

*Theo Bothma*

## **Reflexiones sobre la interrelación entre lexicografía y ciencia de la información: una perspectiva personal**

### **1. Introducción**

La ciencia de la información, por definición, trata de datos, información y conocimiento, además de la interrelación entre estos conceptos que se definen de muchas maneras diferentes. En esta contribución no voy a centrarme en las diferentes definiciones de la disciplina. La ciencia de la información abarca varias subdisciplinas como el manejo de la información y el conocimiento, la organización y extracción de información, el comportamiento y búsqueda de información, la legislación, ética y economía de la información, los sistemas de información indígena y conocimiento indígena, etc.

Lo que es común a todas estas disciplinas es que en todas ellas se estudia la interrelación entre datos/información/conocimiento, humanos y tecnología (de información y comunicación). En algunos casos se resalta en la interrelación entre todos los componentes, en otros casos solo entre dos elementos, por ejemplo la interrelación entre humanos e información, o entre tecnología e información. Humanos se refieren, típicamente, a los usuarios finales, es decir, a las personas que finalmente usan la información para su trabajo, ocio, estudios, etc. Tecnología se refiere, en la mayoría de los casos, a las tecnologías de información y comunicación en su forma moderna de e-tecnología, pero la información solía, y todavía puede, organizarse y manejarse etc. en formatos basados en papel, entre otros.

Cada subdisciplina de la ciencia de la información tiene un número de diferentes teorías que pueden discutirse detalladamente en la literatura, a menudo bastante enconadamente sin mucho consenso. La teoría suele relacionarse con la práctica; las teorías de la ciencia de la información son generalmente, o bien reflexiones sobre cómo se puede manejar, almacenar, extraer etc. información y conocimiento en la práctica, procurando sistematizar la práctica, o bien reglas abstractas que reflejan la práctica y pueden generalizarse para su uso en otras situaciones.

En esencia, la lexicografía también trata de datos, información y conocimiento (aunque la definición de estos conceptos puede diferir de los de la ciencia de la información), humanos y tecnología, y la interrelación entre estos tres componentes. Básicamente, los datos pueden ser datos sobre palabras y palabras en combinación con otras palabras (por ejemplo, colocaciones, expresiones fijas etc.). Como en la ciencia de la información, los humanos son típicamente usuarios finales, y la tecnología empleada puede ser tecnología “antiguada” (papel) o e-tecnología. Hay una serie de teorías de lexicografía, y dentro de estas teorías hay distintos enfoques (todo esto a menudo llamados metalexografía). Sin embargo, también hay lexicógrafos que niegan rotundamente que exista una teoría de la lexicografía argumentando que la lexicografía solo es práctica, en particular la práctica de compilar diccionarios.

A partir de esta breve discusión resulta evidente que hay una serie de puntos comunes entre la ciencia de la información y la lexicografía, lo que recientemente ha sido señalado por varios investigadores. Esto puede implicar que la ciencia de la información pueda aprender de la lexicografía a nivel teórico y práctico, y viceversa, aunque es evidente que los puntos comunes no se dan en todas las subdisciplinas de la lexicografía y de la ciencia de la información.

En ambas disciplinas, el posible uso de los datos por parte de los usuarios finales constituye la razón por la cual los datos están organizados; en ambas disciplinas, los datos son extraídos por los usuarios finales; y en ambas, especialmente ahora, las e-tecnologías, es decir, las modernas tecnologías de información y comunicación, juegan un papel importante en la organización y extracción de los datos. Las dos disciplinas emplean ordenadores –desde su aparición en la década de los 80– para organizar y extraer datos/información; ambas han pasado de los sistemas originales basados en papel a sistemas digitales; ambas usan sistemas de bases de datos para organizar los datos y algoritmos de búsquedas para extraer los datos.

