

ISSN: 0213-1854

## Documentación, edición y traducción: Tesauro de tipografía y edición español-inglés

M. MARCOS ALDÓN/J. PARDO VICARIO  
*Univ. de Córdoba/Univ. de Castilla La Mancha*

Fecha de recepción: 28 de mayo de 2009

Fecha de aceptación: 11 de junio de 2009

**Resumen:** En el presente artículo se analiza la elaboración de un tesauro de tipografía y su correspondiente traducción al inglés. La necesidad de una obra de estas características se justifica por la ausencia de un lenguaje controlado que recoja la terminología existente en lengua española en este ámbito. Se ha desarrollado un tesauro especializado, multidisciplinar y bilingüe con estructura alfabética y sistemática. Para ello se ha contemplado la norma UNE 50-106-90. El proceso de traducción ha seguido una metodología en tres pasos usando obras de referencia en la LO, en la LT y bilingües español-inglés.

**Palabras clave:** Traducción, Lenguajes documentales, Tipografía.

**Abstract:** A study about construction of a Typography Thesaurus and its translation into English is presented. A work of this type is needed due to the non-existence of a controlled language dealing with typographic terminology in Spanish. A specialized, multidisciplinary, bilingual Thesaurus with a systematic and alphabetical structure has been developed. It has been constructed to conform to UNE standard 50-106-90. Translation process has been carried out by means of a three-step method, using reference works both in the SL and the TL as well as bilingual Spanish-English dictionaries.

**Key words:** Translation, Documentary languages, Typography.

### Introducción

La preocupación informacional sobre los lenguajes controlados tiene una amplia repercusión en el campo de los lenguajes controlados<sup>1</sup>, es obvio si pensamos que ese

---

<sup>1</sup> LAGUENS GARCÍA, J. L. "Tesauros y lenguajes controlados en Internet". *Anales de Documentación*, nº 9, 2006, pp. 105-121. Codina, L. "Evaluación de recursos digitales en línea: conceptos, indicadores y métodos". *Revista Española de Documentación Científica*. 2000, vol. 23, nº 1, pp. 9-44. CUEVA MARTÍN, A. "Acceso y utilización de tesauros en Internet". *Revista Española de Documentación Científica*. 1999, vol. 22, nº 4, pp. 531-540. GARCÍA CAMARERO, E.; GARCÍA MELERO, L. A. *La biblioteca digital*. Madrid: Arco/Libros, 2001, p. 381. MARTÍN GONZÁLEZ, J. C.; MERLO VEGA, J. A. "Las revistas electrónicas: características, fuentes de información y medios de acceso". *Anales de Documentación*. 2003, nº 6, pp. 155-186. RUSSELL, R. DAY, M. "HILT: High-Level Thesaurus project [en línea]: automated and manual approaches to the provision of thesauri and subject vocabularies". Ver. 1.0. 5 October 2001. Bath, UK: University of Bath. <<http://www.ukoln.ac.uk/hilt/interfaces/>>. VALLE BRACERO, A.; REY GUERRERO, A.; PÁEZ MANÁ, J. Y VALLE BRACERO, R. "Tesauros en HTML: un modelo de diseño y estructura para su consulta en la malla mundial (WWW)". *Revista Española de Documentación Científica*. 2000, vol. 3, nº 2, p. 159-178. AITCHISON, J., Y GILCHRIST, A.: *Thesaurus Construction, a Practical Manual*, London: ASLIB. 1987. CURRAS, E. *Manual de construcción y uso*, Madrid: Kaher II. 1998. CHAUMIER, J.:

es uno de sus principales objetivos: el control de la información a través de un sistema universal. El mundo de la edición no ha tenido tanta suerte, aunque su importancia en el proceso informacional hoy día es crucial frente a la expansión y desarrollo de la información en los nuevos entornos. Todo tipo de código y mensaje se ha editado a lo largo del tiempo para la comprensión de los mismos, encauzado por cierto tipo de especialistas que las sociedades poseen y que dedican sus esfuerzos al nacimiento, desarrollo, mantenimiento y difusión de la información, ya sea desde un plano teórico, práctico o mítico. Hoy día el proceso de globalización y su difusión han multiplicado los procesos de producción y difusión de la información y es por ello que es necesario introducir herramientas que auxilien ya a las existentes en el proceso de edición<sup>2</sup>, tanto a los profesionales como a los neófitos. Uno de los aspectos que se tratan en el proceso de tratamiento de los datos para su posterior difusión es su presentación gráfica, es por ello que en el marco de creación de un macrotesauro de edición documental basado en descriptores sintagmáticos y con enlaces de acceso y captura a ontologías OWL eran pertinente una serie de elementos que se describen en el presente artículo. Como tal elemento del proceso documental de la edición la tipografía comporta una vertiente artística que ha de complementarse con el rigor técnico. No se puede prescindir de ninguna de estas dos perspectivas sin correr el riesgo de mutilar el trabajo tipográfico. Al

---

*Análisis y lenguajes documentales: El tratamiento lingüístico de la información documental*, Barcelona: Mitre. 1986. GARCÍA GUTIERREZ, A.: *Lingüística documental: Aplicación a la documentación de la comunicación social*, Barcelona: Mitre. 1984. GIL URDICIÁN, B.: *Manual de lenguajes documentales*, Madrid: Noesis. 1996. Maniez, J.: *Los lenguajes documentales y de clasificación: Concepción, construcción y utilización en los sistemas documentales*, Madrid: Fundación Germán Sánchez Ruipérez. 1992. MCLLWAIN, I. C.: *Guía para uso de la CDLJ: Una guía introductoria para el uso y aplicación de la CDU*, Madrid: AENOR. 1994. Slype, G. V.: *Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales*, Madrid: Fundación Germán Sánchez Ruipérez. 1991. UNE 50-106-1990 (1990): *Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües*, Madrid: AENOR. Equivalente a la Norma ISO 2788-1986.

<sup>2</sup> Asociación Española de Normalización y Certificación. *Directrices para el establecimiento y desarrollo de tesauros monolingües*. UNE 50-106-90. Madrid: AENOR, 1990. CURRÁS, E. *Ontologías, taxonomía y tesauros. Manual de construcción y uso*. Trea: Gijón, 2005. ISBN 84-9704-157-7. Gómez Mascaraque, T. *Breve glosario trilingüe sobre edición, producción editorial, derechos y contratos* [pdf]. 1996. Disponible en: <[http://www.acta.es/articulos\\_mf/03103.pdf](http://www.acta.es/articulos_mf/03103.pdf)> [Consulta: 2 junio 2008]. HOFFMAN, P. *Typography & Graphic Design Thesaurus* [en línea]. Ver. 1.1. 31 mayo 2006. Disponible en: <<http://hoffmancommappaul.com/tgdt/>> [Consulta: 17 mayo 2008]. KANE, J. *Manual de Tipografía*. Traducido por M. Dávila. Barcelona: Gustavo Gili, 2005. Traducción de: *A type primer*. ISBN 84-252-1980-9. Martínez de Sousa, J. *Diccionario de edición, tipografía y artes gráficas*. Trea: Gijón, 2001. ISBN 84-95178-96-6. MARTÍNEZ DE SOUSA, J. *Manual de edición y autoedición*. Madrid: Pirámide, 1994. 2ª edición. ISBN 84-368-1931-4. Merriam-Webster, Inc. *Merriam-Webster Online Dictionary* [en línea]. 2005. <<http://www.merriam-webster.com>> [Consulta: 27 mayo 2008]. RENNER, P. *El arte de la tipografía*. Traducido por J.A. Cifuentes y E. Monzó. Valencia: Campgràfic, 2000. Traducción de: *Die Kunst der Typografie* ISBN 84-931677-1-1. Schopp, Jürgen F. "Typography and Layout as a Translation Problem". En: *XVI Congreso Mundial de la Federación Internacional de Traductores*. Vancouver, Federación Internacional de Traductores, 2002. p. 271-275. Sheridan, I. (compilador). *Multilingual Glossary for Art Librarians* [en línea]. 2ª ed. revisada y ampliada. 2006. Disponible en: <<http://www.ifla.org/VII/s30/pub/mg1.htm#spanish>> [Consulta: 4 junio 2008].

igual que muchos otros ámbitos del conocimiento, la tipografía dispone de un extenso vocabulario generado a partir de las distintas facetas que conforman su actividad. Esto lleva a destacar la importancia y la necesidad de un lenguaje controlado que permita un tratamiento sistemático del vocabulario con vistas a la indización y la recuperación terminológica en este ámbito, puesto que actualmente no existe ningún tesoro en lengua española que trate esta materia.

No se ha dejado pasar la oportunidad de conectar esta propuesta de lenguaje documental con la lengua inglesa. Además, parece lógico hacerlo cuando la comunidad angloparlante sí cuenta con varios tesauros dedicados a esta materia. Por otro lado, los aspectos tipográficos son frecuentemente menospreciados en las traducciones. Schopp (2002) señala que actualmente los traductores están asumiendo tareas que antes realizaban expertos en tipografía, así que para desempeñar dicha labor por su cuenta necesitarán adquirir conocimientos básicos de tipografía. Debido a la creencia errónea de que es un elemento textual irrelevante o fácilmente transferible, los resultados a veces son poco profesionales. Este texto pretende, por tanto, ayudar también a los traductores en esta labor interlingüística en la medida de sus posibilidades, con ello se mejora la calidad informativa y, por ende, la calidad del producto de información traductológica como producto global, que puede emplearse tanto en el proceso teórico de información, como en el práctico de documentación, como en el teórico-práctico de la localización traductológica.

### Metodología de la Calidad en el Tesoro

Presentadas las repercusiones sociales de la escasa calidad de la información<sup>3</sup> debemos preguntarnos ¿cómo deberían aproximarse los usuarios a los asuntos relacionados con la calidad de la información en tipografía y edición? Empezaremos detallando algunos de los conceptos fundamentales para la comprensión de la calidad de la información. Con frecuencia, los términos "datos" e "información" se emplean como sinónimos. Ya en la práctica, los usuarios establecen una

---

<sup>3</sup> BAILEY, R., *Human Error in Computer Systems*. Prentice Hall, Englewood Cliffs, 1983. BALLOU, D. y G. TAYL, "Managerial Issues in Data Quality." *Proceedings of The 1996 Conference on Information Quality*. Cambridge, MA, pp. 186-206, 1996. CODD, E. E., "A Relational Model of Data for Large Shared Data Banks," *Communications of the ACM*, 13(6), 1970, pp. 377-387. COREY, D., "Data Quality Improvement in the Military Health Services Systems and the U.S. Army Medical Department", *Proceedings of the 1997 Conference on Information Quality*. Cambridge, MA, pp. 37-62, 1997. COREY, D., L. COBLER, K. HAYNES y R. WALKER, "Data Quality Assurance Activities in the Militan' Health Services System," *Proceedings of the 1996 Conference on Information Quality*. Cambridge, MA, pp. 127-153, 1996. CRC, *Information Quality Assessment Survey: Administrator's Guide*. Cambridge Research Group, Cambridge, MA, 1997. KAHN, B. K., D. M. STRONG y R. Y. WANG, "Information Quality Benchmarks: Product and Service Performance," *Communications of the ACM*. KNIGHT, B., "The Data Pollution Problem," *Computernworld*, September 28, pp. 81-82, 1992. KON, H. B., J. LEE y R. Y. WANG, *A Process View of Data Quality (No. TDQM-93-01)*. *Total Data Quality Management (TDQM) Research Program*, MIT Sloan School of Management, Cambridge, MA, 1993. KWAN, S. y D. ROTEM. "Analysis and Tradeoff Between Data Accuracy and Performance of Databases." *Proceedings of the Conference on Scientific and Statistical Database Management*. Germantown. 1992.

diferenciación intuitiva de ambos al describir la información como unos datos que ya han sido procesados. Salvo indicación contraria, ambos términos se emplean de forma intercambiable.

