



## Journal of the Text Encoding Initiative

Issue 16 | 2023

Treinta años de TEI en español: usos y comunidad

---

# Treinta años de TEI en español: usos y comunidad

## Thirty years of TEI in Spanish: uses and community

Gimena del Rio Riande y Susanna Allés-Torrent

---



### Edición electrónica

URL: <https://journals.openedition.org/jtei/4871>

DOI: 10.4000/jtei.4871

ISSN: 2162-5603

### Editor

TEI Consortium

### Referencia electrónica

Gimena del Rio Riande and Susanna Allés-Torrent, "Treinta años de TEI en español: usos y comunidad", *Journal of the Text Encoding Initiative* [Online], Issue 16 | 2023, Online since 01 August 2023, connection on 07 June 2024. URL: <http://journals.openedition.org/jtei/4871> ; DOI: <https://doi.org/10.4000/jtei.4871>

---

The text only may be used under licence For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.. All other elements (illustrations, imported files) are "All rights reserved", unless otherwise stated.

---

# *Treinta años de TEI en español: usos y comunidad*

Thirty years of TEI in Spanish: uses and community

Gimena del Rio Riande y Susanna Allés-Torrent

---

## 1. Treinta años de TEI en español: usos y comunidad

- 1 En el campo de las Humanidades Digitales, pocas son las iniciativas que han demostrado tener un recorrido tan sólido y una comunidad de usuarios e interesados tan activos como la Text Encoding Initiative (TEI). En funcionamiento desde 1987, el estándar de codificación de la TEI ha sido utilizado para documentos escritos en lenguas tan diversas como el chino, el maya, el copto, el japonés, el español o el árabe, entre otras. Sus Guías Directrices “están dirigidas a cualquier persona que trabaje con cualquier tipo de recurso textual en formato digital.”<sup>1</sup> Es tal vez por ello que el alcance global de la TEI se vuelve a veces tan difícil de aprehender y definir.
- 2 En los países hispanohablantes, al igual que en otras latitudes, el interés creciente por la fijación de textos y su transformación en formato electrónico vino en un primer momento motivado por el análisis de textos más tradicional, como las concordancias, los diccionarios, o la creación de léxicos específicos. Indudablemente, muchos de los recursos desarrollados hasta ese momento vendrían de centros europeos o norteamericanos, como la tesis de doctorado de Leopoldo Sáez Godoy sobre

el léxico de Lope de Rueda, defendida en 1968 en la Universidad de Bonn, o la *Computerized version of the Dictionary of the Old Spanish Language* en el Hispanic Seminary of Medieval Studies (HSMS) de la Universidad de Wisconsin. Más tarde, la adopción del sistema Madison para la transcripción de manuscritos, desarrollado en el HSMS y ampliamente adoptado para textos medievales en España, colaboraría con la necesidad de echar mano de un estándar de codificación que fuera más allá de los formatos y que incluyera tanto información lingüística como semántica. En ese contexto, si bien la TEI surge en ámbito anglosajón en los años ochenta, no tarda en interesar a algunos investigadores activos en los estudios hispánicos durante la siguiente década.

- 3 En 1992, en el marco del Congreso de la Lengua Española celebrado en Sevilla, se organizó el panel “La lengua española y las nuevas tecnologías.”<sup>2</sup> En él se ofrecieron contribuciones que trataban temas todavía hoy esenciales como el procesamiento del lenguaje natural y la inteligencia artificial, diferentes ramas de la lingüística computacional y tecnologías lingüísticas (Jaime Carbonell, Ángel G. Jordán, Luis de Sopena), la creación de bases de datos tanto textuales como acústicos (Joaquín Llisterra, Dolors Poch), o de diccionarios electrónicos del español (Carlos Subirats Rüggeberg). Además, a la siempre presente actividad filológica y de crítica textual, mientras se hablaba de edición crítica de textos históricos (Germán Orduna, Alberto Varvaro), se discutían también los problemas de transcripción electrónica enfrentados en el Hispanic Seminary of Medieval Studies (David McKenzie), o la necesidad de estándares para el entonces pionero ADMYTE (Archivo Digital de Manuscritos y Textos Españoles) (Francisco A. Marcos Marín). En este congreso también se empezó a discutir –quizás por primera vez y gracias a Charles Faulhaber– sobre la TEI para textos en español. Surgían en ese momento problemas hoy superados, como cuestiones relativas al marcado tipográfico y codificación de caracteres (acentos, mayúsculas, minúsculas, etc.), y otros esenciales a los lenguajes de marcado (en ese momento, el Standard Generalized Markup Language o SGML) para distinguir entre aspectos formales y semánticos, así como el establecimiento de una codificación explícita y resolutiva para ambigüedades lingüísticas. No es un dato menor el hecho de que en este evento se presentó la traducción al español de las primeras Guías Directrices de la TEI, elaboradas entre 1988 y 1990. Las “Guidelines for the Encoding and Interchange of Machine-Readable Texts” (TEI P1), traducidas al español por Francisco Marcos Marín y las bibliotecarias argentinas Verónica Zumárraga y Marcela Tabanera, pueden entenderse hoy como claro ejemplo del temprano interés de la comunidad hispanohablante en este estándar.