En el Centro de Lexicografía de la Universidad de Aarhus, Henning Bergenholtz, Sven Tarp y sus colegas han llevado a cabo una gran investigación para desarrollar la Teoría Funcional de la Lexicografía y compilar diccionarios, especialmente monofuncionales, dentro de este marco teórico. La teoría se centra en las necesidades de información de los usuarios finales en una situación específica, es decir, en situaciones comunicativas, cognitivas, operativas e interpretativas. Todo esto se explica detalladamente en el libro de Sven Tarp (2008) y en muchas otras publicaciones de este y otros autores, y también se resume en este número de ELex (Tarp 2015). El Centro ha creado varias plataformas para un debate sumamente crítico como lo describe Rufus Gouws en su contribución a este número (Gouws 2015).

## **2. Aarhus y Pretoria**

Trabajo en el ámbito de la ciencia de la información. Mi campo específico de interés es la organización y extracción de información (y cómo se aplican estos conceptos a la competencia de información y la publicación web y, en la actualidad, también en la e-lexicografía), especialmente relacionado a las necesidades del usuario final. En la ciencia de la información, las necesidades de información no están subdivididas en las cuatro funciones mencionadas con anterioridad. Las necesidades del usuario final y su uso de la información no son, en general, el centro de atención de la investigación, y la necesidad de información puede tratarse de una tarea de trabajo, el ocio, propósitos cognitivos, etc.

Desde hace muchos años estoy convencido de que la información debe estar disponible por encargo, y de que debe ofrecerse al usuario final únicamente la información relevante para sus necesidades específicas en una situación dada y a un nivel de complejidad y detalle deseado por aquel. Ya hace años discutí esta problemática en varias publicaciones (ver, p.ej., Claasen & Bothma 1988, Bothma 1990, 1992a y 1992b).

En 2010 me ofrecieron la oportunidad de demostrar en la práctica cómo la información por encargo y distintas rutas de acceso a exactamente la información requerida pueden

operacionalizarse en una publicación digital. En 2009, la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) me invitó a dirigir un proyecto para el Informe Mundial de la IFLA 2010 (Bothma 2010) y a desarrollarlo –a diferencia de mis anteriores publicaciones basadas en papel– como una publicación interactiva web con múltiples rutas de acceso y presentación de datos por encargo, basado en opciones extensivas de filtrado y diferentes opciones de presentación.

Basado en mi propio interés y experiencia, en 2010 Henning Bergenholtz me invitó a pasar tres meses como catedrático visitante en el Centro de Lexicografía. Una condición para mi visita fue que presentara una ponencia en una Conferencia prevista en Valladolid en 2010 y que presentara esta ponencia de antemano a mis colegas del Centro. Obedientemente, leí cuánto pude de los artículos publicados por los investigadores del Centro y otros colegas que trabajan con la misma teoría, tanto como otras publicaciones que tratan del desarrollo práctico en el campo de la e-lexicografía, para preparar mi propio artículo antes de viajar a Aarhus. Cuando leí esas publicaciones descubrí que la teoría funcional se vinculaba fuertemente a mis propias convicciones e investigaciones previas aunque nunca había intentado formalizarlas en una teoría coherente ya que los principios subyacentes se basan en las doctrinas básicas de hipertextos e hipermedias.

Al llegar a Aarhus, me invitaron a hacer mi presentación, y esta fue la primera vez que experimenté las discusiones críticas a las que Rufus Gouws se refiere en su contribución de esta edición de *Elex* (Gouws 2015). Presenté el título de mi ponencia propuesta en mi primera diapositiva, y en seguida se produjo una profunda discusión de unos 20 minutos sobre el mismo título y las palabras empleadas en el título. Lo que estaba previsto como una presentación de 40 minutos se convirtió en una presentación-más-discusión de más de horas y 45 minutos, distribuidos en dos sesiones.