Tal y como se muestra en la figura<sup>4</sup>

	Fabricación de productos	Producción de información
Entrada	Materia prima	Datos primarios
Proceso	Línea de ensamblaje	Sistema de información
Salida	Productos físicos	Productos de información

Fuente: *Information Technology in Action: Trends and Perspectives*

existen puntos análogos entre los temas relativos a la calidad en la fabricación de un producto y los aplicables a la generación de información. La fabricación de productos se puede ver como un sistema de procesamiento que actúa sobre la materia prima con el fin de fabricar productos físicos. De forma análoga, la creación de información se puede entender como un sistema de procesamiento que actúa sobre datos primarios con el fin de generar un producto de información

Por lo general, las actividades relacionadas con la fabricación de datos o información constan de dos partes<sup>5</sup>. La primera vendría dada por el diseño y desarrollo del producto. De forma habitual, estas tareas las lleva a cabo un grupo dedicado al desarrollo de productos con la colaboración de otras áreas funcionales. La segunda parte sería la de fabricación y distribución del producto. La calidad del producto va en función de ambas partes: la maquinaria de fabricación y los métodos aplicados en ese instante. Ambos conceptos guardan relación con los sistemas de generación de información tal como se plantea acto seguido. Imaginemos que el proceso de edición y tipografía, como parte inexcusable del mismo, no tuviera lugar. Esta confusión, por lo general ya ocurre, implica una desafortunada desorganización que implica una dificultad mayor de la difusión correcta y de su asimilación como conocimiento.

<sup>4</sup> WANG, R. y KON, H. B. "Toward Total Data Quality Management (TDQM)". *Information Technology in Action: Trends and Perspectives*, R. Y. WANG, Ed. 1993, Prentice Hall, 1993, Englewoods Cliffs, NJ, 1993.

<sup>5</sup> LEVITIN, A. V y T. REDMAN, "A Model of the Data (Life) Cycles with Application to Quality," *Information and Software Technology*, 35(4), 1995, p. 7. LIEPINS, G. E., "Sound Data Are a Sound Investment," *Quality Progress*, 22(9), 1989, pp. 61-64. MADNICK, S. y R. Y. WANG. "Integrating Disparate Databases for Composite Answers," *Proceedings of the 21st Hawaii International Conference on System Sciences*. Hawaii, pp. 583-592, 1988. PAUTKE, R. W. y T. C. REDMAN, "Techniques to Control and Improve Quality of Data in Large Databases." *Proceedings of Statistics Canada Symposium 90*. Canada, pp. 319-333, 1990.

Al referirnos, por tanto, a los sistemas de generación de la información, nos remitimos a un sistema que genera productos de información<sup>6</sup>. El concepto relativo a los productos de información hace hincapié en el hecho de que la salida de información de un sistema de este tipo presenta un valor transferible al consumidor, ya sea externo o interno. En este sentido, concretamos cuatro funciones en el marco de nuestro sistema de generación de información:

-Los proveedores de la información son aquellos que crean o recopilan los datos aplicables al producto de información (Productores de información general, Documentalistas, Traductores-correctores).

-Los encargados de la información son aquellos que diseñan, desarrollan o mantienen al día los datos y la infraestructura de sistemas aplicables al producto de información (Participantes del proceso de mantenimiento del programa y del tesaurus).

-Los consumidores de la información son aquellos que la emplean en su trabajo (Correctores-Tipógrafos-Traductores correctores-Documentalistas editores).

Cada función está vinculada a un proceso o tarea: (1) los proveedores con los procesos de producción de los datos; (2) los encargados con el almacenamiento, la conservación y la seguridad de los datos y, (3) los consumidores con los procesos de aprovechamiento de los datos, que bien pueden incluir procesos complementarios de integración y ampliación.

Hay que tener en cuenta todo el ciclo de desarrollo del sistema de información que se ha puesto en marcha en este tesaurus, abarcando desde su diseño e implantación hasta su puesta a disposición de los usuarios, tal como se muestra en la figura.

Puesto que la información generada por el sistema existente ha ido en función de todas las actividades completadas en las diversas fases del ciclo de desarrollo del

---

<sup>6</sup> IBM "Global Services Wins Top Honor For Its Knowledge Management Innovations," IBM, 1998. <http://www.ibm.com/services/pressrel/pr.889566539.html>. Jensen, M. C. y W. H. Meckling, "Specific and General Knowledge, and Organizational Structure" *Journal of Applied Corporate Finance*, 8(2), 1995, pp. 4-18. KELLER, R., y R. R. CHINTA, "International Technology Transfer: Strategies for Success," *Academy of Management Executive*, 4(2), 1990, pp. 33-43. KNIGHTS, D., MURRAY, E. y H. WILLMOT, "Networking as Knowledge Work: A Study of Strategic Interorganizational Development in the Financial Services Industry," *Journal of Management Studies*, 30(6), 1993, pp. 975-995. Polanyi, M., *The Tacit Dimension*. Doubleday, Garden City, NY, 1966. POPPER, K. P., *Objective Knowledge*. Clarendon Press, Oxford, 1972. Potter, M. E., *Competitive Advantage*. Free Press, New York, 1985. Prahalad, C. K., y G. KAMEL, "The Core Competence of the Corporation," *Harvard Business Review*, (May-June), 1990, pp. 79-91. WARIA, *WARIA and Giga Information Group Announce the Finalists of 1998 Giga Excellence Awards for Excellence in Document /Knowledge Management and Workflow/Process Management in North America*. WARIA, 1998 <http://www.waria.com/lawards98.html>. WINTER, S. G., "Knowledge and Competence as Strategic Assets." *The Competitive Challenge*, D. J. TEECE, ed., Ballinger, Cambridge, MA, 1987. SANDBERG, J. "At Thousands of Web Sites, Time Stands Still: Many Web Sites Need Updating," *The Wall Street Journal*, Marzo 11, 1997 p. B1.

sistema, es preciso prestar especial atención a cada fase en lo que respecta a cualquier posibilidad de mejora de la calidad de la información<sup>7</sup>.



Fuente: *Information Technology in Action: Trends and Perspectives*<sup>8</sup>

Entre los problemas de diseño solventados podemos apreciar algunos tan básicos como:

-La información incorporada en el diseño de sistema no coincidió con la exigida por el consumidor debido al análisis incorrecto de los requisitos (problema relativo al diseño).

-No se completaron los ensayos del soporte lógico empleado en el sistema, con lo que el funcionamiento del sistema de generación de información fue errático (problema relativo a la implantación).

-Los encargados de la información no dispusieron de la formación adecuada para el tratamiento de ésta en aspectos tales como la entrada y recuperación de los datos, resultando en datos corruptos (problema relativo a la puesta en marcha).

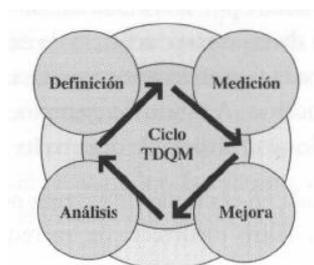
Para solventar estos problemas y otros análogos con los conceptos relacionados con el sistema de generación de información y su ciclo de desarrollo que sientan las bases para la gestión en forma de producto del tesoro empleamos técnicas aplicables al ciclo de Gestión de la Calidad Total de los Datos (TDQM) y al alcance de la calidad de la información<sup>9</sup>.

<sup>7</sup>LAUDON, K. C. "Data Quality and Due Process in Large Interorganizational Record Systems," *Communications of the ACM*, 29 (1), 1986, pp. 4-11. LEE, Y. W. "Learning by Solving DQ Problems: Managing DQ Knowledge," *Proceedings of INFORMS*, Washington, D.C., 1996. MADNICK, S. E., ed. *The Strategic Use of Information Technology*. 1987, Oxford University Press, New York. 206 pages. MADNICK, S. y R. Y. WANG. "Integrating Disparate Databases for Composite Answers," *Proceedings of the 21st Hawaii International Conference on System Sciences*. Hawaii, pp. 583-592, 1988. O'REILLY, C. A. I., "Variations in Decision Makers' Use of Information Sources: the Impact of Quality and Accessibility of Information," *Academy of Management Journal*, 4(25), 1982, pp. 756-771. ORR, K. "Data Quality and Systems Theory," *Proceedings of the 1996 Conference on Information Quality*. Cambridge, MA, pp. 1-15, 1996. ADLER, P. S., "When Knowledge is the Critical Resource, Knowledge Management Is the Critical Task," *IEEE Transactions on Engineering Management*, 6 (30), 1989, pp. 997-1015

<sup>8</sup> WANG, R. y KON, H. B. "Toward Total Data Quality Management (TDQM)". *Information Technology in Action: Trends and Perspectives*, R. Y. Wang, Ed. 1993, Prentice Hall, 1993, Englewoods Cliffs, NJ, 1993.

<sup>9</sup> ANGELES, P. A., *Dictionary of Philosophy*. Harper Perennial, New York, 1981. BALLOU, D. P. y H. L. PAZER, "Modeling Data and Process Quality in Multi-input, Multi-output Information Systems,"

En estas referencias se encuentra el ciclo de Deming<sup>10</sup>, muy extendido en la práctica, para la mejora de la calidad por medio de los conceptos inherentes a los verbos: planifica, realiza, revisa y actúa. Mediante la adaptación del ciclo de Deming, hemos podido desarrollar un ciclo de Gestión de la Calidad Total de los Datos (TDQM) tal como el que se muestra en la figura.



Fuente: *Communications of the ACM*<sup>11</sup>

En el elemento relativo a la definición del ciclo TDQM queda identificado el alcance de la calidad de la información (IQ). Con el elemento relativo a la medición se genera la métrica aplicable a la IQ. Con el elemento análisis se identifica el origen real de los problemas que afectan a la IQ y se calculan las repercusiones de las carencias en la calidad de la información. Por último, en el elemento mejora se incorporan los métodos para la mejora de la IQ. Éstos se aplican a lo largo del alcance del sistema y en función de los requisitos especificados por el usuario (editor, documentalista, corrector, traductor).

Al igual que la calidad del producto presenta múltiples dimensiones, la calidad de la información también tiene varias dimensiones. Entre las que se destacan con más frecuencia estarían la exactitud, el alcance pleno, la coherencia y la puntualidad. La selección de estas dimensiones se basa de forma primordial en la comprensión

---

*Management Science*, 31 (2), 1985, pp. 150-162. BENBASAT, I., D. K. GOLDSTEIN, y M. MEAD, "The Case Research Strategy in Studies of Information Systems," *Management Information Systems Quarterly (MISQ)*, 11 (3), 1987, pp. 369-386. BRODIE, M. L., "Data Quality in Information Systems," *Information and Management*, (3), 1980, pp. 245-258. LAUDON, K. C., "Data Quality and Due Process in Large Interorganizational Record Systems," *Communications of the ACM*, 29 (1), 1986, pp. 4-11. LEE, Y. W, D. M. STRONG, L. PIPINO, y R. Y. WANG, *Methodology-based Software Tool for Data Quality Management (No. TDQM-97-02)*. MITTDQM Research Program, 1997. WEBER, R. P., *Basic Content Analysis*. 2a ed. Sage, Newbury Park, CA, 1990. YIN, R. K., *Case Study Research: Design and Methods*. 2a ed. Sage, Thousand Oaks, CA, 1994.

<sup>10</sup> LEE, Y. W, D. M. STRONG, L. PIPINO, y R. Y. WANG, *Methodology-based Software Tool for Data Quality Management (No. TDQM-97-02)*. MITTDQM Research Program, 1997.

<sup>11</sup> WANG, R. Y, y D. M. STRONG, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems (JMIS)*, 12(4), 1996, pp. 5-34.

intuitiva<sup>12</sup>, la experiencia profesional<sup>13</sup> o la lectura de distintas obras de referencia<sup>14</sup>. Sin embargo, un estudio exhaustivo de las referencias demuestra que no existe un consenso general acerca de las dimensiones de la calidad de la información.

