de Información es muy irregular y en ocasiones incompleto. Algunas entidades federativas incurrir en retrasos considerables (incluso de años) que producen lagunas en nuestro acervo y nos impiden la pronta actualización de nuestra base de datos. Además, un considerable número de estados publican los textos más importantes de su legislatura, en anexos o suplementos que ocasionalmente recibimos.

La diferente periodicidad de aparición de los órganos informativos oficiales de las entidades federativas,<sup>5</sup> nos impide programar adecuadamente el análisis de la legislación estatal. Asimismo, este problema dificulta la tarea de inventario de nuestro acervo.

Más grave resulta la falta de técnica legislativa. Frecuentemente esta carencia incide sobre la denominación o calificación de los diferentes actos legislativos, pues se llega a confundir reglamentos, decretos, acuerdos y leyes. Otras veces se reforman ciertas disposiciones sin publicarse el texto íntegro del artículo reformado y sin hacer referencia al documento-fuente, ni a su fecha de publicación. En otros casos se llega a dar vigencia a decretos, acuerdos, e incluso leyes, con fecha anterior a la de su publicación en el periódico correspondiente.

UNAM-JURE tiene que tomar en consideración todas estas dificultades y adaptarse a las características particulares de la información legislativa estatal, si pretende ofrecer un servicio eficiente. En el apartado relativo al análisis de textos legislativos y a las características del *abstract* (véase B. 2) abordaremos la estrategia adoptada, para hacer frente a las dificultades apuntadas.

## ii) La vigencia de la legislación estatal

Lo dicho sobre los problemas de vigencia en la legislación federal, para efectos de un sistema de recuperación automática de información legislativa, vale para la legislación estatal. Sin embargo, la competencia legislativa local presenta otras dificultades.

<sup>5</sup> Aguascalientes: los domingos. Baja California: los 10, 20, 30 del mes. Baja California Sur: *idem*. Campeche: martes, jueves y sábados. Coahuila: miércoles y sábados. Colima: sábados. Chiapas: miércoles. Chihuahua: miércoles y sábado. Distrito Federal: 1 y 15 del mes. Durango: jueves y domingo. Guanajuato: jueves y domingo. Guerrero: miércoles. Hidalgo: 1, 8, 16, 24 del mes. Jalisco: martes, jueves, sábado. México: *idem*. Michoacán: lunes y jueves. Morelos: miércoles. Nayarit: miércoles y sábado. Nuevo León: *idem*. Oaxaca: sábado. Puebla: martes y jueves. Querétaro: jueves. Quintana Roo: cada semana. San Luis Potosí: domingos y jueves. Sinaloa: lunes, miércoles y viernes. Sonora: miércoles y sábado. Tabasco: *idem*. Tamaulipas: *idem*. Tlaxcala: miércoles. Veracruz: lunes, martes y sábado. Yucatán: de lunes a viernes. Zacatecas: miércoles y sábado.

La existencia de facultades concurrentes de la federación y de las entidades federativas, tienen como consecuencia la duplicidad de la publicación de las disposiciones relativas en los periódicos oficiales federal y de la entidad correspondiente. Sin embargo en algunas ocasiones tiene efectos jurídicos su publicación en el órgano estatal, pues, en general aquélla tiene un mero propósito informativo, de difusión de las disposiciones.

La publicación a nivel estatal tiene efectos primordialmente en dos casos. Nos referimos a la gran diversidad de convenios celebrados por la federación con cada uno de los estados y en los cuales, para su vigencia, se ordena su publicación a nivel federal y estatal.<sup>6</sup> En virtud de la puntual recepción del *Diario Oficial de la Federación*, el análisis de estos convenios se hace tomando como fuente este órgano informativo. La recepción posterior de su publicación a nivel estatal no es objeto de un nuevo análisis, sino que se introduce al sistema por medio de una estructura de análisis preconcebida.

El segundo caso es el de las reformas a la Constitución política federal, para cuya entrada en vigor es indispensable el voto de las dos terceras partes de los presentes en el congreso de la unión, y su aprobación por la mayoría de las legislaturas de los estados (artículo 135 constitucional). Con fundamento en este precepto, las legislaturas de los estados que aprueban la reforma, publican su texto en el periódico oficial local. En virtud de la irregularidad de la publicación de las aprobaciones estatales (que llega a ser posterior a la publicación de la reforma en el *Diario Oficial de la Federación*), UNAM-JURE hace caso omiso de aquéllas, para ocuparse exclusivamente de su análisis a nivel federal.

### iii) La selección de los textos legislativos

En el caso de la legislación federal, y al contrario de lo que sucede con la estatal, el criterio para su selección es más amplio. Debido a la importancia de los textos federales se hacen varias excepciones al principio de no selección de textos cuyos efectos son temporal, espacial o personalmente limitados. De esta manera, los acuerdos que sujetan a permiso la importación de ciertas mercancías o los avisos que determinan el valor del peso mexicano frente a otras monedas extranjeras (efectos temporalmente limitados); los reglamentos de pilotaje y las declaratorias de parques nacionales (efectos espaciales restringidos), y las bases especiales de tributación

<sup>6</sup> Cf. D. O. del 13 de septiembre (Convenio de coordinación para la procuración de justicia entre la federación y los Estados).

o la aprobación de tarifas por sector o rama de actividad (cuya aplicación se limita a un grupo de personas), son todos objeto de análisis.<sup>7</sup>

## B. MÉTODOS DE ANÁLISIS DE LA INFORMACIÓN LEGISLATIVA

Entre los renglones del conocimiento humano en que es imprescindible contar con medios que permitan recuperar la *totalidad* de la información relacionada con un punto de interés, se encuentra, sin duda alguna, la legislación. En efecto, la no recuperación de una pieza del “rompecabezas legislativo” equivale a no tener presente la regulación real de una materia determinada. Con ello, además del desconocimiento absoluto de disposiciones obligatorias —por parte de aquellos a quienes van dirigidas—, se corre el riesgo de que los órganos legislativos elaboren disposiciones contradictorias o paralelas, con las negativas repercusiones sociojurídicas que ello implica.

Durante mucho tiempo, la recuperación de información (incluyendo la legislativa) sólo se efectuó a través de métodos manuales y mecánicos basados en sistemas de lenguaje cerrado. Estos sistemas consisten en la utilización de listas rígidas de descriptores,<sup>8</sup> tanto para elaborar fichas de análisis como para acceder a la información en ellos contenida y por su propia naturaleza no permiten la recuperación de información mediante la integración de conjuntos.<sup>9</sup>

Con la aparición de las computadoras, los sistemas de lenguaje cerrado vieron incrementada, de manera impresionante, su velocidad de operación. Sin embargo, subsistía la imposibilidad de recuperar exhaustivamente la información legislativa. Debido al desarrollo de las técnicas informáticas, los científicos de la computación empezaron a considerar factible la implementación de sistemas que permitieran una recuperación tan exhaustiva como se deseara, sin limitarse al empleo de las listas de descriptores. Surge así la posibilidad de implementar el uso del lenguaje natural, tanto para la elaboración de fichas de análisis como para su recuperación. El

<sup>7</sup> Diarios Oficiales Federales del 18 de enero de 1974 (permisos de importación), 17 de julio de 1980 (declaratoria de parque nacional del anfiteatro de la Bahía de Acapulco), 17 de mayo de 1983 (Regulación Particular de Pilotaje para los amarraderos de Rosario, Baja California), 10 de mayo de 1984 (bases especiales de tributación para agricultores dedicados al cultivo del henequén) y del 29 de marzo de 1984 (tarifas aplicables a servicios de coches-dormitorio).

<sup>8</sup> También denominados *unats*, *unitems*, palabras clave, voces de acceso, palabras llave, etc.

<sup>9</sup> Entendemos por *conjunto legislativo* la totalidad de la información relativa a una determinada interrogación. Por ejemplo: a la interrogación “Municipios, 1976 a 1978”, correspondería como respuesta al conjunto de información integrado por todas las disposiciones en que se aludiera a los municipios, independientemente que se trate de leyes, decretos, circulares, reglamentos, etc.

“nuevo reto informático” abría para el derecho, la esperanza de resolver el problema de recuperación de información en forma completa. Es precisamente este reto el que hemos asumido en el sistema UNAM-JURE.

### 1. Los métodos “tradicionales”

Se ha señalado los tipos de lenguaje de análisis-acceso, y hemos ubicado a nuestro sistema dentro de los de tipo abierto.

En este apartado nos ocuparemos de analizar los métodos de tratamiento de información contenida en documentos-fuente, que pueden calificarse de “tradicionales”. Estos son el método de *indexación* y el método de *full-text*.

#### a. La *indexación*

Como ya señalamos los sistemas de lenguaje cerrado se basan en una lista rígida de descriptores, que son empleados tanto para tener acceso a la información contenida en fichas de análisis, como para la elaboración de éstas. Pues bien, *indexar* consiste, precisamente, en dicha elaboración a través de la calificación de la información contenida en documentos-fuente, mediante el descriptor o descriptores que se consideren apropiados. Por ejemplo:

Se publican reformas al artículo noveno de la Constitución Política Federal (garantía de libertad de asociación). De la lista de descriptores (o voces de acceso), se empleará la que se considere pertinente para calificar la información contenida en la disposición. En este caso pudiera ser “Constitución Federal, reformas”.

En materia legislativa, la rigidez de la lista no lo es tanto, pues se hace necesario, generalmente, agregarle descriptores que reproduzcan el nombre de disposiciones importantes. Sin embargo, mediante este método no es posible la recuperación por materias, cuando son diversas las tratadas por una ley (por ejemplo: la *ley miscelánea*).