- 4 En esta década encontramos también, de manera más aspiracional que real, la creación de ediciones hipertextuales, actividad que para el decenio siguiente florecería a un lado y al otro del Atlántico y reconocería como central la adopción de la TEI en tanto iniciativa internacional para la codificación y el intercambio de cualquier tipo de texto, independientemente de la lengua, época o procedencia.
- 5 Tomando como punto de partida ese congreso y el conjunto de intervenciones académicas que de ahí surgieron, este número especial lleva como título “Treinta años de TEI en español: usos y comunidad.” Creemos que, después de tres décadas de trabajo, aunque en muchos casos aislado y desfinanciado, pueden ya ofrecerse una serie de reflexiones sobre los derroteros que esta práctica digital ha tenido en los estudios hispánicos y que ha generado una red de colaboraciones y una comunidad de práctica con un conjunto de características propias.
- 6 Las seis contribuciones que aquí se publican tienen como hilo conductor el uso de la TEI para la codificación y edición filológica de textos en español y aspiran a ser una muestra representativa del trabajo que se viene llevando adelante en este ámbito en los diferentes hemisferios.
- 7 Cuatro contribuciones se sitúan en el ámbito de la edición filológica digital, pero lo hacen desde perspectivas ecdóticas, contextos de investigación, épocas y géneros literarios, y audiencias diferentes. Gabriel Calarco propone el uso de la TEI para una edición digital de fragmentos de un texto medieval, el *Libro de Alexandre*, en el marco de su tesis doctoral llevada a cabo en la Universidad de Buenos Aires y siguiendo los principios de la “minimal computing.” Antonio Rojas, Grisel Terrón, Alaina Solernou y Tobias Kraft ofrecen un panorama general sobre su proyecto editorial Humboldt Digital, una iniciativa internacional sostenida conjuntamente por instituciones alemanas y cubanas, que ha dado lugar a una plataforma en línea donde se están publicando los escritos de Humboldt relacionados con su estancia en la isla de Cuba. Marta López Izquierdo aborda la edición del género epistolar ofreciendo en primer lugar un informativo panorámico sobre iniciativas de investigación internacionales afines, y trayendo a colación retos y soluciones adoptados en el proyecto “Cartas de Republicanos Españoles Refugiados y exiliados en Francia” (CAREXIL-FR). Desde un abordaje pedagógico, la contribución de Clayton McCarl y su equipo –Lyn Hemmingway, Stacey Lowey-Ball, Sara Menéndez y Georgina Wilson Aranguren– se

sitúa en un contexto académico de estudiantes de grado, concretamente en Estados Unidos, en el que la edición digital se concibe como una herramienta y una actividad de enseñanza que sirve para la apropiación y exploración de una serie de textos coloniales.