La exactitud, una dimensión clave según la mayoría de los estudios realizados en materia de calidad de la información. Si bien es cierto que el término “exactitud” presenta un carácter intuitivo, no existe una definición comúnmente aceptada en cuanto a su significado preciso. A modo de ejemplo, Strong<sup>15</sup> la define como “la veracidad de la información de salida”. Ballou y Pazer<sup>16</sup> la definen como lo que se consigue cuando “el valor registrado es conforme con el valor real.”

Sin duda, el concepto de la calidad de la información depende del uso que se haga de la información en la práctica. Es posible que lo que se entiende como buena información en un caso concreto (un usuario o una aplicación específica) sea insuficiente para otros casos.

Con la Calidad de la Información (IQ) contextual se destaca la obligación de que la calidad de la información se plantee en el contexto de la tarea en cuestión, es decir, que la información debe ser aplicable, puntual, completa y apropiada en cuanto a su volumen como para aportar un valor añadido. Con las dimensiones IQ relativas a la representatividad y la accesibilidad se hace hincapié en las funciones más importantes del sistema. Un sistema debe ser accesible a la par que seguro. Debe presentar la información de tal forma que se pueda interpretar y comprender con facilidad, así como quedar representada de forma concisa y constante<sup>17</sup>.

A lo largo del tiempo, se acumula información acerca de las causas de las incongruencias a partir de las evaluaciones de la exactitud de las distintas fuentes, lo que conduce a la mala reputación de las fuentes menos exactas (una mala reputación en lo que a la calidad se refiere también puede evolucionar a partir de una base menos amparada en los hechos). Según se extiende esa reputación de mala calidad

<sup>12</sup>BALLOU, D. P. y H. L. PAZER, “Modeling Data and Process Quality in Multi-input, Multi-output Information Systems,” *Management Science*, 31(2), 1985, pp. 150-162.

<sup>13</sup>FETTER, R. B., “Diagnosis Related Groups: Understanding Hospital Performance,” *Interfaces*, 21(1), 1991, pp. 626. YIN, R. K., *Case Study Research: Design and Methods*. 2a ed. Sage, Thousand Oaks, CA, 1994.

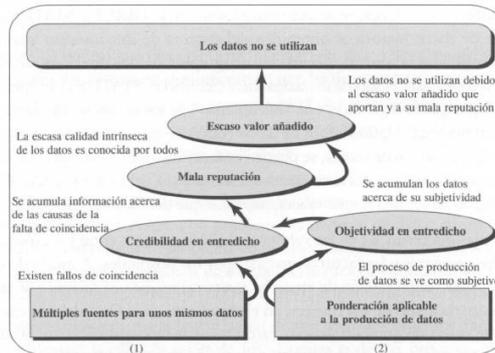
<sup>14</sup>STRONG, D. M., Y. W. LEE, y R. Y. WANG, “Data Quality in Context,” *Communications of the ACM*, 40(5), 1997, pp. 103-110.

<sup>15</sup>*Ibidem*.

<sup>16</sup>BALLOU, D. P. y H. L. PAZER, “Modeling Data and Process Quality in Multi-input, Multi-output Information Systems,” *Management Science*, 31(2), 1985, pp. 150-162.

<sup>17</sup>CRG, *Information Quality Assessment Survey: Administrator's Guide*. Cambridge Research Group, Cambridge, MA, 1997. CRG, *Integrity Analyzer: A Software Tool for TDQM*. Cambridge Research Group, Cambridge, MA, 1997. DEMING, E. W., *Out of the Crisis*. Center for Advanced Engineering Study, MIT, Cambridge, MA, 1986. JURAN, J. M., y F. M. GRYNA, *Quality Planning and Analysis*. 2a ed. McGraw Hill, New York, 1980. MILES, M. B., y A. M. HUBERMAN, *Qualitative Data Analysis: A Sourcebook of New Methods*. Sage Publications, Newbury Park, CA, 1984. MOREY, R. C., “Estimating and Improving the Quality of Information in the MIS,” *Communications of the ACM*, 25(5), 1982, pp. 337-342. WANG, R. Y., y D. M. STRONG, “Beyond Accuracy: What Data Quality Means to Data Consumers,” *Journal of Management Information Systems (JMIS)*, 12(4), 1996, pp. 5-34.

las fuentes de información afectadas se perciben como de escaso valor añadido para el tesoro, con lo que disminuye su uso (tendencia 1).



Fuente: *Communications of the ACM*<sup>18</sup>

La ponderación y la subjetividad aplicadas al proceso de producción de datos suponen otra causa habitual (tendencia 2). A modo de ejemplo, los datos codificados o interpretados se consideran de calidad inferior que los datos en bruto y sin interpretar. En principio, sólo aquellos que conocen los procesos de producción de datos son conscientes de estos problemas en potencia que se plantean como inquietudes relativas a la objetividad de los datos. Con el transcurso del tiempo, se acumula la información acerca de la naturaleza subjetiva de la producción de los datos, quedando en entredicho la credibilidad y la reputación de los mismos y, por tanto, reduciéndose el valor añadido para el usuario. El resultado final pasa por la reducción del uso de esos datos puestos en duda<sup>19</sup>.

Con el fin de garantizar una gestión eficaz, siempre es preciso medir y analizar la materia en cuestión. Ahora bien, siempre que queramos medir algo, tendremos que definir qué es lo que se va a medir. Es por ello por lo que la definición de la calidad de la información (IQ) resulta crucial para permitir que la información se maneje como un producto más. Asimismo, a lo largo de este proceso de definición, los participantes asumirán las tareas de identificación y solución de problemas más

<sup>18</sup>STRONG, D. M., Y. W. LEE, y R. Y. WANG, "Data Quality in Context," *Communications of the ACM*, 40(5), 1997, pp. 103-110.

<sup>19</sup>CODD, E. E., *The Relational Model for Database Management: Version 2*. Addison-Wesley, Reading, MA, 1990. GARVÍN, D. A., "Competing on the Eight Dimensions of Quality," *Harvard Business Review*, 65 (6), 1987, pp. 101-109. JURAN, J. M., *Juran on Leadership for Quality: An Executive Handbook*. The Free Press, New York, 1989. LAUDON, K. C., "Data Quality and Due Process in Large Interorganizational Record Systems," *Communications of the ACM*, 29 (1), 1986, pp. 4-11. WAND, Y., y R. Y. WANG, "Anchoring Data Quality Dimensions in Ontological Foundations," *Communications of the ACM*, 39 (11), 1996, pp. 86-95. WANG, R. Y., y D. M. STRONG, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems (JMIS)*, 12(4), 1996, pp. 5-34.

ajustadas al ámbito de su organización. Es preciso que concreten los procesos organizativos necesarios, así como las alternativas técnicas, para la gestión del producto de información. En función de este objetivo, es preciso que el director de productos de información (DPI) desarrolle la correspondiente métrica, una vez que ya se hayan definido los parámetros IQ, para la medida y el análisis de la calidad de la información, así como para su correspondiente mejora<sup>20</sup>.

¿Hasta qué punto es correcta la información contenida en nuestras bases de datos, almacenes de datos o sistemas de información relativa a la materia? ¿Cómo saldría parada la calidad de nuestra información frente a la del resto del sector? ¿Existe un conjunto de criterios de medida para estas comparaciones? ¿Existe una medida única, útil y de conjunto para la Calidad de la Información (IQ)?

Resolver estas tres cuestiones básicas ha instado a los autores a desarrollar un conjunto adecuado de valores de métrica que permitan realizar las medidas necesarias. La obligación de realizar medidas mantiene una vinculación inseparable de las relacionadas con el análisis y la mejora de la IQ. Como primer problema no existe un criterio universal para la medida de la IQ. Es posible crear unas funciones ponderadas y de conjunto, pero éstas se ceñirán y serán el resultado de una asignación subjetiva de los factores de ponderación. Al desarrollar la métrica de la calidad, ha sido importante que se tener en cuenta los muchos factores que intervienen<sup>21</sup>.

<sup>20</sup>FIRTH, C. P. y R. Y. WANG, *Data Quality Systems: Evaluation and Implementation*. Cambridge Market Intelligence Ltd., London, 1996. GARVIN, D. A., "Competing on the Eight Dimensions of Quality" *Harvard Business Review*, 65(6), 1987, pp. 101-109. HAUSER, J. R. y D. CLAUSING, "The House of Quality," *Harvard Business Review* 66(3), 1988, pp. 63-73. HUH, Y. U., F. R. KELLER, T. C. REDMAN y A. R. WATKINS, "Data Quality," *Information and Software Technology*, 32 (8), 1990, pp. 559-565. KNIGHT, B. "The Data Pollution Problem," *Computernworld*, September 28, pp. 81-82, 1992. KON, H. B., J. LEE y R. Y. WANG, *A Process View of Data Quality (No. TDQM-93-01)*. Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management, Cambridge, MA, 1993. LEPAGE, N., "Data Quality Control at United States Fidelity' and Guaranty Company." *Data Quality Control: Theory and Pragmatics*, G. E. LIEPENS and V R. R. UPPULURI, ed., Marcel Dekker, Inc., New York, 1990. LEVITIN, A. y T. REDMAN, "Quality Dimensions of a Conceptual View," *Information Processing & Management*, 31(1), 1995, p. 8. MCGEE, A. M., *Total Data Quality Management (TDQM): Zero Defect Data Capture (No. TDQM9207)*. Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management, Cambridge, MA, 1992. MCGEE, J. y L. PRUSAK, *Managing Information Strategically*. The Ernst & Young Information Management Series, John Wiley & Sons, Inc., New York, 1993. MEYEN, D. M. y M. J. WILLSHIRE. "A Data Quality Engineering Framework," *Proceedings of The 1997 Conference on Information Quality*. Cambridge, MA, pp. 95-116, 1997. REDMAN, T. G., *Data Quality: Management and Technology*. Bantam Books, New York, 1992. REDMAN, T. C., *Data Quality for the Information Age*. Artech House, Boston, MA, 1996. RYAN, J. E., "Data Quality with LIMS," *Quality*, Mayo, 1988, pp. 12-15. SEGEV, A., "On Information Quality and the WWW Impact." *Proceedings of the 1996 Conference on Information Quality*. Cambridge, MA, pp. 16-23, 1996.

<sup>21</sup>IBM "Global Services Wins Top Honor For Its Knowledge Management Innovations," IBM, 1998. <http://www.ibm.com/services/pressrel/pr.889566539.html>. KNIGHTS, D., MURRAY, E. y H. WILLMOT, "Networking as Knowledge Work: A Study of Strategic Interorganizational Development in the Financial Services Industry," *Journal of Management Studies*, 30(6), 1993, pp. 975-995. LEONARD BARTON, D., *Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation*. Harvard Business School

Amparados en los datos acumulados del estudio y de la práctica de la IQ, planteamos tres clases complementarias de métrica para la evaluación de las necesidades relativas a la Calidad de la Información contenida en el Tesoro:

-Una métrica para la medida de la valoración subjetiva de la IQ por parte de los usuarios.

-Una métrica para la medida de la calidad de la información según variables objetivas y cuantificables, así como independientes de la aplicación en sí (¿Hasta qué punto son exhaustivos, coherentes, actualizados y correctos los datos contenidos en nuestro sistema de información del usuario?)

-Una métrica para la medida de la calidad de la información según variables objetivas y cuantificables, así como vinculadas a la aplicación.

Cuando se emplea de forma combinada, la métrica de cada una de estas clases nos proporciona una información fundamental que va más allá de la evaluación estática de la IQ, para convertirse en una técnica de evaluación dinámica y continua de mejora de la calidad de la información. Con cada clase se mide una serie de aspectos distintos. La métrica que se emplea para conocer la evaluación subjetiva guarda relación con la percepción del individuo en lo que a la calidad de la información se refiere. La métrica independiente de la aplicación va más allá de las aplicaciones específicas y se mantiene independiente del contexto. La métrica vinculada a la aplicación aporta un significado y una pertinencia relacionadas de forma concreta a una aplicación y en función del contexto<sup>22</sup>.