En lo referente a las aplicaciones de la *indexación* (y por tanto, del lenguaje cerrado) en materia jurídica, la opinión predominante en la doctrina y la práctica se ha pronunciado a favor de su implementación para el tratamiento de información bibliográfica, hemerográfica y jurisprudencial. Son muy contados los autores que defienden su aplicación a la legislación.

En cuanto a las ventajas e inconvenientes que suelen predicarse respecto a la *indexación*, podemos citar los siguientes:



Ventajas: la única es de carácter financiero en el sentido de que “re-duce el material de base y por consiguiente, el costo de alimentación”<sup>10</sup> de los datos.

Inconvenientes: condiciona al sistema informático a funcionar con un lenguaje cerrado, haciendo imposible la integración de los conjuntos de información (véase nota 9) —requiere de personal especializado, que designe con los descriptores correctos cada documento-fuente. La califi-cación de información, por parte de los analistas, es meramente subjetiva. Casi siempre es posible encuadrar cierto tipo de información dentro de distintos descriptores, dependiendo del enfoque dado. Esto se agrava si consideramos que no en todos los casos se encuentran descriptores que, con un mínimo de exactitud, puedan servir para calificar determinado tipo de información. Y un documento mal indexado es irrecuperable.

—Debido a que la característica común a los descriptores es la de constituir “calificativos generales”, ante una interrogación por informa-ción específica, se produce el llamado *ruido informático*. Éste consiste en la obtención de gran cantidad de información distinta a la requerida.

—En ciertos sistemas, la interrogación por información específica, no sólo trae aparejada la aparición de *ruido informático*, sino que en realidad no puede obtener respuesta satisfactoria, es decir, puede no dar salida a información o sólo hacerlo parcialmente. Este problema se conoce con el nombre de *silencio informático*.

#### b. El *full-text*

Como se recordará, con la evolución de las técnicas informáticas se dio lugar a las investigaciones tendentes a implementar sistemas de lenguaje abierto o natural. Dichos sistemas para ser viables han requerido la crea-ción de instrumentos lingüísticos que resuelvan, de manera automática, ciertos problemas gramaticales y semánticos, que surgen de las diversas formas en que se puede expresar una misma idea.<sup>11</sup> De igual forma ha sido indispensable crear nuevos métodos, distintos del de *indexación* para el tratamiento de información contenida en documentos-fuente. Es así como surge (a manera de primer intento) el llamado *full-text*, que con-siste en el almacenamiento del texto completo en la computadora, con el fin de recuperar la información contenida en él por cualquiera de las ma-terias a que haga referencia.

<sup>10</sup> Le Boulengue de Galard, Michele, “El tratamiento electrónico de la docu-mentación jurídica”, en *Anuario de la Facultad de Derecho*, Mérida, Venezuela, 1973, núm. 4, pp. 189 y 190.

<sup>11</sup> Véase C. 2. y 3.

En la práctica el método del *full-text* ha tenido poco éxito, pues no ha permitido la utilización del lenguaje natural.<sup>12</sup> Sin embargo, presenta notables ventajas en relación al de indexación.

En cuanto a las aplicaciones de este método, las opiniones, en su mayoría, coinciden en considerarlo ideal para el tratamiento de información legislativa.<sup>13</sup>

A efecto de probar la afirmación anterior, transcribimos algunas opiniones en defensa del *full-text*:

Edgar Salazar Cano: "...en lo referente al ingreso de la información nos parece mucho más deseable, aunque mucho más costoso y difícil, un tratamiento integral de los textos jurídicos (*full-text*), es decir, que ellos no deberían ser abreviados, mutilados o seleccionados a través de procedimientos manuales preliminares".<sup>14</sup>

Fernando Fueyo Laneri: "Al recogerse la información debe hacerse con fidelidad de la fuente, de manera que no se modifique el contenido de esta última, ni se dificulte su comprensión ni aparezca ambigüedad o duda alguna, ni siquiera en cuanto a la individualización de la fuente misma... por lo mismo es que las reproducciones literales íntegras son superiores a las fragmentarias o de síntesis".<sup>15</sup>

Alberto E. Serrano Pirella: Refiriéndose a los casos en que se ha de aplicar el método de la *indexación* o el de *full-text*, señala que: "No se puede sentar un principio general para resolver esta cuestión, la cual —a mi modo de ver— ha de ser resuelta caso por caso. Por ejemplo, en el caso de leyes, reglamentos, etcétera, parece necesario acudir al sistema de texto integral; sin embargo, en el caso de la jurisprudencia, parece suficiente con el almacenamiento de resúmenes y en el caso de opiniones

<sup>12</sup> Cabe aclarar que consideramos *full-text* al almacenamiento del texto con pretensión de recuperar el documento fuente por medio de un lenguaje abierto, sobre las mismas palabras del texto. Por tanto excluimos de tal noción el método empleado por aquellos sistemas que aunque se basan en la transcripción de los textos, únicamente pueden recuperar información a través de un listado o índice de términos empleados por el legislador. Estos métodos los consideramos de *indexación*.

<sup>13</sup> Cabe aclarar que estas opiniones son de carácter doctrinal y se emiten pensando en la aplicación de la transcripción de textos, recuperables gracias al empleo de descriptores. Hasta el momento no se ha tenido noticia de ningún sistema que, utilizando el método de *full-text* propiamente dicho, haya permitido una recuperación precisa de información legislativa, a causa del *ruido informático* que provoca.

<sup>14</sup> Salazar Cano, Edgar, *La informática jurídica: posibilidades y límites en la automatización de la información jurídica y del razonamiento jurídico*, Valencia, Venezuela, Universidad de Carabobo, Facultad de Derecho, 1978, pp. 46 y 47.

<sup>15</sup> Fueyo Laneri, Fernando, "Teoría y práctica de la información jurídica", en *Estudios de Derecho*, año XXXVI, Segunda época, vol. XXXIV, Núm. 87, p. 34, Medellín, Colombia, marzo de 1975.

doctrinarias, bastaría con un resumen o, aun más, una simple referencia bibliográfica".<sup>16</sup>

En cuanto a las ventajas e inconvenientes que se predicán de este método, tenemos:

### i) Ventajas

a) En teoría, elimina la limitación de tener que usar un lenguaje cerrado para acceder a la información, y se puede interrogar a la computadora en un sistema de lenguaje abierto o natural. Esto, como más adelante lo enunciamos, lejos de constituir una ventaja, da lugar al principal inconveniente de este tipo de sistemas, pues origina el llamado *ruido informático*.

b) No es necesario, desde el punto de vista jurídico, un personal especializado que "transcriba" las leyes a la computadora.

c) Elimina el subjetivismo de los analistas que aplican el método de *indexación*.

### ii) Inconvenientes

a) Su costo es muy elevado en comparación con el método de *indexación*, debido al espacio de almacenamiento que requiere.

b) Como se trata de una simple transcripción, no hace caso de los conceptos implícitos del documento-fuente, por los cuales también puede describirse información y, por tanto, interrogarse.

c) Ocupa espacio innecesario en el almacenamiento de la computadora, debido a las inútiles repeticiones de términos que fácilmente podrían reagruparse de manera lógica, a través de un análisis adecuado.

d) El dispositivo de salida de la computadora reproduce todo el texto legislativo lo que, por una parte, hace lenta y difícil la consulta del usuario del sistema para localizar los puntos de interés,<sup>17</sup> y por otra, duplica información que, una vez individualizada, puede fácilmente ser consultada en el soporte documentario original.

e) El *ruido informático* es, sin duda alguna, el mayor inconveniente del uso del *full-text*, y consiste en la obtención de información no pertinente.

<sup>16</sup> Serrano Pirela, Alberto E., *Computadoras y derecho, una introducción a la informática jurídica*, Maracaibo, Venezuela, Universidad del Zulia, 1975, pp. 15 y 16.

<sup>17</sup> Por ejemplo: imagínese el despliegue en pantalla de rayos catódicos de todos los reglamentos interiores de las secretarías de estado, cuando el universo de información que se busca conocer es el de las Direcciones Generales comunes a ellas.

Esta problemática se debe a la conjunción que hace la computadora de términos que, aunque adquieren un sentido lógico y concordante a cierta interrogación, dan lugar a la recuperación de información distinta a la que realmente le interesa. A las combinaciones producto de este tipo de conjunciones se les denomina “combinaciones no pertinentes” y se deben a la no delimitación de los campos, dentro de los cuales la computadora debe formar sus conjuntos, o una delimitación no adecuada.<sup>18</sup> Por ejemplo: alguien que pregunte por *obra pública*, recuperará seguramente el reglamento de la Secretaría de Educación Pública, que entre las facultades que le asigna, comprende algunas relacionadas con obra literaria. La computadora conjunta el *obra* (literaria) y la palabra *Pública* del nombre de la Secretaría, y nos dará esta información no pertinente.

Como consecuencia de la imposibilidad de recuperación de información específica, debido a los problemas técnicos apuntados, no sólo es posible que se produzca *ruido* sino también *silencio informáticos*.

Las ventajas e inconvenientes de los métodos de *indexación* y *full-text* se pueden observar claramente en el siguiente cuadro:

### *Indexación*

### *Full-Text*

#### A. Ventajas

- a) Su economía.

#### B. Inconvenientes

- a) Uso del lenguaje cerrado.
- b) Requiere personal altamente especializado.
- c) Subjetivismo en el análisis.
- d) *Ruido informático*.
- e) *Silencio informático*.

#### A. Ventajas

- a) Permite la recuperación del documento de origen a través del empleo de un sistema de lenguaje abierto.
- b) No requiere de personal especializado.
- c) Elimina el subjetivismo del análisis.