Para mí, todo esto fue, definitivamente, un bautismo de fuego, pero me dio una buenísima percepción de cómo los colegas del Centro enfocaba cualquier problema de investigación, o sea, estimulando una discusión detallada y profunda del tema, nunca apuntada hacia alguien, sino sencillamente definiendo y redefiniendo los temas en cuestión para asegurarse de que se consiga un nivel aceptable de rigor académico. Huelga decir que el resultado de la discusión fue que tuve que reescribir o adaptar la mayor parte de mi ponencia para la Conferencia de Valladolid. Puede que la ponencia aún no cumpliera con todos los requerimientos planteados por los colegas del Centro, pero aun así acabó por ser publicado en el libro que recoge las contribuciones defendidas y analizadas en aquella Conferencia (Bothma 2011).

Esta experiencia dio la tónica para todas mis posteriores interacciones and colaboraciones con los colegas del Centro de Lexicografía: discusiones extensas y relevantes sobre cualquier aspecto de cualquier publicación o ponencia en las que colaboramos.

Siempre he considerado que estas discusiones son extremadamente valiosas y creo que he aprendido, por medio de ellas, a ser mucho más crítico y cuidadoso en el momento de formular mis ideas y pensamientos. Muchas veces tuve que cambiar mis ideas cuando mis colegas me convencieron, otras veces conseguí convencerles de que yo tenía razón, y algunas veces concluimos que era mejor discrepar. Sin embargo, todo esto me enseñó que debo formular mis ideas con más claridad y argumentar de forma más lógica. También pude emplear estas destrezas en mi

colaboración con otros colegas de Pretoria, Stellenbosch y Hildesheim, entre otros lugares, en proyectos relacionados con la e-lexicografía y también en otros campos de interés, por ejemplo en los proyectos de competencia de información en que he participado.

Desde mi primera visita a Aarhus en 2010, he visitado el Centro de Lexicografía varias veces, y tanto Henning Bergenholtz como Sven Tarp han venido a trabajar en la Universidad de Pretoria en varias ocasiones como catedráticos visitantes. La Universidad de Pretoria también reconoce la interrelación entre lexicografía y ciencia de la información, y Henning Bergenholtz ha sido nombrado Catedrático Extraordinario en el Departamento de Ciencia de la Información.

### **3. E-lexicografía y tecnología de información**

En colaboración con mis colegas del Centro de Lexicografía de Aarhus (especialmente Henning Bergenholtz y Sven Tarp), de la Universidad de Hildesheim (especialmente Ulrich Heid), de la Universidad de Stellenbosch (Rufus Gouws) y la Universidad de Pretoria (Danie Prinsloo), he participado en varios proyectos que específicamente tratan de las tecnologías para realzar los diccionarios en línea, proyectos que han resultado en varias publicaciones y ponencias en conferencias, y también en una investigación que todavía sigue.

Mis contribuciones a esta investigación se centran en las tecnologías para diseñar las bases de datos, principios de hipermedias adaptivas y filtrado de datos, y también en adiciones innovadoras para apoyar al usuario de diccionarios en línea. Todas estas contribuciones se han escrito en el marco de la teoría funcional de la lexicografía.

Si un diccionario en línea ha sido desarrollado en el marco de la teoría funcional, el usuario final tendrá típicamente una serie de opciones para acceder a los datos, basadas en los tipos de necesidades de información definidas por la teoría. Esto significa que el lexicógrafo ha decidido que un conjunto predefinido de datos se visualizará en la pantalla –si un usuario necesita información para una necesidad de información relacionada con la producción de textos– y que otro conjunto distinto se visualizará si el mismo usuario tiene una necesidad de información relacionada con la producción de textos.

Muchos de los diccionarios compilados por el Centro de Lexicografía siguen este paradigma (ver, p.ej., Tarp 2008, Nielsen et al. 2013, Fuertes-Olivera & Tarp 2014, Tarp 2015 y las referencias contenidas en estas publicaciones, además de los diccionarios del Centro accesibles en [Ordbogen.com](http://Ordbogen.com)).