### **Métrica subjetiva de la Calidad de la Información**

La métrica subjetiva de la IQ mide la valoración subjetiva de la IQ por parte que un individuo. A partir de las dieciséis dimensiones o parámetros de la Calidad de la Información que se han establecido desde diversos estudios<sup>23</sup>, podemos formular una serie de preguntas (un cuestionario) para conocer la percepción de la situación de la IQ en el tesoro de tipografía y edición. Este tipo de cuestionario fue desarrollado a partir de los estudios acumulados del programa TDQM de MIT<sup>24</sup>.

---

Press, Boston, MA, 1995. LLOYD, T., "ESS: Technical Architecture Project-Stage 1 Technical Report." *Informe Interno de IBM*, 1996. LONG, C. y M. VICKERS-KOCH, "Using Core Capabilities to Create Competitive Advantage," *Organizational Dynamics* (24), 1995, pp. 7-20.

<sup>22</sup>BADARACCO, J., "Knowledge Links," in *The Knowledge Link: How Firms Compete Through Strategic Alliances* 1991, Harvard Business School Press, Boston, MA, 1991. BARLEY, S. R., J. FREEMAN, y R. C. HYBELS, "Strategic Alliances in Commercial Biotechnology," in *Networks and Organizations*, N. NOHRIA, ed., Harvard Business School Press, Boston, MA, 1992. DAVENPORT, T. H., "Saving IT's Soul: Human Centered Information Management," *Harvard Business Review*, 72(2), 1994, pp. 119-131. DAVIS, S., y J. BOTKIN, "The Intelligent Enterprise and New Paradigm," *Academy of Management Executive*, 4(6), 1992, pp. 48-63.

<sup>23</sup>WANG, Y. R. y S. E. MADNICK. "A Source Tagging Theory for Heterogeneous Database Systems," *Proceedings of International Conference on Information Systems*. Copenhagen, Denmark, pp. 243-256, 1990.

<sup>24</sup>PIPINO, L., Y. W. LEE y R. Y. WANG, *Measuring Information Quality (No. TDQM-97-04)*. MIT Sloan School of Management, Cambridge, MA, 1998. PIPINO, L. y R. Y. WANG, *Developing Measurement*

Cada respuesta se evalúa empleando una escala tipo Likert comprendida del cero al diez, en la que el cero supone “en total desacuerdo” y el 10 “completamente de acuerdo”. A modo de ejemplo, una de las preguntas empleadas para analizar la dimensión relativa al pleno alcance de los datos sería: “¿Resulta la información suficientemente completa como para satisfacer sus demandas?”. Este cuestionario se ha empleado de forma eficaz tanto en sectores públicos como privados, se puede emplear para medir la percepción relativa a la calidad de la información. Asimismo, se puede utilizar como un instrumento de análisis para la evaluación de la calidad de la información desde una perspectiva mucho más amplia que la que aporta la exactitud de la misma como elemento aislado. La información obtenida durante esta fase de diagnóstico proporciona la motivación necesaria para la creación de los métodos destinados a la mejora de la calidad de la información según la perciben los consumidores del tesoro.

### **Métrica objetiva de la Calidad de la Información (IQ) en el tesoro de Edición y Tipografía**

La métrica objetiva e independiente de la aplicación sirve para la medida de la IQ en función de variables cuantificables y objetivas. En definitiva, sirve para conocer el alcance pleno, la coherencia, la veracidad y la actualización de la información que se encuentra en el Tesoro. Esta métrica se ampara en la teoría del control de la calidad que se introduce en un sistema. En aquellos sistemas en los que no existen controles en el momento de adquisición de los datos, aún existiría la posibilidad de emplear estas medidas con el fin de conocer hasta qué punto cumple la información existente con la norma aplicable.

El diseño de la mayoría de los sistemas de base de datos obedece a una perspectiva de sistema. Los mecanismos tales como las restricciones de integridad y las teorías de normalización que se emplean para mantener la integridad y la coherencia de la información son necesarios, aunque insuficientes, para garantizar la calidad de la información que exigen los usuarios<sup>25</sup>.

---

*Scales for Data Quality Dimensions (No. TDQM-96-04)*. MIT Sloan School of Management, Cambridge, MA, 1996. REDDY, M. P. y R. Y. WANG. “Estimating Data Accuracy in a Federated Database Environment.” *Proceedings of 6th International Conference, CISMOD (También en Lecture Notes in Computer Science)*. Bombay, India, pp. 115-134, 1995.

<sup>25</sup>COLLINS, H. M., “Humans, Machines, and the Structure of Knowledge,” in *Knowledge Management Tools*, R. L. RUGGLES, ed., ButterworthHeinemann, Boston, MA, 1995. CONSTANT, D., L. SPROULL, y S. KIESLER, “The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice,” *Organization Science*, 7(2), 1996, pp. 119-135. HEIBLER, R. J., “Benchmarking Knowledge Management,” *Strategy & Leadership*, March/April, 1996, pp. 22-29. HUANG, K. T., “Knowledge Is Power: So Use It or Lose It”. IBM, 1998. <http://www.ibm.com/services/articles/inttelcapsun.html>. KOGUT, B., y U. ZANDER, “Knowledge of the Firm Combinative Capabilities, and the Replication of Technology,” *Organization Science*, 3 (3), 1992, pp. 383-397. ORR, J. E., “Sharing Knowledge, Celebrating Identity,” *Collective Remembering*, D. S. MIDDLETON y D. EDWARDS, ed. 1990, Sage, Newbury Park, CA, 1990. PISANO, G. P., “Knowledge, Integration, and the Locus of Learning: An Empirical Analysis of Process Development,” *Strategic Management Journal*, 15, 1994, pp. 85-100.

En este sentido, el doctor Edgar F. Codd<sup>26</sup> planteó las cinco reglas de integridad que debe seguir cualquier tipo de base de datos relacionales. A pesar de que fueron desarrolladas de forma específica para su aplicación al modelo relacional, estas reglas de integridad resultan muy útiles en otros contextos que van desde los sistemas de bases de datos relacionales y de bases de datos en red a los archivos simples. De hecho, muchas de las bases de datos corporativas inician ahora la transición al entorno de bases de datos relacionales tales como el DB2 de IBM<sup>27</sup>, a partir de sistemas de gestión jerárquica de base de datos tales como que el CICS de hace una década. El hecho de garantizar la calidad de la información en esas bases de datos exige una aproximación metódica tal como la que ofrecen las reglas de integridad propuestas por Codd.

Incluso en un entorno relacional, muchas bases de datos relacionales de hoy en día incumplen las reglas de integridad debido a varios motivos: la aplicación específica de base de datos relacionales no dispone de un mecanismo para hacer cumplir con la integridad de los datos, o el fallo en la introducción de los datos específicos, e, incluso, que al programar comprobaciones editoriales en el sistema de gestión de la base de datos del tesouro hemos eliminado controles debido a las prisas.

Para simplificar, las reglas de integridad de Codd garantizan que los datos cumplan con las especificaciones exigidas por el diseñador y el usuario. Por consiguiente, un instrumento IQ cuyo desarrollo se base en las reglas de integridad de Codd ofrece un método formal para definir, medir, analizar y mejorar la calidad de la información. A continuación, repasamos los cinco tipos de integridad propuestas por Codd y que hemos aplicado en la base de datos del Tesouro.

-Integridad de dominio: todos los valores de un campo deben ajustarse a un mismo dominio.

-Integridad de columna: especifica el conjunto de valores aceptables en una columna.

-Integridad de entidad: no admite que a un elemento de clave primaria le falte un valor de cualquier tipo.

-Integridad referencial: por cada clave ajena de una base de datos relacional, debe existir en ésta un valor equivalente de clave primaria del mismo dominio. Si la clave ajena es compuesta, los componentes que a su vez sean claves ajenas deben presentarse en la base de datos como componentes con, como mínimo, un valor de clave primaria procedente del mismo dominio.

-Integridad definida por el usuario: con esto se recogen las prácticas de usuario que deberán quedar reflejadas en la base de datos: las restricciones definidas por el usuario se emplean no sólo para garantizar el estado correcto de la base de datos

---

<sup>26</sup> CODD, E. F., *The Relational Model for Database Management: Version 2*. Addison-Wesley, Reading, MA, 1990.

<sup>27</sup> IBM, "IBM's Planning Guidelines of Leveraging Asset Use in Solutions and Services." *Informe Interno de IBM*, 1996.

sino, también, con el fin de iniciar acciones específicas ante la aparición de unas condiciones concretas en la base de datos.

### **Métrica de la Calidad de la Información (IQ) vinculada a la aplicación del Tesauro de Edición y Tipografía**

Con la métrica vinculada a la aplicación se consigue medir la calidad de la información conforme con unas variables cuantificables y objetivas que son específicas del dominio del tesauro de tipografía y que exigen la participación de expertos en las materias de dicho dominio. Algunos tipos de métrica<sup>28</sup> vinculadas a la aplicación son, relativamente, de carácter intuitivo y de fácil desarrollo, mientras que otras pueden ser muy complejas. Se han iniciado con la revisión de la literatura al efecto y se ha dado por definitiva en el estudio de los elementos integradores de recuperación en la interfaz de usuario.

### **Revisión de la literatura terminológica de tipografía**

Para la detección y recuperación de la terminología española relativa al campo de la Tipografía, se han utilizado diversas obras, pero dos de ellas destacan por su mayor aportación a este trabajo: el *Diccionario de edición, tipografía y artes gráficas de Martínez de Sousa*<sup>29</sup> y el *Manual de Tipografía* de John Kane<sup>30</sup>.

El *Diccionario de edición, tipografía y artes gráficas* recoge definiciones en el ámbito de la tipografía. Se trata de la obra más exhaustiva en nuestro idioma sobre esta materia, que abarca además las artes gráficas y el proceso de creación de un libro. Las entradas lexicográficas recogen información enciclopédica en aquellas cuestiones en las que es necesaria una explicación o aclaración para una mejor comprensión. Esta característica ha sido muy útil en la elección del término principal dentro de las relaciones de equivalencia. A todo esto se suma otra virtud, puesto que el diccionario incluye con frecuencia el término equivalente en francés, en inglés, o en ambos idiomas. Incluye además remisiones en un sentido similar al de los no descriptores de un tesauro. Sin embargo, la obra adolece, en algún que otro concepto, de la profunda idiosincrasia de su autor. Esto quiere decir que no siempre se ha utilizado como criterio de inclusión en este tesauro la opción preferente de Martínez de Sousa, puesto que se aleja del uso y criterio de otros profesionales del ámbito que también se ha tenido en cuenta.

El *Manual de Tipografía* es una obra de marcado carácter práctico y bastante accesible para personas que se inician en la materia, y por tanto una fuente de extracción terminológica útil, ya que el libro inicia su exposición con las partes de la

<sup>28</sup>WANG, R. Y. y D. M. STRONG, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems* (JMIS), 12(4).

<sup>29</sup>MARTÍNEZ DE SOUSA, J. *Diccionario de edición, tipografía y artes gráficas*. Trea: Gijón, 2001.

<sup>30</sup>KANE, J. *Manual de Tipografía*. Traducido por M. DÁVILA. Barcelona: Gustavo Gili, 2005. Traducción de: *A type primer*.

letra y avanza pasando por los niveles de frase, párrafo y texto, para terminar con nociones de composición.

En lo referente tanto a la recuperación terminológica como a la traducción al inglés, la obra principal usada en este trabajo ha sido *Typography & Graphic Design Thesaurus* de Paul M. Hoffman<sup>31</sup>. Es un tesoro que contiene alrededor de 1900 descriptores. Cuenta con una presentación sistemática de nueve facetas y otra alfabética.