#### B. Inconvenientes

- a) Costo de almacenamiento.
- b) No alude a conceptos implícitos.
- c) Ocupación innecesaria de espacio de almacenamiento.
- d) Hace difícil la consulta.
- e) *Ruido y silencio informáticos*.

<sup>18</sup> La delimitación que se hace, al menos en algunos sistemas de informática jurídica que emplean *full-text*, parte de la identificación de signos delimitantes, por parte de la computadora, concordantes con signos de puntuación del soporte de origen. Por ejemplo: cada punto marcará el principio y fin de cada campo de conjugación. A estos signos delimitantes se les conoce en computación con el término de *restrictores de distancia*.

Hasta aquí hemos expuesto lo relativo a los métodos de *indexación* y *full-text*. En el siguiente apartado haremos referencia al *abstract legislativo*, método que pretende superar los inconvenientes de sus análogos.<sup>19</sup>

## 2. El abstract legislativo

### a. Consideraciones generales

Como una tercera alternativa para el tratamiento de la información documental, se presenta la utilización del método que culmina con la formulación de un *abstract*. El sistema UNAM-JURE adoptó este método de análisis de los documentos legislativos, considerando que, con el mismo, no sólo resultan superados los inconvenientes que plantea la aplicación de los métodos de *indexación* y *full-text*, sino que, además, presenta varias ventajas importantes, como son —entre otras— las siguientes:

a) El uso de un lenguaje abierto, libre, tanto en la interrogación del banco de datos como en la elaboración de las fichas de análisis.

b) La posibilidad de explicitar, en las fichas de análisis, las figuras jurídicas implícitas en el documento-fuente.

c) La discriminación de información que no tiene relevancia jurídica, o que es repetitiva.

d) Otorgar al analista la flexibilidad necesaria para adaptarse, en forma adecuada, a las características o particularidades de la información documental legislativa.

e) Permite reagrupar la información relacionada con un mismo concepto, órgano, autoridad o institución, que se encuentra desarrollada en diversos artículos o capítulos. Igualmente cuando una misma disposición hace alusión a dos o más conceptos, es posible clasificar en forma conveniente la información, utilizando *párrafos* distintos del *abstract*, uno para cada concepto (véase B.2.b.i.).

f) Puesto que son frecuentes los casos de reglamentación excesiva de instituciones jurídicas, lo cual origina duplicidad de disposiciones que es necesario corregir, el *abstract* constituye el medio idóneo para ello.

g) En la redacción de las leyes, y del derecho en general, el legislador emplea el lenguaje natural, lo que en muchos casos origina la falta de coincidencia entre la idea que se quiere comunicar y aquello que realmente se manifiesta. Ello se debe, por una parte, a la riqueza de dicho

<sup>19</sup> Los puntos relectivos a las ventajas e inconvenientes de los métodos de *indexación* y *full-text*, así como las opiniones de los doctrinarios en relación a uno y otro, son tomados de, Cáceres, Enrique, *La teoría del abstract legislativo*, Tesis profesional, México, Facultad de Derecho de la UNAM, 1984.

lenguaje y, por la otra, a las limitaciones del mismo. La riqueza del lenguaje natural se pone de manifiesto tanto por las posibilidades y variedad de formas en que una idea puede expresarse —en observancia de las distintas reglas de la sintaxis—, como por la existencia de una gran cantidad de sinónimos o análogos. Las limitaciones del lenguaje están referidas a la existencia de palabras que son vagas o que son polisémicas, es decir, tienen más de un significado o que pueden referirse a más de un objeto. Los problemas de técnica legislativa que se originan por lo anterior, son afrontados exitosamente por el *abstract* con el auxilio de los instrumentos lingüísticos: *Léxico* y *Thesaurus* (véase C.2. y 3.).

Aun cuando el Programa de Automatización de la Legislación Mexicana, UNAM-JURE, es pionero en la aplicación del *abstract* para el tratamiento de la información documental legislativa, el IRETIJ (*Institut de Recherches et d'Etudes pour le Traitement de l'Information Juridique*), de Montpellier, Francia, equipo asociado al Centre Nationale de la Recherche Scientifique lo introdujo y utiliza con magníficos resultados en materia de jurisprudencia. En 1983 el instituto francés y el Instituto de Investigaciones Jurídicas de la UNAM, celebraron un convenio de colaboración en materia de informática jurídica, que ha permitido un valioso intercambio de experiencias en la instrumentación de sus respectivos programas.

No es necesario insistir sobre las importantes diferencias existentes entre la jurisprudencia y la legislación (véase A.1.), baste considerar en cuanto a la forma la distinta presentación, estructura y extensión, de cada una; y en cuanto al fondo, la necesaria concretación de la primera, oponible a la abstracción, generalidad e impersonalidad, de la segunda; así como otras diferencias sustanciales relacionadas con la obligatoriedad, reformabilidad y vigencia de una y otra.

Cabe advertir que si bien el *abstract legislativo* del sistema UNAM-JURE, tomó como punto de partida el *abstract jurisprudencial* del IRETIJ, aquél se ha apartado de éste, haciéndolo marcadamente diferente; circunstancia que al propio tiempo planteó la necesidad de comenzar a desarrollar una *teoría del abstract legislativo* y de introducir importantes modificaciones en los aspectos estructurales y de operación de los instrumentos lingüísticos, así como en sus mismas finalidades.

## b. Estructura y función de la *ficha de análisis*

### i) El *abstract*

Con el término *abstract legislativo* hacemos referencia al documento cuya información, obtenida de un documento-fuente, es organizada en

forma lógica a través del empleo de restrictores de distancia con el fin de lograr su recuperación, así como su presentación sintética. Al *abstract* se le han asignado dos funciones interrelacionadas:

a) La recuperación automática de la información legislativa, con la indicación del tipo y ubicación del documento-fuente.

b) Proporcionar al usuario una idea más precisa de los datos que se encuentran en el documento-fuente, y su localización en él.

Estas finalidades son las que determinan el tipo de tratamiento al que se somete la información del documento-fuente y, en última instancia, también los criterios que aplica el analista para la elaboración del *abstract* (véase B.2.c).

El *abstract* se estructura en una serie de *párrafos* separados por puntos; cada *párrafo*, a su vez, se divide en *frases*, separadas por diagonales. El principio y el fin del texto del *abstract* se señala con asteriscos.

/            / restrictores de distancia entre *frases*.  
./           / . restrictores de distancia entre *párrafos*.  
\*/    /./    /\* inicio y fin de *abstract*.

El documento-fuente, normalmente, trata de una serie de conceptos, problemas, facultades y obligaciones, instituciones jurídicas, etcétera, pero no siempre en forma sistemática y organizada (por ejemplo, en uno o varios artículos o capítulos). El *párrafo* y la *frase* del *abstract* reorganizan y sistematizan los datos contenidos en el documento-fuente, formando con ellos conjuntos y subconjuntos.

#### a') El *párrafo* y la *frase*

El *párrafo* es el desarrollo de una unidad de información del documento-fuente, es decir, el *párrafo* —en principio— desarrolla la información referente a un problema, concepto o figura jurídicos.

La primera *frase* de cada *párrafo* contiene el nombre y *status* de la disposición. La segunda *frase*, denominada “cabeza de párrafo”, es la que indica el concepto, materia o facultad, que se desarrolla en el *párrafo*. Por ejemplo:

./LEY DEL NOTARIADO ESTATAL, REFORMAS Y ADICIONES,  
ARTÍCULOS 13 Y 14/GOBERNADOR DEL ESTADO, FACULTADES Y OBLIGACIONES/ ASIGNAR A LOS NOTARIOS PÚBLICOS DISTRITOS PARA EL EJERCICIO DE SUS FUNCIONES/.

La *frase* contiene un subconjunto del conjunto desarrollado en un *párrafo*, esto es, la frase enuncia ideas, facultades o aspectos, que tienen relación con la unidad de información que abarca el *párrafo*. Por ejemplo:

/CLASIFICACIÓN DE FRACCIONAMIENTOS/  
/OTORGAMIENTO DE LICENCIAS DE CONSTRUCCIÓN/

b') Los niveles de interrogación

El *párrafo* y la *frase* constituyen los dos niveles en que se puede interrogar la información contenida en el *abstract*. La interrogación en el *nivel párrafo* permite relacionar la información de los subconjuntos que forman un *párrafo*.

El *nivel frase* permite la recuperación de un concepto o problema específico. En consecuencia, de no agruparse adecuadamente la información del documento-fuente (tratando un solo concepto o figura en el *párrafo*) es fácil obtener documentos no pertinentes (*ruido*).

Es evidente que cierta clase de información, y los conceptos que la complementan, pueden situarse únicamente en uno de ambos *niveles* (por ejemplo, el nombre de organismos o de disposiciones en el *nivel frase*), lo que simplifica el problema del nivel en que ha de interrogarse.

En este punto, conviene insistir en lo importante que resulta la posibilidad que tiene el usuario de interrogar el banco de datos utilizando un lenguaje abierto, libre, natural, siendo necesaria la determinación previa del perfil de interés, por el que se precisen las materias o temas, motivo de la investigación.

Asimismo, el uso de los operadores lógicos, la interrogación a *nivel párrafo* o *frase*, y la posibilidad de relacionar —para efectos de recuperación— cualquier parte del *abstract* con la *zona de campos fijos*, cuyo mecanismo de funcionamiento se explica en el siguiente apartado, permiten la recuperación completa, precisa y pertinente, de la información legislativa.

El *abstract legislativo*, pues, pretende analizar y tratar toda la información de un documento-fuente, mas no reproducir simplemente el contenido de éste (que equivale al *full-text*). Podríamos afirmar que intenta aproximarse a un sistema que llamaríamos de *full-information*.