Ya en 2009, Tarp propuso que las necesidades del usuario volvieran a redefinirse:

- “la lexicografía científica estaría antes de nada interesada en saber en qué situaciones –p.ej. recepción y producción– estas necesidades pueden producirse. Luego se plantearía la tarea de descubrir las necesidades de los usuarios en el restante 20 por ciento de las consultas, es decir, en una de cada cinco consultas. Y no se pondría contenta con esto pues intentaría profundizar

aún más en el problema para descubrir las necesidades que solo ocurren en una de cada cien o una de cada mil consultas [...]” (Tarp, 2009b:292).

- “[...] o aún más raramente, para poder concebir diccionarios capaces de satisfacer todas las necesidades de los usuarios en situaciones específicas,” (Tarp, 2009b: 292).
- “[...] artículos dinámicos [...] estructurados de diferentes formas de acuerdo con cada tipo de criterio de búsqueda”. “[...] artículos especialmente adaptados [...]”; “[...] definir su propio perfil [...]”; “[...] una dirección será probablemente la ‘individualización’ del producto lexicográfico adaptándolo a las necesidades concretas de un usuario concreto [...]” (Tarp, 2009a: 57, 59, 61).

Estas citas indican claramente que Tarp se imaginaba refinamientos ulteriores de la presentación de datos por encargo. He intentado, p.ej. en Bothma (2001), mostrar las tecnologías de las que actualmente se disponen para que estas ideas sean viables, tanto para limitar los datos mediante filtrado y perfilado como para ayudar al usuario a obtener más datos apoyando situaciones específicas de uso.

Toda esta investigación se apoya en discusiones profundas entre colaboradores que a menudo resultan en varios subproyectos adicionales. En todas las publicaciones se combinan conocimientos procedentes de la lexicografía y de la ciencia de la información (y a veces también conocimientos de la lingüística y del procesamiento de lenguajes naturales) con la esperanza de ofrecer un sustrato teórico con el que podamos construir nuevos y mejores tipos de diccionarios en línea entendidos como herramientas de información. El objetivo es que los usuarios más fácilmente puedan encontrar información relevante a un nivel de complejidad y detalle idóneo para sus necesidades específicas de información.

Las cuestiones investigadas incluyen:

- Presentación adaptada a las necesidades y algunas implicaciones de esto; ver Bergenholtz & Bothma (2011) y Bothma & Bergenholtz (2013).
- Un experimento práctico en la creación de múltiples diccionarios monofuncionales de expresiones fijas en afrikáans, basados en la teoría funcional; ver, p.ej., Bergenholtz, Bothma & Gouws (2011).
- El papel que puede jugar la teoría de relevancia de la ciencia de la información en la lexicografía; ver Bothma & Tarp (2012, 2014).
- Evaluación de los e-diccionarios conectados a los e-libros de Kindle para apoyar la recepción de textos; ver Bothma & Prinsloo (2013).
- Cómo los datos de corpus procesados y sin procesar pueden añadir valor por encargo para los usuarios de diccionarios en línea; ver, p.ej., Heid, Prinsloo & Bothma (2012) y Prinsloo et al. (2012).

- Cómo las herramientas de apoyo adicional pueden integrarse en los e-diccionarios para ofrecer apoyo por encargo a los usuarios para resolver necesidades de información sobre estructuras gramaticales complejas; ver, p.ej., Prinsloo, Bothma & Heid (2014).
- Incluso hemos aplicado la teoría funcional para evaluar un diccionario del siglo 18; ver Tarp & Bothma (2013). La colaboración investigadora continúa en la mayoría de estos proyectos, y nuevos proyectos se incorporan de forma regular. El único factor limitador de la colaboración es la falta de suficiente tiempo y personas para atender todos los desarrollos fascinantes que surgen a partir de nuestra colaboración y discusiones críticas. La colaboración con los colegas del Centro de lexicografía de Aarhus fue para mí el catalizador que abrió la puerta a muchas nuevas e interesantes posibilidades para poder aplicar el conocimiento teórico de la organización y extracción de información a un campo nuevo, la e-lexicografía. También tuve la gran suerte de poder extender mi investigación colaborativa a varios otros colegas del campo de la e-lexicografía. Considero que es muy ilusionante y un verdadero privilegio poder participar en este desarrollo.