Para analizar otros elementos relacionados con la Gestión de Calidad Total de los Datos, teniendo como norte que partíamos de un control terminológico exhaustivo con fuentes de reconocida fiabilidad como hemos referido anteriormente, empleamos en la gestión transversal de los datos en consulta con *Integrity Analyzer* (IA). Es uno de los productos resultado de los estudios realizados por MIT en materia de gestión de la calidad total de los datos (TDQM)<sup>32</sup>, que presenta una metodología TDQM en la que se combinan los principios del ciclo aplicable a este tipo de gestión con los correspondientes a las restricciones de integridad en las bases de datos relacionales, tal como se muestra en la tabla:

	Integridad del dominio	Integridad de la entidad	Integridad referencial	Integridad de columna	Integridad definida por el usuario
Definir	Definir los dominios empleados en la base de datos. Especificar el dominio asociado a cada columna.	Especificar la clave primaria para cada tabla y cualquier clave potencial	Especificar todas las claves ajenas. Especificar la clave primaria asociada a cada una.	Especificar las normas de los valores aceptables para cada columna	Especificar toda norma empresarial no recogida por la integridad de entidad, referencial y de columna
Medir	Comprobar la existencia de infracciones, por ejemplo, de valores de una columna que no proceden del dominio correspondiente.	Comprobar la existencia de infracciones, por ejemplo, claves nulas o no exclusivas.	Comprobar la existencia de infracciones, por ejemplo, valores de claves ajenas que no coinciden con el valor de la clave primaria.	Comprobar la existencia de infracciones, por ejemplo, valores de columna que presentan valores inaceptables.	Comprobar la existencia de infracciones, por ejemplo, el registro de los casos que no se ciñen a estas normas.
Analizar	Analizar la medida realizada.	Analizar la medida realizada.	Analizar la medida realizada.	Analizar la medida realizada.	Analizar la medida realizada
Mejorar	Revisar el registro de infracciones y cambiar los valores según sea necesario.	Revisar el registro de infracciones y cambiar los valores según sea necesario.	Revisar el registro de infracciones y cambiar los valores según sea necesario.	Revisar el registro de infracciones y cambiar los valores según sea necesario.	Revisar el registro de infracciones y cambiar los valores según sea necesario.

<sup>31</sup>HOFFMAN, P. *Typography & Graphic Design Thesaurus* [en línea]. Ver. 1.1. 31 mayo 2006. Disponible en: <<http://hoffmancommappaul.com/tgdt/>>

<sup>32</sup>WILSON, P., "Information Retrieval and Cognitive Authority." *Second Hand Knowledge*, Greenwood Publishing Group, Westport, CT, 1983. WEISS, A. R., y P. BIMBAUM, "Technological Infrastructure and the Implementation of Technological Strategies," 35(8), 1989, pp. 1014-1026. QUINN, J. B., P. ANDERSON, y S. FINKELSTEIN, "Managing Professional Intellect: Making the Most of the Best," *Harvard Business Review*, 2(74), 1996, pp. 71-80. WANG, R. Y. y H. B. KON, "Towards Total Data Quality Management (TDQM)." *Information Technology in Action: Trends and Perspectives*, R. Y. WANG, ed., Prentice Hall, Englewood Cliffs, NJ, 1993. RYAN, J. E., "Data Quality with LIMS," *Quality*, Mayo, 1988, págs. 12-15. MEYEN, D. M. y M. J. WILLSHIRE. "A Data Quality Engineering Framework," *Proceedings of The 1997 Conference on Information Quality*. Cambridge, MA, pp. 95-116, 1997.

Cada columna de la tabla se correspondía con uno de los cinco principios de la integridad de Codd: dominio, entidad, referencial, columna e integridad definida por el usuario. La integridad de dominio exigía que todos los valores dispuestos en la columna de unas tablas procediesen de un mismo dominio. La integridad relativa a la entidad exigía que cada entidad (tabla) disponga de una clave primaria que conste de una o más columnas. La clave primaria debía ser única sin que le fallase ningún valor. La integridad referencial exige que, por cada clave ajena única de una base de datos relacional, existiese un valor equivalente de clave primaria en la base de datos perteneciente al mismo dominio. Con la integridad de columna se limitan aún más los valores que pueden extraerse de un dominio para una columna concreta. En resumen, con la integridad de columnas se concreta el conjunto de valores aceptables para la columna. Es posible especificar estos valores en función de exigencias de exclusividad de valores positivos, un margen de valores aceptados o una lista de valores aceptados. Con la integridad definida por el usuario se especifican las normas complementarias con las que debían cumplir los valores de las columnas y que se adaptaron a las diferentes necesidades de los diferentes usuarios, desde el traductor hasta el tipógrafo profesional. Con frecuencia, estas normas incluyeron condiciones que van en función de los valores que presentan otros campos. Estas cuatro acciones - definir, medir, analizar, y mejorar - de la IQ forman las filas de la tabla. Asimismo, se prestan para satisfacer los requisitos de dominio, entidad, referencial, columna e integridad definida por el usuario-consulta. Las celdas de la tabla muestran la aplicación de una acción IQ a un tipo determinado de requisito de integridad en un entorno de lenguaje documental controlado.

Esta herramienta ofreció varios instrumentos adicionales de apoyo del análisis de Calidad de información (IQ). De forma más concreta, uno de los más útiles es el de las comprobaciones de frecuencia, que ofrece datos acerca de los valores y la frecuencia con la que aparece cada valor en una columna. A menudo, las comprobaciones de frecuencia se emplean de forma conjunta con la integridad de columna. Por tanto podemos extraer del número de frecuencias vectores de uso y couso que permiten asegurar en otras herramientas terminológicas el peso de estas frecuencias en relación a los conceptos más utilizados y consultados por los usuarios profesionales y prever el uso e interés de estos usuarios en una próxima consulta.

Es posible que el usuario defina una o más comprobaciones de frecuencia específicas o que edite una especificación anterior relativa a estas comprobaciones. Al elegir la función medición, el usuario inicia una comprobación concreta de la frecuencia. El usuario puede visualizar los resultados de la evaluación al seleccionar la opción analizar. El *Integrity Analyzer* funcionó tanto con las restricciones formales de integridad de Codd como con las restricciones definidas por el usuario. La evaluación de la integridad de dominio, de columna, de entidad y referencial permanecieron sin cambios en todas las aplicaciones. Todos los sistemas de gestión de base de datos relacionales, de forma idónea, deberían seguir estas normas.

Asimismo, todas las bases de datos, al margen del tema que recojan, deberían adherirse a las normas que imponen este tipo de restricciones.

No obstante, las decisiones definidas por el usuario quedan vinculadas a la aplicación. Éstas varían en función de las distintas aplicaciones, así como entre distintos sectores. Asimismo, las restricciones vinculadas a la aplicación evolucionan en función del tiempo. Debemos destacar que el analizador de integridad va más allá de la mera implantación de las restricciones de integridad de los programas de bases de datos relacionales. Al contrario que lo que ocurre con los habituales programas comerciales de sistemas de gestión de bases de datos relacionales (RDBMS) que comprueban la adherencia a las restricciones de Codd cuando se introducen los datos en la base, *Integrity Analyzer* supuso una herramienta de diagnóstico para la evaluación del grado de adhesión de las bases de datos existentes a las restricciones definidas por Codd, así como a las normas definidas por el usuario en este caso fueron las normas UNE e ISO que queden vinculadas a la aplicación. Como herramienta de diagnóstico, el analizador proporcionó un análisis del estado actual de la calidad de los datos recogidos y destacó las zonas susceptibles de mejora.

Puesto que la información obedece a una construcción social, no queda garantizada su neutralidad ni su objetividad<sup>33</sup>. Asimismo, la información que necesitan los usuarios tampoco tiene por qué ser objetiva, pero sí universal. La información recogida y empleada por el Tesouro constituye un reflejo del conocimiento aplicado a la generación de los datos y del contexto en el que se enmarcan sus objetivos y el ámbito del conocimiento. Los tesauros dejan al descubierto el conocimiento que se esconde tras la información y desarrollan nuevos conocimientos a partir de dicha información. Éste es el motivo por el que se debe prestar un mayor grado de atención a la producción de información de calidad, así como a todo el proceso global de generación de datos. Una vez que el contexto en el que se enmarcan los datos y el conocimiento empleados para producir la información queda separado de la información en sí, especialmente, cuando se transfiere a otra base de datos o emplazamiento, resultará mucho más difícil identificar y reproducir el conocimiento inherente a esta información.

Es por ello por lo que debe reconocerse de forma explícita las distintas funciones que desempeñan los datos y su medida en un sistema. Empleando tanto una descripción multinivel de los datos, contextos e informaciones, así como de usuarios, se facilita la comprensión de las distintas modalidades del conocimiento y con ello podemos reconciliar e integrar las diferentes perspectivas de la información relativa a un mismo producto o servicio, es decir, aunar en macrotesouro de edición información relativa que puede ser, a su vez, integrada en por captura o adquisición en formas más complejas como ontologías en lenguaje OWL o mapas de conocimientos visuales. En este sentido nos vimos obligados a identificar tres

---

<sup>33</sup>DENNING, E. y P. J. DENNING, "Data Security," *A CM Computing Surveys*, 1 1(3), 1979, pp. 227-250.

modalidades del conocimiento que son aplicables a la formación de las capacidades organizativas:

- Know-what (saber el qué), que se refiere al conocimiento de los hechos.
- Know-how (saber el cómo), que se refiere al conocimiento de los procedimientos.
- Know-why (saber el por qué), que se refiere al conocimiento axiomático. En este aspecto se incluye de forma explícita el conocimiento de los motivos y de los supuestos axiomáticos que subyacen en las prácticas. Es decir, Manuales de procedimiento, comprobación de datos, comprobación de información y flujo, etc.

### **Confección del tesoro**

El proceso seguido se explica aquí dividido en distintas fases. Sin embargo, la práctica ha demostrado que éstas no pueden ser seguidas de manera rígida, puesto que eso significaría que en un determinado momento se decide dejar de incorporar términos porque ya estamos en la fase de establecimiento de las relaciones semánticas. Esta flexibilidad en la implementación de la planificación coincide con lo que afirma Currás (2005): “Lo que sí es aconsejable es ir realizando las distintas etapas simultáneamente. El trabajo resulta más fluido, más eficaz y menos monótono<sup>34</sup>.”

### **Definición y delimitación del problema**

Una de las primeras cuestiones que se plantean en la elaboración de cualquier tesoro tiene que ver con la delimitación terminológica, es decir, dónde se pone la frontera entre lo que hay que incluir en él y lo que es irrelevante para el posible usuario de este recurso documental.

A priori, una posible solución puede encontrarse en una definición de tipografía (Martínez de Sousa, 2001): “Procedimiento de impresión con formas o moldes cuyos motivos impresores están en relieve. Estética de lo impreso, especialmente relacionada con la composición y la compaginación. Aspecto general de un texto, dependiente del ojo y tipo de letra empleados en su composición<sup>35</sup>.”

Se puede ver en estas tres acepciones escogidas que hay dos aspectos diferenciados dentro de esta disciplina, uno más técnico y otro orientado hacia lo creativo. Renner (2000) afirma que: “Técnica, práctica y estética no constituyen compartimentos separados en la tipografía, sino que representan facetas distintas, constituyen tareas parciales que contribuyen a una sola obra y que se plantean de forma recurrente en su realización<sup>36</sup>.” Esto lleva a pensar que sería conveniente una flexibilidad en cuanto a la búsqueda y recuperación terminológica. Así pues, los tipos de letra o la maquetación, que condicionan la estética y la capacidad para

---

<sup>34</sup>CURRÁS, E. *Ontologías, taxonomía y tesauros. Manual de construcción y uso*. Gijón: Trea, 2005.

<sup>35</sup>MARTÍNEZ DE SOUSA, J. *Diccionario de edición, tipografía y artes gráficas*. Trea: Gijón, 2001.

<sup>36</sup>RENNER, P. *El arte de la tipografía*. Traducido por J.A. Cifuentes y E. Monzó. Valencia: Campgràfic, 2000.

atraer de un texto, son aquí tan importantes como las actividades de producción editorial.

Aun así no deja de ser una selección altamente subjetiva, puesto que lo que un autor puede considerar como perteneciente al campo de la tipografía, puede ser considerado por otro como un tema más o menos relacionado.