A través del *abstract*, el sistema UNAM-JURE solamente permite la localización automática de información legislativa. El acceso al documento-fuente es indirecto. El *abstract* tiene por objetivo la selección de la información y nunca sustituye al documento-fuente, al que el usuario deberá remitirse siempre.



ii) La zona de campos fijos y el resumen

Son partes complementarias de la *zona de abstract* en las fichas de análisis, las zonas de campos fijos y de resumen. Sólo la zona de campos fijos, junto con la de *abstract*, pueden ser objeto de interrogación.

I. *Zona de campos fijos*: contiene la información descriptiva general del documento y los datos de identificación de la ficha. Comprende los siguientes datos codificados:

1. Número de referencia (campos 1-6): número de 6 dígitos que sirve para identificar cada una de las fichas elaboradas. Los 2 primeros dígitos corresponden al año de elaboración de la ficha y los 4 restantes se asignan en forma progresiva.

2. Procedencia geográfica (campos 7-8): determina el estado de la República de donde proviene el documento. Se ha elaborado una tabla de correspondencias en la que se ha asignado una clave de dos dígitos a cada uno de los estados, así como a la Federación. De este modo  $\phi\phi$  indica procedencia federal;  $\phi 1$ , de Aguascalientes;  $\phi 2$  Baja California, y así sucesivamente hasta el número 32 que corresponde a Zacatecas (ver anexo.).

3. Tipo de documento (campos 9-10): describe con dos letras el tipo de documento de que se trata. Se ha elaborado una tabla que pretende codificar todos los tipos de documentos que se han encontrado. Por ejemplo LE = LEY; DE = DECRETO, etcétera. (Ver anexo 2.)

Existía el problema de determinar si la codificación del documento debía atender al carácter formal o al contenido material del mismo. Se adoptó el criterio material, ya que permitía mayor uniformidad y objetividad en el uso del descriptor, así como mayor congruencia con la información que se pretendía dar. De este modo si se tiene un decreto que reforma una ley, el codificador debe ser LE, ya que es el contenido material del documento.

En caso de que un mismo decreto reforme dos tipos de documento, ejemplo una ley y un decreto, deberán hacerse dos fichas, una por cada disposición, a excepción de la "Ley miscelánea" que aparece anualmente.

4. Nombre de la publicación (campos 11-12): nombre de la publicación en la que aparece el documento analizado. Por ejemplo Diario Oficial, Periódico Oficial, etcétera. Se ha elaborado una tabla de codificación. (Ver anexo 3.)

5. Fecha de publicación (campos 13-18): en este espacio se señala con 6 dígitos la fecha de publicación del documento analizado. Los dos pri-

meros indican el día, los dos siguientes el mes y los dos últimos el año de la publicación.

6. Campo extraordinario (campo 19): espacio en el que se indica con una letra si el documento se encuentra, en un alcance, suplemento, número extraordinario o sección especial de la publicación. Este dato es importante, pues en muchos casos, sobre todo a nivel estatal, los documentos relevantes no se publican en los números ordinarios. Este campo no siempre se usa, sólo en caso de que sea necesario. Existe tabla de codificación. (Ver anexo 4.)

7. Clave de microfilme (20-28): lugar en el que se indica con 8 dígitos el número de rollo y los pies de éste en donde se encuentra el documento. Los dos primeros espacios corresponden al rollo y los seis siguientes al número de los cuadros de éste en donde se encuentra el documento. Ejemplo de zonas de campos fijos:

8 3 8 7 5 6  $\phi$  1 A C P O 1 2  $\phi$  5 7 8  
A A A 1 4 5 1 4 8

1. Número de referencia
7. Estado: Aguascalientes
9. Tipo de documento: Acuerdo
11. Publicación: Periódico Oficial
13. Fecha de publicación del documento: 12 de mayo de 1978
19. Campo extraordinario: Alcance
20. Microfilme: Rollo AA, cuadros del 145 al 148

II. *El resumen*: es una descripción breve e informal del documento, con el objeto de que el lector tenga una guía del contenido general, extensión y fecha de promulgación del documento fuente. Esta parte tiene fines de información para el usuario, pero los datos asentados en él no son susceptibles de recuperación por interrogación. Aparecen en pantalla cuando el documento ha sido localizado a través de la interrogación a la *zona de abstract* o de campos fijos.

El resumen se toma, por regla general, del sumario de la publicación, eliminando las frases o palabras que no aporten nada significativo. Por ejemplo, en lugar de "Honorable Congreso de la Unión" o de "Ciudadano Gobernador del Estado Libre y Soberano de Chiapas", basta con señalar: "Congreso de la Unión" o "Gobernador de Chiapas".

En ciertos casos basta con señalar el nombre y *status* de la disposición para dar idea del contenido de ésta.

En muchas ocasiones las leyes o decretos, etcétera, vienen numerados. No es necesario poner el número de la ley o decreto, pues éstos responden

a muy diversos criterios, creando muchas veces sólo confusión sin añadir información relevante.

En caso de que la publicación no tenga sumario, el analista debe formular el resumen. En estos casos es útil leer el primer artículo del documento pues, por lo general, es posible redactarlo a partir de éste.

Después del nombre y contenido general se pone punto, y se indica a continuación el número de artículos, incluso los transitorios, de que consta el documento, y las páginas en las que se encuentra dentro de la publicación. Para este efecto en caso de reformas a leyes, reglamentos, etcétera, en el resumen se indica de cuántos artículos consta el decreto de reformas y no los artículos reformados.

Finalmente se indica con seis dígitos la fecha de promulgación del documento, que se refiere al día en que el ejecutivo ordena la publicación del texto, y no a la fecha de aprobación del órgano legislativo. Por regla general estos datos se encuentran al final del documento fuente, después de los artículos transitorios.

En caso de convenios internacionales, en la primera parte del resumen se pone la fecha de firma del documento, y en la última la fecha de promulgación del mismo en el *Diario Oficial*.

A continuación se dan algunos ejemplos de resumen:

LEY DE FOMENTO GANADERO DEL ESTADO DE QUINTANA ROO, CONSTA DE 256 ARTÍCULOS Y 3 TRANSITORIOS Y VA DE LA PÁGINA 1 A LA 43 DEL PERIÓDICO OFICIAL. 161276.

DECRETO QUE REFORMA DIVERSOS ARTÍCULOS DEL REGLAMENTO INTERIOR DE LA SECRETARÍA DE REFORMA AGRARIA, CONSTA DE 4 ARTÍCULOS Y ESTÁ EN LA PÁGINA 5 DEL DIARIO OFICIAL. 23ϕ581.

DECRETO QUE PROMULGA LA TERCERA CONFERENCIA DE LAS NACIONES UNIDAS SOBRE EL DERECHO DEL MAR FIRMA EL 25ϕ479 EN LA CIUDAD DE GINEBRA. CONSTA DE 354 ARTÍCULOS Y 4 ANEXOS. VA DE LA PÁGINA 12 A LA 84 DE LA SEGUNDA SECCION DEL DIARIO OFICIAL. 23ϕ481.

ACUERDO POR EL QUE SE AMPLÍA HASTA EL 311283 EL PLAZO PARA CONTINUAR UTILIZANDO MARBETES EN LOS ENVASES DE BEBIDAS ALCOHÓLICAS NACIONALES DISTINTAS DE LAS CERVEZAS. CONSTA DE 6 ARTÍCULOS Y VA DE LA PÁGINA 4 A LA 5 DEL DIARIO OFICIAL. ϕ2ϕ883.

DECRETOS POR LOS QUE SE DECLARAN NULAS LAS ELECCIONES DE LOS AYUNTAMIENTOS DE LOS MUNICIPIOS QUE SE INDICAN. CONSTA DE 3 ARTÍCULOS Y ESTÁ EN LA PÁGINA 22 DEL PERIÓDICO OFICIAL. 151282.

### c. Los criterios de análisis

En función de las finalidades del *abstract* se han desarrollado una serie de criterios para la elaboración de las fichas de análisis. A continuación se enuncian los lineamientos generales que el analista debe seguir:

A. *El párrafo*: La estructura del párrafo se guía por los siguientes lineamientos:

1. Una primera frase que contenga el nombre y *status* (si se trata de publicación o reformas, y en este caso qué artículos). Ejemplo:

/LEY DE HACIENDA ESTATAL, PUBLICACIÓN/  
/LEY DE HACIENDA ESTATAL, REFORMAS Y ADICIONES,  
ARTÍCULO 95/

En el caso de reformas debe referirse al número de los artículos reformados.

2. *La cabeza de párrafo* (véase B.2.b) es la que indica el concepto, problema, materia o facultad que se va a desarrollar en las frases siguientes. La elección de esta frase es de la mayor importancia y corresponde al género o conjunto, cuyas especies o subconjuntos se explicitan en el desarrollo posterior del párrafo.

3. Las frases siguientes son el desarrollo del género o conjunto anunciado en la cabeza de párrafo. Estas frases van siempre ligadas o referidas a la cabeza, debiendo ser la primera la más cercana a aquélla y la última la que tenga menor relación. En otras palabras, las frases deben ir de lo general a lo particular, pero siempre relacionadas con la cabeza de párrafo. Si esto no sucede significa que se trata de un concepto distinto, al que se le debe dar tratamiento en un párrafo independiente. Ejemplos de párrafos:

/LEY ORGÁNICA DEL BANCO DE MÉXICO, ARTÍCULO 27. REFORMA/BANCO DE MÉXICO, FACULTADES/ESTABLECIMIENTO DE REGLAS GENERALES PARA EL ORDEN EN QUE HAN DE PRACTICARSE OPERACIONES CON INSTITUCIONES DE CRÉDITO/LÍMITES AL VOLUMEN GENERAL DE CRÉDITO/. /LEY ORGÁNICA DEL BANCO DE MÉXICO, ARTÍCULO 56. REFORMA/BANCO DE MÉXICO, JUNTA DE GOBIERNO, FACULTADES/APROBACIÓN DEL PROYECTO DE PRESUPUESTO DEL BANCO DE MÉXICO/DETERMINACIÓN DEL SUELDO DEL DIRECTOR GENERAL/GRATIFICACIONES ANUALES DE FUNCIONARIOS Y EMPLEADOS/.