#### 4. Conclusiones

Nuestra investigación, y una serie de otras publicaciones, han reforzado mi convicción de que la ciencia de la información y la lexicografía pueden aprender mucho la una de la otra, y de que la investigación inter- y multidisciplinaria en la que participamos conducirá a herramientas de información innovadoras.

Puede que una parte, incluso la mayor parte, de la investigación discutida con anterioridad no sea comercialmente viable de momento. Sin embargo, si los investigadores no experimentan con conceptos y tecnologías que no parecen realistas y viables en la actualidad se disminuirá la innovación con las herramientas de información relacionadas con internet. Este tipo de investigación imaginativa podría con el tiempo desembocar no solo en herramientas de información en línea gradualmente mejores sino en herramientas diferentes mediante innovación disruptiva (Bergenholtz, Bothma & Gouws 2011).

Por eso, todos los proyectos de investigación –y muchos otros– ofrecen plataformas para experimentar con tecnologías disruptivas, desarrollar prototipos y averiguar hasta qué punto estas tecnologías pueden añadir valor para el usuario ofreciéndole acceso a la información desde la perspectiva de sus necesidades específicas de información en una determinada situación de uso. Esta investigación imaginativa es absolutamente esencial para asegurar el desarrollo de tipos de herramientas no solo mejores sino distintas. Los usuarios podrán ayudar a ir mejorando los diccionarios en línea, pero solo la investigación fundamental en la metalexografía, las necesidades del usuario, las tecnologías de bases de datos y los principios de organización, acceso y extracción de información resultará en tipos diferentes de e-herramientas (Bergenholtz, Bothma & Gouws 2011).

Estas temas acentúan la naturaleza inter- y multidisciplinaria de la investigación que necesita realizarse en la intersección entre lexicografía y ciencia de la información, en conjunto con otras

disciplinas como la lingüística, el procesamiento de lenguajes naturales, la ciencia de computación, la interacción humano-computadora, el diseño de interfaces, etc.