### **Recuperación terminológica**

En este paso, se ha utilizado el método inductivo-deductivo. Esto implica una primera fase cuyo objetivo es la recopilación del mayor número posible de términos para su posterior inclusión. Para ello en un primer momento se recurre a información online, en aras de una mayor inmediatez y una información más actualizada. No obstante, las obras recuperadas a través de la red, aunque numerosas (no faltan glosarios, enciclopedias, etc.), no son todo lo completas que sería deseable, y se decide acudir a manuales, monografías y diccionarios de tipografía en soporte impreso. También han servido como fuentes de información para la elaboración de este tesoro artículos extraídos de diferentes bases de datos, manuales didácticos para la formación de correctores, otros tesauros en otros idiomas, etc. De éstos, por tanto, sólo los tesauros suponen una fuente terminológica previamente estandarizada.

Esta variedad de fuentes ha sido útil no sólo en la fase de recopilación, sino también a la hora de contrastar información terminológica cuando existía disparidad de criterios o cuando se quería confirmar la veracidad del uso de un término detectado. Una vez obtenido un número satisfactorio de términos, se han establecido las entradas de acuerdo con las características de los términos, haciendo esta selección de acuerdo con las relaciones lógicas que se establecen entre ellos. Esta operación funciona en un sentido término genérico – término específico.

El método deductivo se ha empleado una vez que las entradas del tesoro estaban definidas. Como esta categorización no finaliza el proceso, se siguió detectando términos que había que incluir en las entradas dispuestas (o bien crear una nueva categoría retomando así el proceso deductivo). En el proceso deductivo la inclusión funciona en un sentido término específico – término genérico.

### **Selección: criterios**

En principio, el criterio más importante a la hora de decidir la inclusión o no de un término ha sido el de uso frente a desuso en el sentido de actualidad. Es decir, se han descartado términos que han dejado de emplearse por una u otra razón, como puede ser el caso de *linotipia*. Hecha esta excepción, el resto de términos han sido incluidos puesto que su presencia no hace sino enriquecer la función de recuperación de información que esta herramienta pretende.

Otro aspecto relacionado con la selección terminológica afecta a las categorías gramaticales de las palabras susceptibles de ser incluidas. Se ha optado por la

primacía del sustantivo frente a otras clases, aunque se puede observar que en muchas ocasiones éste deriva de una actividad expresada mediante un verbo.

### **Establecimiento de relaciones**

Las relaciones entre términos se han realizado de acuerdo con lo establecido en la norma UNE 50-106-90<sup>37</sup>. Para designarlas se ha optado por la notación con siglas en lugar de símbolos. Por ejemplo, **UP** en vez de = para designar en la lista alfabética a los no descriptores. En la lista sistemática se han señalado mediante la notación **ND**. En inglés se han utilizado las siglas en este idioma.

Se puede ver el modo en que se han establecido las relaciones en el apartado 1.9. Ejemplo de entrada.

### **Inclusión de términos compuestos**

La inclusión de términos compuestos en un tesoro parece algo lógico cuando se están introduciendo segmentos de un lenguaje natural en un lenguaje controlado. Aunque se puede optar por incluir tan sólo unitérminos, esta decisión acarrea una complicación adicional puesto que elimina la delimitación conceptual que el término compuesto posee. Por tanto aparecen términos compuestos en el presente tesoro.

En cuanto a su fraccionamiento, tanto el fraccionamiento semántico como el sintáctico han sido descartados. No se ha incluido, por ello, **PIE + FOTOGRAFÍA** como **ND** del término *pie de fotografía*. Los términos aparecen tal y como se pueden encontrar en los documentos originales puesto que no se estima que su fraccionamiento pueda aportar ninguna mejora a la indización y recuperación terminológica en el campo que este tesoro abarca.

### **La cuestión del número**

En la confección de tesoros la inclusión de los términos en su forma singular o plural está también determinada por la normativa existente, que establece cuatro posibles casos:

- El término significa lo mismo en singular que en plural: en este caso se opta por uno u otro dependiendo de diversos factores:
  - o Si el sustantivo es contable se elige su forma plural. Un ejemplo de esto en este tesoro podría ser el término *Erratas*.
  - o También se incluye el término en plural cuando forma parte de un todo y esta parte se encuentra más de dos veces. Esto sucede en la entrada *Partes de la letra*
  - o Si el sustantivo es incontable, se elige la forma singular. Esto hace referencia a cualidades, propiedades, actividades o disciplinas. Se encuentran aquí ejemplos como: *Legibilidad, Corrección, Maquetación*, etc.

---

<sup>37</sup>Asociación Española de Normalización y Certificación. *Directrices para el establecimiento y desarrollo de tesoros monolingües*. UNE 50-106-90. Madrid: AENOR, 1990.

- El significado varía de una forma a otra: debe incluirse el término que responda a nuestras necesidades, ya sea en plural, en singular, o en ambas formas. Es el caso del término compuesto *dos puntos*, cuyo significado se modifica si se compara con el singular *punto*, puesto que su función como signo de puntuación difiere del punto.
- Ambas formas tienen el mismo significado y se pueden usar indistintamente: no se han encontrado
- La forma singular y la forma plural se escriben con distinta ortografía: no se han encontrado

Cabe mencionar que en este tesoro han aparecido términos afectados por este fenómeno lingüístico que exceden lo establecido por las directrices anteriormente expuestas. Es el caso de términos que no varían en su forma al pasar de singular a plural, como por ejemplo paréntesis.

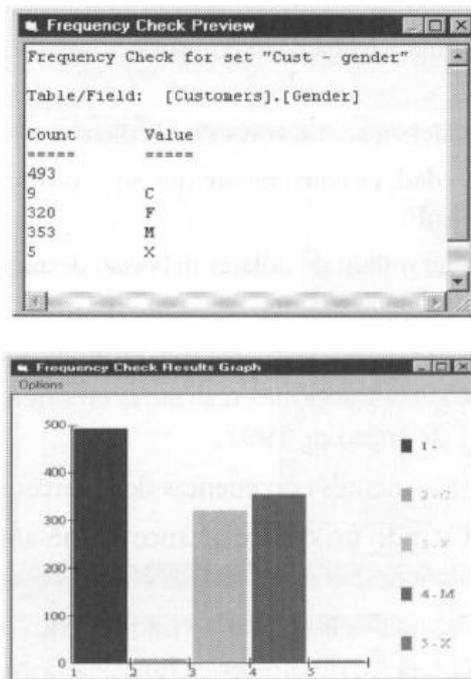
### La extracción de datos y conocimiento

La extracción de datos es el proceso con el que se extrae información comprensible y previamente desconocida a partir de grandes bases de datos para su uso en los procesos de toma de decisiones. Esta técnica se reconoce cada vez más como un instrumento esencial de inteligencia destinada a revelar la información que necesita un usuario para mejorar su posición antes de elaborar el producto de información y facilitar así la retroalimentación. Por medio del empleo de un conjunto de instrumentos destinados al usuario final, la extracción de datos es accesible a todos, no sólo a los analistas de datos sino, también, a los traductores, correctores, ortotipógrafos, etc. Este proceso modifica la forma en la que los productores de información (editores, traductores, tipógrafos, etc.), toman decisiones. La extracción de conocimientos, parecida a la de datos, es un proceso de abstracción de conocimientos válidos y previamente desconocidos, que acabarán siendo comprensibles a partir de una variedad de fuentes de información. La extracción de conocimientos mejora de forma drástica las capacidades de búsqueda de información al ampliar la diversidad de las fuentes de información. Asimismo, se hace posible desvelar información de forma puntual y exhaustiva gracias a la integración de sucesos actuales y datos de transacciones con las distintas fuentes. Esto permite recolectar la información que antes quedaba excluida de las bases de datos relacionales<sup>38</sup>.

---

<sup>38</sup>BEIZER, B., *Software Testing Techniques*, 2.a ed., Van Nostrand Reinhold, 1990. BRILLIANT, S.S., J.C. KNIGHT, y N.G. LEVENSON, "The Consistent Comparison Problem in N-Version Software", *ACM Software Engineering Notes*, vol. 12, n.º 1, enero 1987, pp. 29-34. DEUTSCH, M., Verification and Validation, *Software Engineering*, R. JENSEN y C.TONIES (eds.), Prentice-Hall, 1979, pp. 329-408. FOSTER, K.A., "Sensitive Test Data for Boolean Expressions", *ACM Software Engineering Notes*, vol. 9, n.º 2, Abril 1984, pp. 120-125. HOWDEN, W.E., "Weak Mutation Testing and the Completeness of Test Cases", *IEEE Trans. Software Engineering*, vol. SE-8, n.º4, julio 1982, pp. 371-379. KNIGHT, J., y P. AMMANN, "Testing Software Using Multiple Versions", *Software Productivity Consortium*, Report n.º

Las técnicas de extracción de conocimientos emplean las habituales técnicas de depuración de datos tales como la aglomeración, la clasificación, la predicción de valores, el descubrimiento de asociaciones, el descubrimiento de patrones secuenciales y de secuencias de tiempo similares. También incorpora nuevas tecnologías tales como la compilación, la simbolización, la abstracción por palabra clave, el análisis semántico y el reconocimiento de conceptos/características, así como las técnicas de análisis de metadatos tales como el análisis de contexto y semántica<sup>39</sup>.



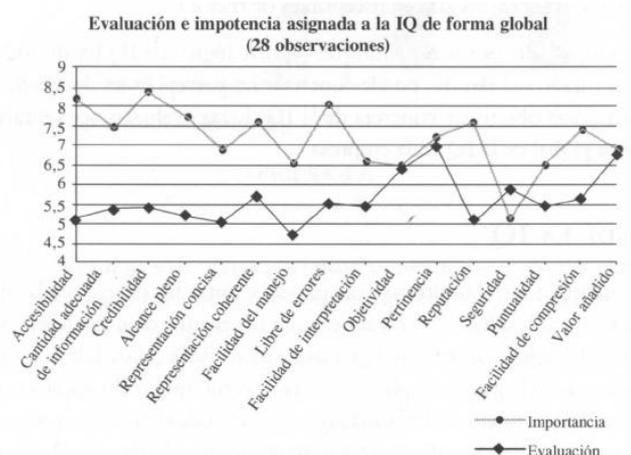
Fuente: Tesauro de Información y Edición. Comprobación de frecuencia.

89029N, Reston, VA, Junio 1989. Tai, K.C., "What to Do Beyond Branch Testing", *ACM Software Engineering Notes*, vol. 14, n.º 2, Abril 1989, pp. 58-61.

<sup>39</sup>BOWAN, J.P., y M.G. HINCHLEY, "Ten Commandments of Formal Methods, Computer", vol. 28, n.º 4, Abril 1995. GRIES, D., y F.B. SCHNEIDER, *A Logical Approach to Discrete Math*, Springer-Verlag, 1993. GUTTAG, J.V., y J.J. HORNING, *Larch: Languages and Tools for Formal Specifications*, Springer-Verlag, 1993. HALL, A., "Seven Myths of Formal Methods", *IEEE Software*, Septiembre 1990. LISKOV, B.H., y V. BERZINS, "An Appraisal of Program Specifications", publicado en *Software Specifications Techniques*; eds.: N. GEHANI y A.T. MCKITTRICK, Addison-Wesley, 1986, p. 3. SPIEVY, J.M., *The Z Notation: A Reference Manual*, Prentice-Hall, 1992. SPIEVY, J.M., *Understanding Z: A Specification Language and Its Formal Semantics*, Cambridge University Press, 1988. YOURDON, E., "Formal Methods", *Guerrilla Programmer*, Cutter Information Corp., Octubre 1994. AITCHISON, J., y GILCHRIST, A. *Thesaurus Construction, a Practical Manual*, London: ASLIB, 1987.

Una aplicación característica de extracción de conocimientos consta de 4 fases: (1) recogida de información para recolectar y cribar la información, (2) el reconocimiento de características para generar las características/metadatos, (3) el análisis de metadatos para la abstracción de los conocimientos y, (4) la presentación de los conocimientos a los usuarios.