4. Puede suceder que la cabeza de párrafo corresponda a un conjunto muy amplio, es decir, que la información contenida en el texto-fuente sobre un problema determinado sea muy extensa. Agrupar toda esta in-

formación en un solo párrafo haría que éste fuera demasiado grande con el peligro de crear *ruido* en el momento de la interrogación.

Para evitarlo es posible desarrollar varios párrafos sobre un mismo problema, en donde las dos primeras frases serán iguales y la tercera será la que delimite el problema específico a desarrollar:

/LEY DE FRACCIONAMIENTOS ESTATAL, PUBLICACIÓN/  
CLASIFICACIÓN DE LOS FRACCIONAMIENTOS/FRACCIONA-  
MIENTOS URBANOS .../. /LEY DE FRACCIONAMIENTOS  
ESTATAL, PUBLICACIÓN/CLASIFICACIÓN DE LOS FRACCIO-  
NAMIENTOS/ FRACCIONAMIENTOS INDUSTRIALES.../

5. *Las cabezas de párrafo* deben seguir, en la medida de lo posible, las divisiones y el lenguaje del legislador. Sin embargo, no es obligatorio. De hecho, en la mayoría de los casos es necesario que el analista formule cabezas de párrafo que no correspondan a las divisiones y lenguaje del legislador, reagrupando materias, problemas, conceptos que se encuentren dispersos. Estas cabezas corresponden, por ejemplo, a las facultades de determinado órgano o autoridad, a las sanciones, el procedimiento o figuras específicas que no son conceptos desarrollados explícitamente en el texto.

6. Es deseable que dentro de un mismo párrafo se utilice el mayor número de análogos posible, bien en una misma o en distintas frases, con el fin de ampliar las posibilidades de recuperación:

\*/PUBLICACIÓN DE LA LEY ESTATAL DE DEUDA PÚBLICA/  
REQUISITOS Y PROCEDIMIENTO PARA LA CONTRATACIÓN  
DE DEUDA PÚBLICA, FINANCIAMIENTOS, CRÉDITOS Y EM-  
PRÉSTITOS/.

B. *La frase*: 1. Al final de cada frase debe indicarse el número del, o de los artículos del documento fuente en donde se encuentran los problemas tratados, encerrando entre paréntesis el número o números correspondientes (información no recuperable).

Lo anterior permite al usuario saber el lugar del texto que debe consultar, aspecto que resulta necesario sobre todo en el caso de textos largos o mal estructurados:

/OBRAS PÚBLICAS, EJECUCIÓN (ART. 34 AL 38) AGUA PO-  
TABLE Y DRENAJE (ART. 36)/.

2. El objeto del *abstract* no se limita a la recuperación de la información, sino que busca también informar al lector sobre el contenido del documento, por ello es necesario redactar cada frase de modo que se en-

tienda con claridad cuál es la información que contiene el texto-fuente. Así, el usuario podrá determinar si el documento es, o no, pertinente.

3. Por la misma razón es también necesario que dentro de cada frase los conceptos vayan calificados, con el objeto de lograr precisión en la interrogación. Un ejemplo aclarará lo anterior:

**\*/REGLAMENTO DE CONSTRUCCIONES ESTATAL, PUBLICACIÓN/OBRAS PÚBLICAS/CONSTRUCCIÓN DE CALLES Y AVENIDAS/.**

En este caso, el concepto /OBRAS PÚBLICAS/ no debe aparecer solo, pues indica un concepto muy amplio que puede crear mucho *ruido* en la interrogación. Es necesario calificarlo, por ejemplo:

**\*/REGLAMENTO DE CONSTRUCCIONES ESTATAL, PUBLICACIÓN/OBRAS PÚBLICAS, EJECUCIÓN/CONSTRUCCIÓN DE CALLES Y AVENIDAS/.**

4. Lo anterior es especialmente válido respecto de las palabras polisémicas. Una palabra de este tipo nunca debe ir sola en una frase, pues provoca *ruido* e imprecisión en la interrogación.

C. El primer párrafo del *abstract*: 1. En principio, todos los documentos cuyo *status* sea "publicación" (en especial leyes y decretos) deben iniciarse con un primer párrafo que indique los conceptos más generales del documento que se analiza. Estos datos son, cuando los haya:

- a) En caso de documentos federales, las siglas de las dependencias que los expiden de acuerdo a la codificación respectiva.
- b) Nombre del documento-fuente.
- c) Ámbitos de validez personal, territorial, temporal y material.
- d) Principales autoridades responsables de la aplicación.
- e) Fundamentación siempre y cuando se encuentre dentro de los artículos del documento fuente.
- f) Conceptos generales implícitos.

Ejemplos:

**\*/LEY DE FRACCIONAMIENTOS ESTATAL, PUBLICACIÓN/ INTERVENCIÓN DE LA ADMINISTRACIÓN PÚBLICA EN DIVISIÓN DE PREDIOS O TERRENOS/GOBERNADOR Y SECRETARÍA DE OBRAS PÚBLICAS, FACULTADES, OBLIGACIONES/DESARROLLO URBANO/FRACCIONADORES, DERECHOS Y OBLIGACIONES/.**

**\*/SHCP/REGLAMENTO DEL ARTÍCULO 85 DEL CÓDIGO FISCAL DE LA FEDERACIÓN/DIRECCIÓN GENERAL DE INGRE-**

## SOS, FACULTADES, OBLIGACIONES/INGRESOS FEDERALES, EJERCICIO FISCAL 1983/.

2. En el caso de que el *status* de la disposición sea “reformas” no es necesaria la elaboración de este primer párrafo.

3. En ciertos casos el primer párrafo agota el contenido de la disposición, y por ello es suficiente la elaboración de un solo párrafo. Ejemplo:

**\*/SECRETARÍA DE PATRIMONIO Y FOMENTO INDUSTRIAL, APROBACIÓN DE NORMAS PARA PRODUCTOS INDUSTRIALES, DEROGACIÓN DE LA NORMA DGN “J” 14 195φ/\***

D. Aspectos generales: 1. Es necesario procurar hacer expresas las figuras implícitas en el documento que se analiza. Para lograrlo es deseable la especialización de los analistas, con el fin de tratar uniformemente dichas figuras, que varían de una materia a otra.

2. En caso de que un texto derogue o abroge explícitamente uno anterior, debe indicarse en un párrafo independiente al final del *abstract*.

**\*/PUBLICACIÓN DE LA LEY GENERAL DE DEUDA PÚBLICA/ DEROGACIÓN DEL ARTÍCULO 26 DE LA LEY ORGÁNICA DE NACIONAL FINANCIERA SA/DESAPARICIÓN DE LA COMISIÓN ESPECIAL DE FINANCIAMIENTOS EXTERNOS/\***

3. Nunca deben abreviarse las palabras dentro de las frases. En caso de abreviaturas de nombres propios, éstas deben ponerse después de escribir el nombre completo de la institución.

**/PETRÓLEOS MEXICANOS, PÉMEX/ORGANIZACIÓN DE LAS NACIONES UNIDAS, ONU/.**

4. Las fracciones, capítulos, apartados, secciones, títulos o volúmenes se escriben con números romanos o letras entrecomillados:

**/CONSTITUCIÓN POLÍTICA DE LOS ESTADOS-UNIDOS-MEXICANOS. REFORMAS Y ADICIONES, ARTÍCULO 123, APARTADO “B”, FRACCIÓN “XIII”/.**

Aquí resulta conveniente recordar que la estructura y redacción del *abstract* responden a las necesidades de la interrogación, o a la estructura de la ley (aunque en ocasiones coincidan), por lo que el analista está en posibilidad de reformular aquéllas de la manera que crea conveniente.

#### d. Problemas de forma y de fondo

##### i) De forma

Se refiere a la calidad o estilo de expresar las figuras jurídicas que contiene el documento-fuente, tomando en cuenta los problemas de redacción y puntuación:

##### a') Redacción

La redacción del *abstract* presenta como problema fundamental mantener el equilibrio entre las palabras superfluas —que hay que eliminar—, y la lectura clara del *abstract* sin menoscabo de la información jurídicamente relevante. Por ejemplo:

**\*/STPS/INSTITUTO NACIONAL DE PRODUCTIVIDAD, EXTINCIÓN, ORGANISMO PÚBLICO DESCENTRALIZADO/SECRETARÍA DEL TRABAJO Y PREVISIÓN SOCIAL, ASUNCIÓN O SUBROGACIÓN EN LOS DERECHOS DEL INSTITUTO/ TRANSFERENCIA DE BIENES, DERECHOS Y OBLIGACIONES, REUBICACIÓN DE TRABAJADORES DEL ESTADO./INSTITUTO NACIONAL DE PRODUCTIVIDAD, DESAPARICIÓN/FIDEICOMISO PARA EL CENTRO NACIONAL DE PRODUCTIVIDAD AC, EXTINCIÓN/\***

##### b') Signos de puntuación

El problema que presenta es determinar los signos que se deben incluir en el *abstract* y sus funciones, para facilitar su lectura y comprensión. Para ello se han establecido las reglas siguientes:

1) *Abreviaturas*. La regla general es que no se deben abreviar las palabras en el *abstract* excepto las abreviaturas *ART* para artículo, y *FR* para fracción, que se utilizan sólo entre paréntesis; y en caso de las siglas de organismos públicos y privados, se elabora actualmente un listado, para su recuperación tanto por el nombre como por las siglas.