- 8 Las dos contribuciones restantes ofrecen puntos de vista alternativos pero complementarios a los anteriores. Por un lado, José Calvo Tello y Nanette Rißler-Pipka tratan el problema del uso de textos digitales publicados con algún tipo de derecho de autor y que por consiguiente no pueden ser reutilizados libremente. La solución propuesta es la creación de datos derivados en TEI. Los autores exponen algunas estrategias útiles para tal fin siguiendo la filosofía de los criterios FAIR (en sus siglas en inglés, Findable, Accesible, Interoperable, Reproducible). A modo de cierre, las editoras de este volumen, Susanna Allés-Torrent y Gimena del Rio Riande, ofrecen un panorama genérico sobre el uso de la TEI en el mundo hispanohablante. La mayoría de los resultados obtenidos y discutidos en el artículo proceden de una encuesta que se llevó a cabo durante 2022 gracias a la financiación del proyecto “Communicating the TEI to a Multilingual User Community” por parte de la Mellon Foundation.
- 9 El *Journal of the Text Encoding Initiative* y las editoras invitadas apuestan con este primer número especial en español por el multilingüismo y la apertura a diferentes comunidades académicas, con la esperanza de que se dé la bienvenida a otros números en diferentes lenguas y que la revista ejerza así una función de observatorio real de los usos de la TEI a nivel global.

## 2. Thirty years of TEI in Spanish: uses and community

- 10 In the field of the Digital Humanities, few initiatives have such a well grounded background and an active community of users and practitioners as the Text Encoding Initiative (TEI). Active since 1987, the TEI encoding standard has been used for documents written in languages as diverse as Chinese, Mayan, Coptic, Japanese, Spanish, and Arabic, among others. Its Guidelines “are addressed to anyone who works with any kind of textual resource in digital form.”<sup>3</sup> This is perhaps why the global scope of TEI is so difficult to grasp and define.
- 11 In the Spanish speaking countries, as well as in other latitudes, the growing interest in text standardization and its transformation into electronic format arrived initially motivated by traditional text analysis, such as concordances, dictionaries, or the creation of specific lexicons. Undoubtedly, many of the resources developed so far would come from European or North

American centers, such as Leopoldo Sáez Godoy's doctoral thesis on the lexicon of Lope de Rueda, defended in 1968 at the University of Bonn, or the *Computerized version of the Dictionary of the Old Spanish Language* at the Hispanic Seminary of Medieval Studies (HSMS) at the University of Wisconsin at Madison. Afterwards, the adoption of the Madison system for manuscript transcription, developed at the HSMS and widely adopted for medieval texts in Spain, would insist on the need for an encoding standard that needed to go beyond formats and include both linguistic and semantic information. In this context, although the TEI emerged in the Anglophone world in the late 1980s, it did not take long for some researchers active within Hispanic studies to become interested in it during the following decade.

- 12 In 1992, within the framework of the Congreso de la Lengua Española held in Seville, a group of participants organized the panel "The Spanish language and new technologies."<sup>4</sup> Some contributions dealt with topics that are still essential today, such as natural language processing and artificial intelligence, different branches of computational linguistics and linguistic technologies (Jaime Carbonell, Ángel G. Jordán, Luis de Sopeña), the creation of both textual and acoustic databases (Joaquín Llisterra, Dolors Poch), or electronic dictionaries of Spanish (Carlos Subirats Rüggeberg). In addition to the ever-present interest for philology and textual criticism in these type of conferences and popular discussions about critical editing of historical texts (Germán Orduna, Alberto Varvaro), issues emerged surrounding electronic transcription at the Hispanic Seminary of Medieval Studies (David McKenzie), along with the need for standards for the then pioneering ADMYTE project (Archivo Digital de Manuscritos y Textos Españoles) (Francisco A. Marcos Marín). At this conference, discussions began to take place -perhaps for the first time and thanks to Charles Faulhaber- about the Text Encoding Initiative for texts in Spanish. The challenges that arose at that time have already been overcome nowadays, such as issues related to typographic markup and character encoding (accents, capitalisation, lowercase, etc.). Others, instead, pointed to essential notions of markup languages (at that time, the Standard Generalized Markup Language or SGML) to distinguish between formal and semantic textual aspects, the establishment of an explicit encoding, and the resolution of linguistic ambiguities. It is important to note that for the first time the translation in Spanish of the TEI Guidelines, drafted between 1988 and 1990, was presented at this event. "The Guidelines for the Encoding and Interchange

of Machine-Readable Texts” (TEI P1), translated into Spanish by Francisco Marcos Marín and the Argentinean librarians Verónica Zumárraga and Marcela Tabanera, demonstrate the enormous early interest of the Spanish-speaking community for this standard.