## 5. Bibliografía

- BERGENHOLTZ, Henning y Theo J.D. BOTHMA (2011): «Needs adapted data presentation in e-information tools», *Lexikos 21*, pp. 53-77.
- BERGENHOLTZ, Henning, Theo J.D. BOTHMA y Rufus H. GOUWS (2011): «A model for integrated dictionaries of fixed expressions», en I. Kosem and K. Kosem, eds., *Electronic Lexicography in the 21st Century. New Applications for New Users. Proceedings of eLex 2011, Bled, 10-12 November 2011*. Bled, Trojina, pp. 34-42.
- BOTHMA, Theo J.D. (1990): «Teaching Biblical languages online», en *Proceedings of the 14th International Online Information Meeting, London, December 1990*. Oxford, Learned Information, pp. 97-104.
- BOTHMA, Theo J.D. (1992a): «Presenting grammatical information in Biblical information systems», *Proceedings of the 16th International Online Information Meeting, London, December 1992*. Oxford, Learned Information, pp. 163-172.
- BOTHMA, Theo J.D. (1992b): «Hypermedia as information model in studying Biblical languages and cultures», en *Proceedings of the Third International Colloquium: Bible and computers: Interpretation, Hermeneutics, Expertise, Tübingen, 26-30 August 1991*. Paris-Genève, Champion-Slatkine, pp. 329-350.
- BOTHMA, Theo J.D., ed., (2010): *IFLA World Report 2010*. (IFLA/FAIFE World Report Series vol. VIII.) The Hague: IFLA. [Online] Accesible en: <http://www.ifla-world-report.org>.
- BOTHMA, Theo J.D. (2011): «Filtering and adapting data and information in the online environment in response to user needs», en Pedro A. Fuertes-Olivera and Henning Bergenholtz, eds., *e-Lxicography: The Internet, Digital Initiatives and Lexicography*. London y New York, Continuum, pp. 71-102.
- BOTHMA, Theo J.D. y Henning BERGENHOLTZ (2013): «“Information needs changing over time”: a critical discussion», *SAJLIS 79(1)*, pp. 22-34.
- BOTHMA, Theo J.D. y Daan J. PRINSLOO (2013): «Automated dictionary consultation for text reception: a critical evaluation of lexicographic guidance in linked Kindle e-dictionaries», *Lexicographica 29*, pp. 165-198.
- BOTHMA, Theo J.D. y Sven TARP (2012): «Lexicography and the relevance criterion», *Lexikos 22*, pp. 86-108.
- BOTHMA, Theo J.D. y Sven TARP (2014): «Why relevance theory is relevant for lexicographers», *Lexicographica 30*, pp. 350-378.
- CLAASSEN, Walter T. y Theo J.D. BOTHMA (1988): «Structuring diverse types of information in Hypertext: the case of Biblical information», en *Online Information '88, Proceedings, vol. 1*. Oxford: Learned Information, pp. 83-90.
- FUERTE-OLIVERA, Pedro A y Sven TARP (2014): *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*. Berlin y Boston, De Gruyter.
- GOUWS, Rufus H (2015): «Towards a critical lexicographic discussion culture», este número de ELex.
- HEID, Ulrich, Daan J. PRINSLOO y Theo J.D. BOTHMA (2012): «Dictionary and corpus data in a common portal: state of the art and requirements for the future», *Lexicographica 28*, pp. 269-291.
- NIELSEN, Sandro, Pedro A. FUERTES-OLIVERA, Henning BERGENHOLTZ, Pablo GORDO GÓMEZ, Marta NIÑO AMO, Angeles SASTRO RUANO, Angel LOS RIOS RODICIO, y Marisol VELASCO SACRISTÁN (2013): «The Construction of Online Specialized Dictionaries: Theory and Practice», *Fachsprache 2013(1-2)*, pp. 23-41.

- PRINSLOO, Daan J., Ulrich HEID, Theo J.D. BOTHMA y Gertrud FAAß (2012): «Interactive, dynamic electronic dictionaries for text production», *Lexikos* 22, pp. 290-320.
- PRINSLOO, Daan J., Theo J.D. BOTHMA y Ulrich HEID (2014): «User support in e-dictionaries for complex grammatical structures in the Bantu languages», en A. Abel, C. Vettori and N. Ralli, eds., *Proceedings of the XVI EURALEX International Congress: The User in Focus. 15-19 July 2014*. Bolzano/Bozen: EURAC Research. [Online] Available [http://www.euralex.org/elx\\_proceedings/Euralex2014/euralex\\_2014\\_062\\_p\\_819.pdf](http://www.euralex.org/elx_proceedings/Euralex2014/euralex_2014_062_p_819.pdf).
- TARP, Sven (2008): *Lexicography in the borderland between knowledge and non-knowledge: General lexicographical theory with particular focus on learner's lexicography*, Tübingen, Niemeier.
- TARP, Sven (2009a): «Reflections on Data Access in Lexicographic Works», en Sandro Nielsen y Sven Tarp, eds., *Lexicography in the 21st Century. In honour of Henning Bergenholtz*. Amsterdam y Philadelphia, John Benjamins, pp. 43-62.
- TARP, Sven (2009b): «Reflections on Lexicographical User Research», *Lexikos* 19, pp. 275-296.
- TARP, Sven (2015): «La Teoría Funcional en Pocas Palabras», este número de ELex.
- TARP, Sven y Theo JD BOTHMA (2013): «An alternative approach to enlightenment age lexicography: The Universal Dictionary of Trade and Commerce», *Lexicographica* 29, pp. 222-284.