2) *La función gramatical de la coma*. Para evitar complicaciones innecesarias en la redacción y en la lectura del *abstract*, se ha asignado a la coma, como único signo de puntuación, las funciones que desempeñan el punto y coma, los dos puntos, etc. Por ejemplo:

**\*/CONSTITUCIÓN POLÍTICA ESTATAL, ARTÍCULO 3, REFORMA/MUNICIPIO DE JALAPA, CAMBIO DE DENOMINACIÓN, AHORA JALAPA DE MÉNDEZ/\***



3) *Paréntesis: contenido y función.* El paréntesis contiene información no recuperable, y su función es la de ampliar o aclarar la información del *abstract*, sin entorpecer su lectura. Por ejemplo:

**\*/LEY ORGÁNICA DEL PODER JUDICIAL ESTATAL, REFORMAS, ARTÍCULOS 19, 33, 34, 6~~o~~/JUZGADOS FAMILIARES Y MIXTOS DE PRIMERA INSTANCIA, COMPETENCIA/RESOLVER ASUNTO DE PATERNIDAD, RECONOCIMIENTO DE HIJOS, ESTADO CIVIL (ART 19)/.**

**\*/CONSTITUCIÓN POLÍTICA DE LOS ESTADOS-UNIDOS-MEXICANOS, REFORMAS, ARTÍCULO 73/PODER LEGISLATIVO FEDERAL, FACULTADES/CAMBIAR LA RESIDENCIA DE LOS SUPREMOS PODERES DE LA FEDERACIÓN (FRACCIÓN "V")/\***

## ii) De fondo

Para el tratamiento de fondo de los documentos-fuente es necesario determinar qué información es irrelevante y cuál es susceptible de interrogar. Para enfrentar este problema se ha intentado adaptar el *abstract* a la legislación, y, por otra parte, señalar qué aspectos de la legislación deben adaptarse al *abstract*.

### a') Adaptación del *abstract* a la legislación

Consiste el establecimiento, en las fichas de análisis, de las diferentes etapas por las que atraviesa una ley (publicación, reformas, fe de erratas, derogaciones, abrogación) y evitar la duplicidad de análisis entre el ámbito estatal y federal. Para solucionar este problema se establecieron las siguientes reglas:

- Selección de documentos que están publicados a nivel estatal y federal.
- Determinar si la publicación estatal es requisito para la vigencia del documento, o es meramente informativa.

### b') Adaptación de la legislación al *abstract*

Para organizar la información y lograr su rápida y eficaz recuperación, se ha tratado de adaptar la legislación al *abstract*, afrontando la explicitación y la determinación de la profundidad en el análisis.

1) *La explicitación.* Consiste en expresar de manera clara las figuras jurídicas, para facilitar la interrogación de aquellos conceptos que se encuentran implícitos en las disposiciones. Por ejemplo:

En el Código civil para el Distrito Federal se establece la *gestión de negocios ajenos* (artículo 1896): “El que sin mandato y sin estar obligado a ello se encarga de un asunto de otro, debe de obrar conforme al interés del dueño del negocio”.

Para lograr la explicitación en los *abstracts* es necesario que las figuras implícitas sean tratadas de manera uniforme por los analistas (al igual, claro, que toda la información jurídicamente relevante), y para ello es deseable su especialización por materias.

2) *La profundidad del análisis*. La profundidad en el análisis del documento-fuente es uno de los problemas importantes que afronta el *abstract* legislativo. Del criterio que se adopte para su solución dependerá conocer qué tan específica es la información recuperable por el sistema.

Para solucionar el problema de la profundidad, se han establecido algunos principios que son guía para el analista:

—En principio, *toda* la información contenida en el documento-fuente debe estar en el *abstract*. Lo anterior lleva a un análisis general, de modo que todos los aspectos tratados por el legislador, aun los particulares, sean susceptibles de recuperar a través de la interrogación. Es necesario subrayar que se trata de toda la *información relevante* que debe tener el *abstract* y no de todo el contenido del documento-fuente.

—Las disposiciones que se expiden con frecuencia y que no cambian en su estructura general (por ejemplo: convenios de coordinación, presupuestos, declaración de validez de elecciones, etcétera) se les ha dado un tratamiento diferente, pues se han elaborado estructuras de análisis que permiten al analista tener una guía de la estructura y cantidad de información que debe contener cada *abstract*. Por ejemplo:

\* /LEY DE INGRESOS ESTATAL, EJERCICIO FISCAL 1983, PUBLICACIÓN/IMPUESTO, DERECHOS, PRODUCTOS, APROVECHAMIENTOS, PARTICIPACIONES FISCALES, ESTATALES, DETERMINACIÓN/\*

Importa, pues, dar cuenta *del qué y del quién*, pero no del cómo o del por qué. Por ejemplo:

/LEY DE OBRAS PÚBLICAS/CALLES, TIPOS Y MEDIDAS/

sería suficiente, ya que otra que indicara:

/LEY DE OBRAS PÚBLICAS, PUBLICACIÓN/TIPOS Y MEDIDAS DE CALLES/CALLES PRIMARIAS. 16 METROS DE ANCHO/CALLES SECUNDARIAS, 5 METROS DE ANCHO/

resulta innecesaria y excedería el objeto del *abstract*: el decir *cómo* y no limitarse al *qué*. En efecto, si se interrogara por *medidas de las calles*, se

obtendría la información necesaria para remitirse al texto donde se encuentra la respuesta que busca el usuario.

La profundidad debe ser, pues, adecuada a la interrogación. Así, el analista, al elaborar el *abstract*, debe valorar si la información puede ser relevante y de interés para un hipotético usuario.

## C. LOS INSTRUMENTOS LINGÜÍSTICOS

### 1. *Los problemas lingüísticos y su solución*

Determinar la forma de tratamiento del lenguaje es uno de los problemas básicos de cualquier sistema de recuperación automática de información. Como ya explicamos, se decidió que el sistema UNAM-JURE adoptara un lenguaje abierto.

La adopción de un lenguaje abierto, en cualquier sistema de recuperación automática, presenta problemas lingüísticos que se derivan, además de las características mismas del lenguaje, de las diferencias que existen entre los procesos de pensamiento y uso del lenguaje humano, y el funcionamiento de una computadora.

Estos problemas se manifiestan, sobre todo, al momento de redactar un *abstract*, ya que el analista puede expresar una frase de distintas formas sin variar su contenido semántico. Por ejemplo:

/FACULTADES DE LA DIRECCIÓN GENERAL JURÍDICA/  
INTERVENCIÓN EN LOS PROCEDIMIENTOS DE IMPUGNACIÓN/.

/ATRIBUCIONES DE LA DIRECCIÓN GENERAL JURÍDICA/  
INTERVENIR EN LOS PROCESOS EN QUE SE IMPUGNEN  
RESOLUCIONES DE LA SECRETARÍA/

Se puede afirmar que existen dos clases de problemas lingüísticos: los gramaticales y los semánticos (esta división es meramente para fines de exposición, pues la relación entre ambos es estrecha).

En los problemas gramaticales se ubican las cuestiones de sintaxis y morfología; en los semánticos se tratan las cuestiones de sinonimia, analogía y polisemia. Enseguida se expondrá, brevemente, cada uno de ellos.

#### a. Gramaticales (sintaxis y morfología)

El estudio de toda gramática comprende tanto la forma (morfología) como la función (sintaxis) que guardan entre sí las palabras en una frase.

En tanto no existen restricciones en el uso del lenguaje, es posible redactar una misma idea utilizando palabras que morfológicamente son diferentes, sin modificar su contenido semántico. Por ejemplo: la idea de *creación de una empresa*, puede ser redactada de varias maneras:

UNA EMPRESA FUE CREADA  
UNA EMPRESA SE CREÓ  
SE CREARÁ UNA EMPRESA  
SERÁ CREADA UNA EMPRESA  
SE HA CREADO UNA EMPRESA

El problema para el sistema consiste en que debe ser capaz de recuperar frases o ideas que aunque morfológicamente sean diferentes, semánticamente sean iguales.

Al respecto es necesario, pues, que el analista conozca la función que tiene cada una de las palabras dentro de una frase y ser capaz de expresar, sin alterar su contenido semántico, las ideas que se encuentren en el documento de origen.

#### b. Semánticos

La semántica es la materia que se encarga del estudio del significado de los signos (no lingüísticos y lingüísticos). Los signos que nos interesan, por supuesto, son los lingüísticos. Y los problemas que presentan son:

##### i) La sinonimia

Consiste en la existencia de dos o más palabras que tienen el mismo significado o contenido semántico. Por ejemplo:

NIÑEZ-INFANCIA  
PROPIETARIO-DUEÑO

##### ii) La analogía

Indica la posibilidad de referirse a un mismo concepto con palabras que tienen un contenido semántico similar. Por ejemplo, la misma idea de *creación de una empresa*, puede expresarse también como:

CONSTITUCIÓN DE UNA EMPRESA  
ESTABLECIMIENTO DE UNA EMPRESA  
FORMACIÓN DE UNA EMPRESA

### iii) La polisemia

Expresa la existencia de una palabra con más de un significado. Por ejemplo:

BANCO —DE ARENA  
—PARA SENTARSE  
—PARA DEPOSITAR DINERO  
—DE SANGRE

En el aspecto anterior, el problema tiene, en nuestro sistema, dos características que lo hacen aún más complejo.