- 13 In this decade we also find, more as an aspiration than a reality, the creation of hypertext editions, an activity that would flourish in the following decade on both sides of the Atlantic. Hence, the adoption of the TEI would be recognised as central, because it provided an international initiative for the codification and exchange of any type of text, regardless of language, time or origin.
- 14 Taking that conference and the group of academic contributions that emerged from it as a starting point, this special issue is entitled “Thirty years of TEI in Spanish: uses and community.” We believe that, after three decades of work, albeit in many cases isolated and unfunded, we can now offer a set of reflections on the directions that this digital practice has taken in Hispanic studies. Also, we aim to highlight the networks of collaborations and the distinct community of practice that have appeared during these years.
- 15 The six contributions published here focus on the use of the TEI for the encoding and scholarly editing of Spanish texts. They aim to be a representative sample of the work being done in this field in the different hemispheres.
- 16 Four contributions are situated within the field of digital scholarly editing, but they do so from different ecdotal perspectives, research contexts, literary periods and genres, and are conceived for different audiences. Gabriel Calarco proposes the use of the TEI for a digital edition of fragments of an Iberian medieval text, the *Libro de Alexandre*. His work is done within the framework of his doctoral thesis carried out at the University of Buenos Aires and follows the principles of “minimal computing.” Antonio Rojas, Grisel Terrón, Alaina Solernou and Tobias Kraft give an overview of their editing project Humboldt Digital, an international initiative jointly supported by German and Cuban institutions, which offers an online platform where Humboldt’s writings related to his stay on the island of Cuba are published. Marta López Izquierdo addresses the editing of the epistolary genre by, first, offering an informative overview of related international research initiatives, and bringing up challenges and solutions adopted in her project “Letters of Spanish Republican Refugees and Exiles in France” (CAREXIL-FR). From a pedagogical approach, the contribution by Clayton McCarl and his team -Lyn Hemmingway, Stacey Lowey-Ball, Sara Menéndez, and Georgina

Wilson Aranguren- is situated in an academic context of undergraduate students, specifically in the United States, in which digital editing is conceived as a teaching device and practise that serves for the appropriation and exploration of Spanish colonial texts.

- 17 The two remaining contributions offer alternative but complementary points of view to the previous ones. On the one hand, José Calvo Tello and Nanette Reißler-Pipka address the problem of using digital texts published under some form of copyright and which therefore cannot be freely reused. The proposed solution is the creation of derived data in TEI. The authors outline some useful strategies for this purpose following the philosophy of the FAIR criteria (Findable, Accessible, Interoperable, Reproducible). In closing, the editors of this volume, Susanna Allés-Torrent and Gimena del Rio Riande, offer a general review of the TEI use in the Spanish-speaking world. Most of the results obtained and discussed in the article come from an online survey carried out during 2022 thanks to the funding of the project “Communicating the TEI to a Multilingual User Community” by the Mellon Foundation.
- 18 With this first special issue in Spanish, the *Journal of the Text Encoding Initiative* (JTEI) and the guest editors of this issue commit to multilingualism and opening the dialogue to different scholarly communities, in the hope that the journal will welcome other issues in different languages and will thus serve as a real observatory of the uses of the TEI at a global scale.

## NOTAS

1 Fecha de consulta 16 de julio de 2023, <https://www.tei-c.org/Vault/P5/4.6.0/doc/tei-p5-doc/en/html/AB.html>.

2 Fecha de consulta 16 de julio de 2023, <https://cvc.cervantes.es/obref/congresos/sevilla/indice.htm>.

3 Accessed July 16, 2023, <https://www.tei-c.org/Vault/P5/4.6.0/doc/tei-p5-doc/en/html/AB.html>.

4 Accessed July 16, 2023, <https://cvc.cervantes.es/obref/congresos/sevilla/indice.htm>.



## AUTORS

### **GIMENA DEL RIO RIANDE**

Investigadora Adjunta. IIBICRIT, CONICET

### **SUSANNA ALLÉS-TORRENT**

Profesora. Department of Modern Languages and Literatures de la University of Miami (Fl, Estados Unidos)