La obligación de automatizar la generación de conocimientos con el fin de limitar la carga de trabajo de los humanos se ha visto intensificada según entramos en la edad de la información. Con la proliferación de nuevos productos y servicios la intensidad relativa al conocimiento de los productos y servicios (especialmente, del sector IT) genera cada vez más unos elevados costes en la metodología de evaluación y de cálculo de esfuerzos. El ciclo de vida del conocimiento sigue acortándose cada vez más. Además de lo anterior, también se acorta el plazo de tiempo disponible para la toma de decisiones y los encargados de dichas decisiones cada vez se encuentran más dispersos o en situación de movilidad constante, por lo que el tesoro actual viene a resultar indispensable dada su presencia en la WEB y su integración en el Laboratorio de Documentación Aplicada.



Clasificación de la Evaluación y la importancia relativa en Calidad de Información en el Tesoro

### El almacén de datos

Los almacenes de datos sirven como fuentes de información primaria una vez señalada la calidad terminológica de las fuentes y evaluada su fiabilidad y presencia constante como referencia, esta fiabilidad es básica para la actualización y reforma para su integración en unidades superiores de información y conocimiento como las Ontologías Web y las Sociedades de Conocimiento, en las que la información se

consigue por medio de agentes que, de forma activa, recogen, depuran, almacenan y actualizan la información. Entre las fuentes de información se incluyen los documentos, archivos, noticias, registros, transacciones y diversos tipos de bases de datos, así como niveles superiores de información en estos datos: peso, uso, vectores, trascendencia. Por lo general, el almacén de datos del tesoro de edición y tipografía contiene bases de datos relacionales e instrumentos de gestión, herramientas de análisis y de recuperación de los datos. Con ello se facilita que los usuarios dispongan de acceso fácil a una información que les aporta una perspectiva en cuanto a las tendencias y pautas del funcionamiento informacional y de su evolución en la interacción con otros usuarios. El almacén de datos sirve como un depósito para la recogida e integración de los datos procedentes de múltiples fuentes, aportando una estructura adecuada para su uso en la resolución de problemas con los usuarios más profesionales.

Los instrumentos disponibles en la actualidad nos han permitido construir bases de datos de información que contienen datos sólo de lectura que han sido ajustados o derivados a partir de múltiples fuentes. Estas bases de datos contienen la información inicial para las aplicaciones de apoyo a la toma de decisiones en su evolución a plataformas superiores<sup>40</sup>. A su vez, estas aplicaciones analizan con más detalle estos datos y los presentan en los formatos exigidos por los trabajadores del conocimiento. Los resultados que arrojan unas múltiples fuentes a partir de un gran número de formatos llevan a la obligación de centrarse en el proceso de gestión de la información. Por ello debíamos facilitar un proceso normalizado para la gestión de los requisitos relativos a la información, así como para determinar cuáles son los datos adecuados para su posterior uso:

La infraestructura del almacén de datos facilita la creación de las siguientes prestaciones a los trabajadores del conocimiento:

- La información disponible y la forma de acceder a ella.
- Las vinculaciones existentes entre la información estructurada y la no estructurada.
- El acceso a datos/información heterogénea.
- La extracción periódica de datos a partir de fuentes funcionales.
- La mejora y la actualización de los datos (creación de datos ajustados o derivados).
- La distribución y reproducción de los datos a múltiples emplazamientos.
- El apoyo a la infraestructura de integración y gestión.

### **Agentes de Conocimiento como intermediarios en captura informacional para Tesoros**

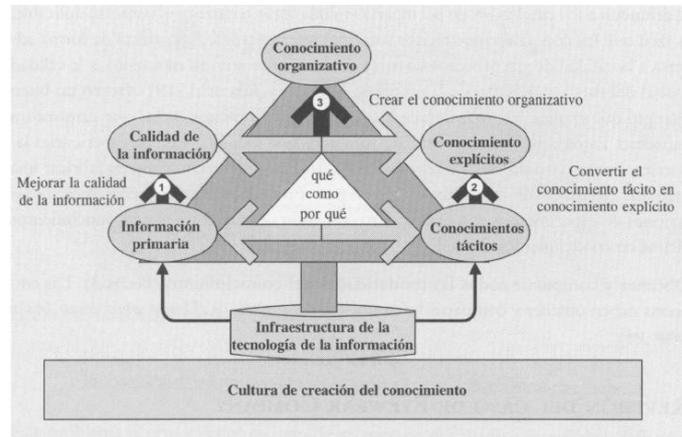
---

<sup>40</sup>HOARE, C.A.R., *Communicating Sequential Processes*, Prentice-Hall International, 1985. JONES, C.B., *Systematic Development Using VDM 2*. 2ª ed., Prentice-Hall, 1991. ROSEN, K.H., *Discrete Mathematics and Its Applications*, 3.- ed., McGraw-Hill, 1995. SPIEVY, J.M., *The Z Notation: A Reference Manual*, Prentice-Hall, 1992. SPIEVY, J.M., *Understanding Z: A Specification Language and Its Formal Semantics*, Cambridge University Press, 1988.

El auge del volumen de nuevos datos que nos llega a diario presenta un reto importante tanto para la actualización como para la integración de la información como conocimiento. Se calcula que la cantidad de nueva información generada se ha ido duplicando cada 20 años a lo largo del último siglo y que este ritmo se acelerará hasta duplicarse cada 10 años. La mayor parte de la información se presenta en forma de documentos. Es interesante destacar que la mayor dificultad con la que se topan los usuarios que buscan información en Internet o, incluso, en vía intranet, no radica en dar con los datos, sino en asimilar la tremenda abundancia de datos/información que recuperan. Esto se debe a la falta de mecanismos de medida de la calidad para la evaluación de los datos y de un posterior cribado para abstraer los más cruciales y obtener conocimientos comprensibles. Este problema repercute en la calidad de la información y puede llegar a afectar hasta a los usuarios de almacenes de información.

De forma reciente, los agentes de red han surgido como un nuevo mecanismo para la adaptación de las búsquedas, filtrando datos inconexos y extrayendo los conocimientos que resultan útiles para los usuarios de una página web. Un agente de red es una pieza de soporte lógico que ayuda al usuario a recoger y a procesar la información procedente de múltiples fuentes. De forma habitual, un agente de soporte lógico al uso dispone de la capacidad de adaptación personal de búsquedas de información, la capacidad dinámica necesaria para completar las tareas al margen de las actividades del usuario y la capacidad de adaptarse a partir de sus acciones previas y del entorno en el que trabaja.

Un agente de red es un agente de soporte lógico que dispone de movilidad en la red. Los agentes de red (u objetos de red) son objetos que disponen de un código (comportamiento), un estado de ejecución y un itinerario. Éstos están agrupados de forma lógica y se activan como una única unidad. Lo anterior contrasta con los objetos estacionarios que disponen de código (comportamiento) y datos. En ambos casos, el comportamiento viene dado por las diversas interfases, sin embargo, en el caso de los objetos estacionarios, dado que son estáticos y no dinámicos tanto en su composición como en los datos, pueden actuar como testigos de datos en un tiempo relativo, por lo menos, es comprensible hasta que alcancen la fase de obsolescencia. Eso sí, tanto el código como los datos pueden depender de una sola plataforma en la que el agente actúe, frente al agente de red que su dinamismo le llevan a desplazarse, por lo que el motor y su código, datos y estado de ejecución de itinerario informacional deben ser portátiles, actuando en su itinerario de búsqueda y recuperación como una muestra para el introductor y gestor de calidad que podrá así comprobar la eficacia y solidez de la información acumulada y capturada, como por ejemplo, a través de búsquedas repetitivas de información.



Fuente: *Cambridge Research Group*.

Creación de conocimiento a través de recuperación con agentes

### Propuesta de Tesauro

El objetivo de este Tesauro ha sido desde un primer momento elaborar un lenguaje controlado a partir de la terminología existente en el lenguaje natural. Se ha optado por su estructuración en entradas no facetadas, puesto que no se ha estimado que esta técnica se adapte a las características de este tesauro. Se puede ver un ejemplo de tesauro con estructura facetada en Hoffman (2006).

Las entradas o familias que se han establecido recogen diversos campos, actividades o procesos encontrados en los documentos consultados, y son las siguientes:

- Elementos tipológicos diacríticos
- Elementos tipográficos
- Signos diacríticos
- Abreviaciones
- Puntuación
- Corrección
- Maquetación
- Software
- Unidades de tipometría
- Partes de la letra
- Actividades de producción

Esta organización surge de las características de los términos encontrados, que se adaptaban a una entrada ya existente o requerían la creación de una nueva. Este

proceso es parte de la naturaleza dinámica de la creación de un lenguaje controlado, y ya ha sido explicado en el apartado de Recuperación terminológica.

### **Requisitos**

A la hora de su construcción, se han contemplado las condiciones que ha de cumplir un tesoro para que se considere como tal (Currás, 2005):

- Se trata de un lenguaje especializado
- Normalizado, mediante un proceso postcontrolado
- Las unidades lingüísticas adquieren la categoría de términos puesto que proceden de campos concretos del conocimiento
- Los términos o palabras clave se relacionan entre sí por relaciones de equivalencia, jerárquicas o asociativas
- Se trata de un lenguaje terminológico orientado a la recuperación de información

En su posterior desarrollo habrá que comprobar, no obstante, si permite la introducción o supresión de términos para evitar que quede obsoleto.

### **Clase de tesoro**

Los diferentes ámbitos del saber imponen sus características a los lenguajes controlados que pretenden estructurarlos, por lo que existen diferentes categorías o tipos de tesoros. El que se describe en estas páginas se caracteriza por ser:

Especializado, debido a que pretende abarcar un campo específico.

Multidisciplinar. Aunque podría verse la maquetación como un paso más en la búsqueda de la estética en la edición de textos, se ha considerado esta disciplina suficientemente relevante como para distinguirse de la tipografía, por lo que otorga a esta obra la categoría de multidisciplinar. Por otro lado, a pesar de que las abreviaciones son un fenómeno lingüístico más que tipográfico, no parecería razonable no incluirlas ya que aspectos como su ortografía atañen a esta obra.

Estructura alfabética y sistemática. Al contrario que la estructura gráfica, que se ha estimado irrelevante y trabajosa para el objetivo de esta obra, las estructuras alfabética y sistemática son imprescindibles en cualquier tesoro. En primer lugar se presenta la estructura sistemática, que en este caso es jerárquica. Tras ésta se encuentra la estructura alfabética, donde se incluyen también las relaciones entre los términos para una recuperación terminológica más rápida y eficiente, puesto que se incluye el **TG** al que pertenece el término, y sus **TE** y **TR** correspondientes, así como las posibles **NA**.

Bilingüe. Por último, se ha considerado la categoría a la que pertenece esta obra de acuerdo con los idiomas en que se presenta. La última sección del tesoro es un glosario bilingüe en el que se hallan las correspondencias español-inglés de los términos incluidos en la estructura sistemática. El proceso de traducción al inglés se explica más adelante.

### Caracterización de entrada

Para explicar la estructura del Tesoro se tomará como ejemplo una de las entradas de la presente obra.

Dado que se entiende que la unidad básica es el *término*, la elaboración de una entrada requerirá decidir las relaciones que se establecen entre ellos, así como determinar cuál será el término más amplio dentro del que se incluirán otros más específicos o relacionados. También, como es el caso en este ejemplo, puede suceder que encontremos un término y sea conveniente englobarlo en uno más general que actuará de Término de Cabecera (TC). Así pues, ante el ejemplo *Corrección*, se entiende que esta actividad está enmarcada dentro de un proceso más amplio que sería el *Trabajo editorial*. Se establece así una relación jerárquica partitiva entre estos dos términos, puesto que se considera que la corrección forma parte en el proceso de elaboración de una obra. Nos situamos entonces en el TC Trabajo editorial, al cual pertenecen diversos TE como Cálculo del texto, Composición, Casado, Corrección, etc.

Las Notas de Aplicación (NA) tienen dos funciones. O bien definen el concepto en cuestión, o bien aclaran a qué área o actividad pertenece el término, probablemente porque coincida en forma con otro concepto. Así en *Corrección* se ha usado la siguiente NA: Operación o conjunto de operaciones con que se trata de perfeccionar los textos y cada una de las partes que forman un libro o una publicación periódica (Martínez de Sousa, 2001).