En primer lugar, debe hacerse notar que, en principio, el lenguaje jurídico es, en la mayoría de los casos, morfológicamente semejante al lenguaje ordinario, pero semánticamente distinto, y en donde uno de los sentidos tiene una significación especial. Por ejemplo: el término *Derecho* tiene un sentido especial y diverso al que tiene en el lenguaje común.

En segundo lugar, el sistema, por sus características técnicas, no permite el uso de acentos, dando lugar a que muchas palabras que no son polisémicas, aparezcan como tales. Por ejemplo: secretaría es una dependencia de gobierno, secretaria es una persona. Sin embargo ambas se escriben con mayúsculas SECRETARIA, sin permitir en primera instancia determinar a cuál de los dos sentidos se refiere.

Para resolver estos problemas planteados se han desarrollado dos instrumentos lingüísticos: el *léxico* y el *thesaurus*. A continuación explicaremos en qué consisten y los problemas específicos que resuelven.

## 2. El léxico

El *léxico* permite la comunicación entre el usuario y la computadora. De su estructura depende que se logre una localización rápida y pertinente de los documentos con la menor cantidad de *ruido* o *silencio* posible.

Tomando como punto de partida el *léxico* del IRETIJ,<sup>20</sup> nuestro *léxico* se encarga de organizar las palabras no nulas<sup>21</sup> contenidas en el *corpus* con base en criterios morfológico-semánticos, principalmente; y sintácticos cuando es necesario acudir al contexto de la ficha de análisis, para ubicar el significado de la palabra.

<sup>20</sup> Cfr. Mazet Gou y Claude Bernard. "Aspects linguistiques du traitement automatique de l'information juridique" en *L'informatique juridique: du reue a l'instrument*, Ed. Les Editions Themis inc. — C. N. R. C. 1975.

<sup>21</sup> "Palabras nulas" son aquellas que son irrelevantes en la interrogación, por ejemplo conjunciones, preposiciones, artículos, etc.

Con el *léxico* se pretende resolver los problemas lingüísticos de sintaxis, sinonimia y algunos aspectos de polisemia.

En principio y en razón de la estructura morfológica de las palabras, se puede sostener que cuando éstas tienen la misma raíz presentan significados iguales o semejantes. Por ello, el léxico se estructura en nociones. Una noción es la agrupación de las palabras que tienen la misma raíz. Esto con el objeto de que al momento de la interrogación, y a través de un comando,<sup>22</sup> sea posible obtener toda la información recuperable por las palabras contenidas en la noción.

Cada noción admite hasta 10 subnociones. En cada una de ellas se colocan las palabras que semánticamente se consideran equivalentes. Por regla general, los accidentes gramaticales de género y número (singular, plural, masculino y femenino) de una misma palabra se consideran como tales y se colocan en la misma subnoción. Lo mismo sucede con el infinitivo, las formas verbales, el participio pasado y el sustantivo que indica la acción del mismo verbo.

La finalidad es resolver el problema de sintaxis, ya que sin importar que se interrogue por “creación de una sociedad” o por “crear una sociedad”, se asegura la recuperación de la información. En realidad la palabra “crear” o “creación” realizan la misma función semántica y gramatical dentro de cada una de las frases.

A continuación presentamos el desarrollo de varias nociones con sus subnociones:

- 323680 CONTRATO  
CONTRATOS
- 323681 CONTRATEN  
CONTRATADOS  
CONTRATAR  
CONTRATADO  
CONTRATARLO  
CONTRATACIÓN  
CONTRATACIONES
- 323682 CONTRATISTA  
CONTRATISTAS
- 323683 CONTRATO-LEY  
CONTRATOS-LEY
- 323684 SUBCONTRATACIÓN  
SUBCONTRATACIONES

<sup>22</sup> Instrucción dada a la computadora para que realice cierta operación, en este caso traer todos los documentos en donde existan las palabras contenidas en la noción.

- 323685 CONTRATANTE  
CONTRATANTES
- 323686 SUB-CONTRATISTA  
SUBCONTRATISTAS  
SUBCONTRATISTA  
SUB-CONTRATISTAS
- 323687 CONTRACTUAL  
CONTRACTUALES
- 323688 EXTRA-CONTRACTUALES  
EXTRA-CONTRACTUAL  
EXTRACONTRACTUAL  
EXTRACONTRACTUALES
- 327950 INFORMAR  
INFORMACIÓN  
INFORMACIONES
- 327951 INFORME  
INFORMES
- 327952 INFORMANTE  
INFORMANTES
- 327953 INFORMATIVA  
INFORMATIVAS
- 327910 INDUSTRIA  
INDUSTRIAS
- 327911 INDUSTRIAL  
INDUSTRIALES
- 327912 INDUSTRIALIZAR  
INDUSTRIALIZACIÓN  
INDUSTRIALIZADA  
INDUSTRIALIZADO  
INDUSTRIALIZADOS
- 327913 INDUSTRIALIZADORA  
INDUSTRIALIZADORAS

Las reglas generales que enunciamos anteriormente admiten una serie de excepciones que mencionamos a continuación:

Para resolver el problema de la sinonimia se colocan en una misma subnoción palabras que, aunque no tengan la misma raíz, poseen exactamente el mismo contenido semántico. Ejemplo:

322320 PANTEÓN  
PANTEONES  
CEMENTERIO  
CEMENTERIOS

Un segundo caso es el denominado “ruptura semántica total”. Esto sucede cuando una palabra, aun teniendo la misma raíz, no comparte ningún sentido con el resto de las palabras de la noción. Ejemplo: Las palabras brazo y braza provienen del latín *brachium*. Es notorio que en la actualidad no tienen relación semántica. Por esta razón estas palabras, aunque tienen la misma raíz, deben tratarse en nociones distintas. En este caso la palabra debe aislarse en una noción distinta.

Un tercer caso, cuyo origen está en la polisemia, es el que se denomina “ruptura semántica parcial”. Dentro de este rubro es posible identificar dos situaciones. La primera, cuando una o más de las palabras de la noción, además de compartir el significado común del resto de la noción, tiene otro u otros sentidos distintos. En estos casos, la palabra polisémica debe aislarse en una subnoción, dentro de la noción general. Ejemplo:

322320 PRISIÓN  
PRISIONES

322321 PRISIONERO  
PRISIONERA

322322 PRESO  
PRESOS

322323 PRESA  
PRESAS

Como puede observarse, la palabra “presa” tiene tres sentidos. El primero corresponde al de la idea de “prisión”, el segundo al de una obra artificial para almacenar agua, el tercero se refiere a una “presa” de caza. Por lo anterior, la palabra en cuestión se aísla en una subnoción distinta a la de “preso”, que es la que le correspondería normalmente. Esto evita que, si alguien interroga por “preso”, aparecieran documentos que no corresponden al sentido de la interrogación. Por otro lado, permite que, si se interroga por el comando que comprende toda la noción, no se pierde el sentido de “prisionera”. El riesgo de ruido existe, pero generalmente se elimina con el contexto de la interrogación.

La segunda situación es cuando existe una palabra con dos o más sentidos, los cuales se encuentran desarrollados previamente en el léxico en dos nociones distintas, pero que tienen la misma raíz (véase ruptura semántica total). Este problema lo hemos denominado “polisemia puen-



te”, ya que la palabra puede ubicarse en cualquiera de las dos nociones. Un ejemplo aclara lo anterior:

La palabra CONSTITUCIÓN puede ubicarse en la noción de CONSTITUIR o en la de CONSTITUCIONAL. Agrupar la palabra en las dos nociones ocasiona un serio problema técnico, por lo que se decidió que este tipo de palabra, cuyos dos sentidos existen en el *léxico*, debe aislarse en una noción distinta, y para cada una de las tres nociones señalarse el aislamiento semántico. La finalidad de lo anterior es permitir al usuario, en el momento de la interrogación, darse cuenta de que la palabra ha sido aislada. Por ejemplo:

322660  
COBRAR

322760  
COBRE

322650  
CUPRÍFERO  
CUPRÍFEROS

322651  
COBRADOR  
COBRADORES

322662  
COBRANZA  
COBRANZAS

La regla de la “polisemia puente” puede tener excepciones en razón al contexto sintáctico en que ésta se utiliza. Consideremos por ejemplo la palabra “importar”, la cual podría colocarse en las nociones de “importación” o de “importante”, sin embargo en una muestra de 1123 palabras, en más del 95% apareció en el sentido de importación. Por esta razón se prefirió dejarla con “importación” pues el riesgo de silencio es reducido.

El último problema es el de la “polisemia paralela”. En este caso, todas las palabras de la noción comparten más de un sentido idéntico, por lo cual ninguna de las palabras debe aislarse. Ejemplo:

322650 CIRCULACIÓN  
CIRCULACIONES

322651 CIRCULAR  
CIRCULARES  
CIRCULANDO

322652 CÍRCULO  
CÍRCULOS

Después de enumerar las excepciones a las reglas generales de construcción del *léxico*, a continuación se mencionan algunas reglas prácticas de formación de nociones.

Cuando aparece una palabra nueva, debe consultarse el listado de palabras por orden alfabético, con el objeto de saber si ya existe una noción que contenga la raíz de dicha palabra nueva. Una vez localizada la noción, debe colocarse en la subnoción que le corresponde. Debe tenerse cuidado en dos sentidos:

a) Si aparece una palabra nueva debe buscarse en el listado de palabras nuevas con los prefijos y sufijos que admita.

b) Si aparece una palabra que contenga un prefijo o sufijo, debe buscarse la palabra sin éstos.

Lo anterior tiene el objeto de que no se generen nociones nuevas con palabras cuya raíz ya existe. Ejemplo: es fácil crear una nueva noción con CAPACIDAD si no se tiene el cuidado de buscar en el listado alfabético las posibles derivaciones de la palabra CAPACIDAD, ya que es posible que previamente haya aparecido la palabra INCAPACIDAD.

En general los adverbios, adjetivos y sustantivos derivados de una misma raíz deben colocarse en subnociones distintas.

En caso de que una palabra sea al mismo tiempo forma verbal y sustantivo, la palabra debe dejarse con las formas verbales y no separarse en una subnoción distinta.

En caso de que aparezcan adjetivos y algunos sustantivos nuevos, debe añadirseles automáticamente el singular, plural, masculino y femenino que correspondan, excepto cuando una de las derivaciones sea una palabra polisémica. En este caso debe esperarse a que aparezca como palabra nueva en el *corpus*. Por ejemplo: si aparece la palabra ALTA debe añadirse ALTAS, pero no ALTO y ALTOS, porque estas dos son polisémicas y deben ubicarse en una subnoción distinta dentro de la misma noción. Lo mismo sucede si aparece CABALLO; debe añadirse CABALLOS y esperar a que aparezca YEGUA, pero ésta debe colocarse en una noción distinta, ya que pertenece a otra raíz.

Asimismo, cuando aparezca una forma verbal, se debe añadir automáticamente el infinitivo, los participios pasados y el sustantivo que indique la acción del verbo. Si aparece el sustantivo, debe añadirse el verbo.

Los gerundios, aún siendo formas verbales, se colocan en una subnoción distinta.

Para tratar los nombres propios tanto de personas como de poblaciones se utiliza una subnoción por cada uno sin importar lo que se encuentre en las otras subnociones. Lo anterior porque cada una de ellas es independiente de las otras y no hay forma de agruparlas de acuerdo con criterios semánticos o morfológicos. Ejemplo:

342530 SANTA-ROSA  
342531 TAMASULA  
342532 CIUDAD-GUZMÁN  
342533 ELOTA

- 342534 MIGUEL-HIDALGO  
MIGUEL-HIDALGO-Y-COSTILLA
- 342535 GUANAJUATO  
GUANAJUATENSE
- 342536 FRANCIA  
FRANCÉS  
FRANCESA  
FRANCESES  
FRANCESAS
- 342537 GUASAVE
- 342538 GUSTAVO-A-MADERO
- 342539 LA-BARCA

Como podrá observarse, los nombres propios compuestos de más de una palabra se unen con guiones con el objeto de que el sistema los trate como una unidad. Cuando existen varias formas en que pudieran aparecer se colocan todas dentro de una misma subnoción a fin de asegurar que se recupere toda la información sobre ese asunto. Ejemplo: MIGUEL-HIDALGO; MIGUEL-HIDALGO-Y-COSTILLA.

Cuando aparece algún país o entidad federativa los gentilicios se colocan dentro de la misma subnoción.

Para tratar los números tanto arábigos como romanos, se ha asignado a cada uno una subnoción especial. Para el caso de arábigos hasta el 5 000 y en el caso de los romanos hasta el "L". En este último caso éstos se deben escribir siempre entre " " para evitar confundirlos con las letras.

Existen otra serie de casos especiales que no mencionamos. En última instancia lo que determina su colocación es la posibilidad de ruido o silencio que pueden producir. En todos los casos se debe preferir el ruido al silencio.

### 3. *El thesaurus*

Debemos advertir, en primer lugar, que con este nombre se designa generalmente las listas de descriptores empleados en los sistemas de lenguaje cerrado (véase B.1.). En el sistema UNAM-JURE nos referimos al *thesaurus* para denominar al instrumento lingüístico que tiene por objeto la reagrupación de las palabras contenidas en el *léxico* con aquellas que comparten un mismo sentido semántico (analogía y antonimia).

La finalidad del *thesaurus* es auxiliar al usuario a resolver los problemas de analogía, antonimia y de aislamiento semántico en caso de polisemia puente (véase C.2.) al momento de interrogar.

En efecto, hemos explicado que en el léxico se organizan las palabras de acuerdo con criterios morfológico-semánticos y, en algunos casos, sintácticos; sin embargo, las nociones de aquél son independientes unas de otras, por lo que el usuario no puede relacionarlas con el fin de ampliar o limitar el contenido de su interrogación. El *thesaurus* permite esta relación señalándole las palabras análogas y antónimas de aquélla, a partir de la cual realiza su pregunta; si fuera polisémica le señalará sus sentidos y además le permite conocer y reintegrar, en su caso, las palabras que han sido aisladas semánticamente.

Para la construcción del *thesaurus* en UNAM-JURE se han considerado dos etapas. La primera supone su elaboración en forma manual, y la segunda su implementación automática. En ambos casos, la consulta al *thesaurus* es optativa para el usuario.

El *thesaurus* se encuentra actualmente en su primera etapa manual, de ella nos ocupamos a continuación.

El *thesaurus* se ha construido a partir de cada una de las nociones del léxico. A cada una de ellas se le relaciona con las palabras análogas y antónimas. Ejemplo:

	NOCIÓN	ANALOGÍA
320290	ACUSACIÓN ACUSAR	DENUNCIA QUERRELLA INCUPLPAR IMPUTAR DELACIÓN INCRIMINAR
328530	JORNAL JORNALLES	SALARIO SUELDO RETRIBUCIÓN PAGA REMUNERACIÓN
328531	JORNALERO JORNALEROS	BRACERO TRABAJADOR OBRERO ARTESANO ASALARIADO PEÓN LABRIEGO LABRADOR CAMPEÑINOS

En la primera etapa hemos considerado todos los análogos posibles, tanto desde el punto de vista del lenguaje ordinario, como del de técnica jurídica, con el fin de ampliar lo más posible la utilidad del *thesaurus*.

Se han relacionado asimismo cada noción con sus antónimos cuando no son morfológicamente semejantes. Lo anterior tiene sentido si se considera que en la redacción basta una negación antepuesta a un sustantivo para que su antónimo le sea semánticamente igual, por ejemplo:

### NO PERMITIDO = PROHIBIDO

En el caso de palabras polisémicas tanto los análogos como los antónimos se organizan en capas semánticas. A cada capa corresponde un sentido específico, en donde análogos y antónimos están agrupados dependiendo del grado de semejanza que tengan con la noción. Ejemplo:

	NOCIÓN	ANALOGÍA
320120	ACCIÓN ACCIONES	1) TÍTULO VALOR, BONO, CERTIFICADO 2) HECHO 3) VÍA, DEMANDA, PRETENSIÓN

En muchas ocasiones, alguna o algunas de las subnociones de una noción contienen palabras polisémicas cuyos sentidos no los comparten el resto de las palabras de la noción. A fin de diferenciar las capas semánticas se numeran y se indican a cuáles corresponde cada subnoción. Por ejemplo:

	NOCIÓN		
328360	INTRASMISIBILIDAD	1	1) INALIENABLE. INTRASPASABLE PERSONAL. PRIVATIVO. INDIVIDUAL
328362	TRANSMISIÓN TRANSMISIONES	2, 3, 4,	2) TRASPASO, TRANSFERENCIA. TRASLADO, CESIÓN. ENAJENABLE
328362	RETRANSMISIÓN RETRANSMISIONES	1, 4	3) MECANISMO. COMUNICACIÓN. EMBRAGUE
328363	INTRASMISIBLE INTRASMISIBLES	1	4) DIFUSIÓN, EMISIÓN.

328364	TRANSMISIBLE TRANSMISIBLES	2, 5	5) RADIACIÓN, PROGRAMA INFECCIOSO, CONTAGIOSO
--------	-------------------------------	------	--

Una segunda función del *thesaurus*, como ya explicamos, es relacionar las palabras que siendo de la misma raíz, han sido separadas bien un caso de polisemia puente, bien que la palabra ha sido aislada porque tiene un sentido diverso y especial al del resto de la noción. Del primer caso:

NOCIÓN	CUIDADO CON
325310	CONSTITUCIÓN CONSTITUCIO- NES
323520	CONSTITUCIONAL CONSTITUCIONALES
323521	CONSTITUCIONALISTA CONSTITUCIONALISTAS
323522	CONSTITUCIONALIDAD
323523	INCONSTITUCIONALIDAD INCONSTITUCIONALES
323524	CONSTITUYENTE CONSTITUYENTES
323530	CONSTITUTIVA CONSTITUTIVAS CONSTITUTIVO CONSTITUTIVOS
323540	CONSTITUIR CONSTITUIDOS CONSTITUIDAS CONSTITUYEN CONSTITUYE

Del Segundo:

NOCIÓN	CUIDADO CON
320120	ACCIÓN ACCIONES
320140	ACCIONARIA ACCIONARIAS ACCIONARIO ACCIONARIOS
320141	ACCIONISTA ACCIONISTAS
320130	ACCIONAMIENTO ACCIONAMIENTOS

La función del *thesaurus* en este aspecto es evitar el riesgo de *silencio* relacionando la palabra puente con las otras nociones de las que ha sido separada.

## 64 EL SISTEMA UNAM-JURE: UN BANCO DE DATOS LEGISLATIVOS

En realidad nuestro *thesaurus* está apenas implementado en su primera etapa. Falta aún mucho por investigar a partir del contexto en que se encuentren las palabras para de ahí deducir las relaciones que establecen entre sí y las posibles relaciones significativas al momento de interrogar.

En última instancia nuestra idea final es lograr un solo instrumento lingüístico que cumpla, a la vez, las funciones que actualmente desarrollan el *léxico* y el *thesaurus*.