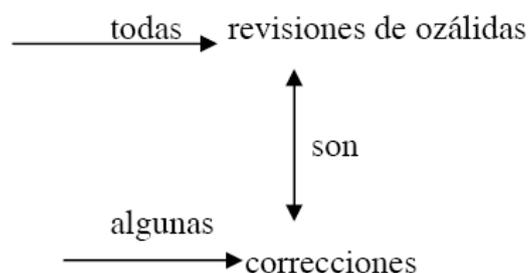
Las relaciones entre los términos se denominan relaciones semánticas, que pueden ser de tres tipos: de equivalencia, de jerarquía y de asociación.

Las relaciones de equivalencia, como su nombre indica, se dan al mismo nivel entre descriptores y no descriptores de un mismo concepto. De entre las distintas causas que normalmente dan lugar a esta relación, la más frecuente en esta entrada ha sido la de la de términos más actuales frente a otros ya obsoletos. No parece raro esto en un campo que ha sido modificado “recientemente” gracias a la tecnología informática. Así, *corrección de galeradas*, donde la galerada es la trozo de composición que cabe en una galera (Martínez de Sousa, 2001), ha caído en desuso a favor del término *corrección de primeras pruebas*, término que alude al lugar que toma la revisión en el proceso y no a una medida que actualmente es menos usada. Sin embargo, esto no debe llevar a eliminarlo de la clasificación puesto que su uso, aunque menor, sigue estando presente, y la inclusión del término como no descriptor puede ayudar en la recuperación de información por parte del usuario.

Otro tipo de relaciones de equivalencia menos frecuente pero que en este Tesouro, y en concreto en esta entrada, tiene una presencia significativa, es la correspondencia entre el nombre otorgado a cada una de las marcas de corrección, derivado de la objetivo que pretende, y la marca no verbal que lo representa, ambos empleados en el ámbito de la corrección tipográfica. Se ha preferido otorgar la categoría de término principal o preferente al concepto en forma de palabra en detrimento de la forma no verbal puesto que sería complicado para el usuario la recuperación de la información a partir de esta última.

Las relaciones de jerarquía (TG – TE) son relaciones asimétricas en las que un término es superior o genérico de otro. Pueden ser relaciones género-especie, parte-todo, o de ejemplo.

Se encuentran en esta entrada algunos ejemplos de relación género-especie, que tienen que ver con tipos de actividades. Se establece pues entre la actividad corrección y *corrección de primeras pruebas* o *revisión de ozálidas* entre otros. Cumple el criterio de que ambas pertenezcan a una misma categoría fundamental (actividades) y además cumple el criterio de comparación todos-alguno. Por ejemplo:



Sin embargo, se verá más abajo que el término pruebas ozálidas no podría constituir una relación jerárquica con correcciones, puesto que el primero es un elemento y el segundo una actividad.

No se encuentran relaciones jerárquicas partitivas puesto que ninguna de las relaciones existentes se amoldan a los cuatro grupos de éstas que pueden serlo: sistemas y órganos del cuerpo, localidades geográficas, ciencias y disciplinas o estructuras jerárquicas sociales.

Las relaciones de asociación (TR) son relaciones recíprocas (un término es relativo de otro y viceversa), son mucho más variadas que las anteriormente mencionadas. En general se han incluido aquí todos los términos que no cumplían las reglas para ser catalogados dentro de una relación jerárquica, teniendo en cuenta todos los requisitos existentes. Es el ejemplo de *correcciones* y *pruebas ozálidas*, que se incluiría en el Tesouro como una relación asociativa proceso – instrumento.

Otras relaciones presentes en la entrada que aquí se describe son:

- ocupación – persona que la ejerce: es el caso de *corrección de primeras pruebas y corrector de primeras pruebas*
- acción – sujeto paciente: en este caso podría citarse *corrección y erratas*
- actividad – propiedad del instrumento: aquí encontramos *corrección y canon de página*

TG Corrección

- TE Corrección de primeras pruebas
- TE Corrección de segundas pruebas
- TE Corrección de terceras pruebas
- TE Revisión de ozálidas
  - TR Pruebas ozálidas
- TE Corrección de concepto
- TE Corrección de estilo
- TE Corrección tipográfica
  - TR Marcas de Corrección

- ND Corrección de galeradas
- ND Corrección de compaginadas

TE Llamadas

ND

TE Suprimir

ND

TE Separar

ND

TE Unir	ND		
TE Cambiar orden	ND		
TE Insertar	ND		
TE Cursiva	ND		
TE Negrita	ND		
TE Caja alta	ND		
TE Caja baja	ND		
TE Versalita	ND		
TE Añadir tilde	ND		
TE Superíndice	ND		
TE Subíndice	ND		
TE Falta texto	ND		
TE Quitar sangría	ND		
TE Sangrar texto	ND		
TE Recorrido	ND Calle		ND
TR Lectura gramatical			
TR Lectura de maquetación			
TR Lectura ortotipográfica			
TR Corrección de autor			
TR Erratas			
TR Canon de página			ND Página de corrector
TR Evaluación de material			
TR Cálculo de matrices (Corrección)			
TR Corrector de primeras pruebas			ND Corrector de galeradas
TR Labor del corrector de primeras pruebas			
TE Limpieza			
TE Unificación			
TE Normalización			

### Traducción de descriptores y términos en el tesoro

El otro objetivo a la hora de elaborar este tesoro ha sido el de su ampliación a la categoría de tesoro bilingüe, puesto que actualmente no se ha desarrollado ninguno que relacione el español con otra lengua en lo relativo a la Tipografía.

Sin querer entrar en disquisiciones acerca del concepto de equivalencia, que competen a la Traductología, se puede decir que un tesoro *bilingüe* es un caso especial de tesoro monolingüe al que se le han añadido las equivalencias de los términos en una segunda lengua (Currás, 2005).

Dado que la labor de traducción de un tesoro requeriría por sí sola un artículo completo dedicado a ésta, lo que se pretende aquí es describir brevemente la metodología seguida con este fin.

### Materiales y métodos

Se ha optado de nuevo por las dos obras más exhaustivas en este tema que han servido también para la confección monolingüe del tesoro. Éstas son, *Typography & Graphic Design Thesaurus* y *Diccionario de edición, tipografía y artes gráficas*. La obra española ha sido empleada tanto para un uso de diccionario monolingüe como para el uso de bilingüe, ya que incorpora una sección de correspondencias francés/inglés-español.

Se ha seguido un método de trasvase terminológico en tres fases. En primer lugar se realiza la búsqueda del término en el diccionario u obra de consulta monolingüe de la LO. El siguiente paso es la búsqueda de equivalencias del mismo término en un diccionario bilingüe. Por último, se lleva a cabo la búsqueda del término elegido en la fase anterior en un diccionario u obra de consulta monolingüe de la LT. El sentido de los pasos 1 y 3 es sencillamente el de asegurar una equivalencia adecuada entre ambos términos o conceptos, ya que la búsqueda exclusiva en un diccionario bilingüe puede conducir a desajustes debidos a la multitud de acepciones que a veces posee un término o al criterio del lexicógrafo, que puede diferir del que aquí se sigue.

Las fichas utilizadas para el trasvase de la LO a la LT en tres pasos son las que siguen:

### Traducción directa

Diccionario monolingüe:	Diccionario de edición, tipografía y artes gráficas
Término:	Punto final
Definición:	Signo (.) que se emplea para cerrar una oración cuando tiene sentido completo o un período.

Diccionario bilingüe	Diccionario de edición, tipografía y artes gráficas
Equivalencias:	Period, full stop
Elección de equivalencia:	Period

Término:	Period
Diccionario monolingüe consultado:	<i>Typography &amp; Graphic Design Thesaurus</i>
Número de acepciones:	SN: In American usage, refers to the hash mark.
Sinónimos:	Full stop

Diccionario monolingüe:	Diccionario de edición, tipografía y artes gráficas
Término:	Bigote
Definición:	Filete ahusado o combinado con otros que se utiliza como adorno para separar dos partes de un texto

Diccionario bilingüe	Diccionario de edición, tipografía y artes gráficas
Equivalencias:	French rule
Elección de equivalencia:	French rule

Término:	French rule
Diccionario monolingüe consultado:	<i>Typography &amp; Graphic Design Thesaurus</i>
Número de acepciones:	SN: Double rules divided by diamonds
Sinónimos:	---

Diccionario monolingüe:	Diccionario de edición, tipografía y artes gráficas
Término:	Pica
Definición:	Unidad de medida tipográfica anglonorteamericana, de 12 puntos como el cícero didot o europeo, pero basada en la pulgada inglesa, ligeramente más pequeña que la europea.

Diccionario bilingüe	Diccionario de edición, tipografía y artes gráficas
Equivalencias:	Pica [EUA], em, eme [RU]
Elección de equivalencia:	Pica

Término:	Pica
Diccionario monolingüe consultado:	Merriam Webster Online
Número de acepciones:	1 : 12-point type 2 : a unit of about 1/6 inch used in measuring typographic material 3 : a typewriter type providing 10 characters to the linear inch and six lines to the vertical inch
Sinónimos:	---

Podría considerarse como una cuarta fase la que se ha realizado en las ocasiones en la que los resultados obtenidos no eran completamente satisfactorios. Este procedimiento ha consistido en la contrastación del resultado-término obtenido con glosarios, enciclopedias, y manuales. Ejemplos de esto son:

- *Manual de edición y autoedición* (Martínez de Sousa, 1994)
- *Breve glosario trilingüe sobre edición, producción editorial, derechos y contratos* (Teresa Gómez Mascaraque, 1996)
- *Multilingual Glossary for Art Librarians* (Sheridan, 1996),

### Problemas traductológicos

El ámbito de aplicación tan restringido de este tesoro ha hecho que muchos glosarios, diccionarios y enciclopedias queden descartados, total o parcialmente,

para su uso en la traducción siguiendo el método que se ha elegido. Esto ocurre por ejemplo con casi cualquier diccionario bilingüe español-inglés no especializado u otros diccionarios monolingües de cualquiera de las dos lenguas. Se puede hacer una sencilla prueba con el *Merriam Webster Online* (2005), en el que al ser remitidos al término inglés *widow*, aparece entre sus acepciones “a single usually short last line (as of a paragraph) separated from its related text and appearing at the top of a printed page or column”. Sin embargo, al buscar un término tan cercano semánticamente al anterior, *orphan*, la búsqueda es infructuosa.

No obstante, este mismo diccionario ha servido en algunos casos, como es el del término *pica* que se incluye entre los ejemplos de traducción. Con ello se demuestra que si bien su uso no es primordial, tiene una utilidad parcial en la justificación de una determinada traducción de un término, especialmente cuando se trata de términos de uso “más común”.

Otro problema de índole metodológica, puesto que invierte el proceso, pero de gran utilidad en cuanto a lo pragmático, surge a partir del hecho de que una de las fuentes usadas para la traducción (*Typography & Graphic Design Thesaurus*, 2006), constituye también, como se ha visto anteriormente, una fuente de extracción terminológica para el tesoro en español. Así pues, sucede que en la recuperación a partir del citado tesoro es necesaria su traducción al español para que sea incluido, y esto conduce a tener solventado el problema de traducción.

### Traducción inversa

Diccionario monolingüe:	<i>Typography &amp; Graphic Design Thesaurus</i>
Término:	Copyfitting
Definición:	Method of calculating the amount of space (pages, lines, etc.) a text will require

Diccionario bilingüe	Diccionario de edición, tipografía y artes gráficas
Equivalencias:	Cálculo tipográfico
Elección de equivalencia:	Cálculo tipográfico

Término:	Cálculo tipográfico
Diccionario monolingüe consultado:	Diccionario de edición, tipografía y artes gráficas
Número de acepciones:	Determinación de la cantidad de texto que cabrá en un documento o en una de sus partes teniendo en cuenta las características de la composición: tipo, ojo, cuerpo, medida, etc.
Sinónimos:	Calibrado