¿Limpia, fija o da esplendor? El dilema de Mr. Proper en la era digital

Jorge Muriel Martín
Folkuniversitetet Göteborg (Suecia)

INTRODUCCIÓN

Las redes digitales constituyen un elemento clave para el desarrollo cultural y lingüístico de los diferentes grupos sociales que en ella se desenvuelven. El propio carácter de la información digital y de las tecnologías de la información y la comunicación (TIC) presenta varias ventajas de valor incalculable para el diseño de estudios lingüísticos y, en especial, dentro del campo denominado *Lingüística computacional* o *Programación de lenguajes naturales*. Este campo presenta aplicaciones asequibles a la mayoría de miembros de la comunidad de profesionales de las humanidades, que pueden derivar, entre otros, en nuevos y más eficientes métodos para la enseñanza del español como lengua extranjera (ELE).

Este artículo tiene por objetivo llamar la atención sobre estas interesantes aplicaciones, al tiempo que propone un crítico repaso histórico a las circunstancias que justificaron la creación de las redes digitales. El dilema de Mr. Proper se revela, por tanto, como una metáfora, cuyo fin es llamar la atención sobre el hecho de que, hoy día, la discusión no debe estar en si los usos de una lengua en la red son correctos o incorrectos, sino en que la red constituye un medio social por derecho propio, en el que hay posibilidades ilimitadas para los estudios humanísticos.

Esta riqueza se explora en el tercer y último apartado, en el que se ofrece una visión general sobre algunas de las herramientas y aplicaciones digitales con las que se pueden iniciar proyectos de humanidades digitales que contribuyan al mejor conocimiento de comunidades lingüísticas en la red, y que pueden ser empleadas en el desarrollo de estrategias de enseñanza de ELE.

1. LENGUAJE Y TECNOLOGÍA

1.1. Lenguaje y Tecnología. El presente

En los últimos años, nombres como Assange, Manning o Snowden se han convertido en símbolos de una corriente que critica el empleo que se hace de las redes digitales por parte de las instituciones gubernamentales y de las empresas privadas para socavar las libertades civiles de los ciudadanos que emplean dichos medios de comunicación.

En la prensa abundan también noticias que se relacionan de una u otra forma con estos fenómenos. Las reclamaciones por parte de empresas como Facebook para lograr que el hecho de que pulsar el botón *Me gusta* constituya un acto de *libertad de expresión*, o las disputas entre las grandes empresas de internet contra las exigencias de privacidad por parte de los países de la UE constituyen, entre otros, ejemplos de uno de los mayores obstáculos encontrados en la evolución reciente de las redes digitales. Si bien, por una parte, las redes han permitido imitar, de forma creativa, gran parte de las estructuras sociales, culturales y lingüísticas que consideramos esenciales para el establecimiento de una identidad grupal que dé pie al establecimiento de una comunidad, por otra parte, la propia naturaleza de la información digital impide que ciertos derechos constitucionales se vean reproducidos de forma igualmente efectiva. Los *usuarios* medran en un entorno en el que se permiten gran parte de las actividades que caracterizan la interacción humana, pero no son *propietarios* de dicho entorno. Así, en tanto que el espacio de interacción es propiedad de un tercero, las garantías para que la actividad social digital se desarrolle de forma que imite a la perfección su homólogo físico se convierten en un *valor de cambio* para el que existe oferta y demanda y, también de forma automática como ya advirtiera Ivan Illich, «un mercado negro»[1].

Sin embargo, un breve análisis histórico de la evolución de las relaciones del lenguaje y la lengua con la tecnología puede permitirnos apreciar que la presente tormenta mediática en torno a los casos de espionaje indiscriminado, las revelaciones sobre los aspectos más atroces e hipócritas de las relaciones internacionales o de la guerra, o la propia inexistencia de una verdadera privacidad en las comunicaciones digitales, entre otras, no han de sorprender a nadie, pues son, como veremos, características intrínsecas de dicha tecnología.

La existencia de una tecnología cualquiera obedece siempre al concurso de varios factores entre los que destacan, siguiendo la propuesta que hiciera Thomas Kuhn[2], el contexto de descubrimiento y el contexto de justificación. En el caso de las tecnologías de la información y de la comunicación, el contexto de justificación es la «necesidad política de disponer de medios que permitan identificar y mantener un registro de aquellos elementos que son esenciales para una comunidad».

Un ejemplo de esta tecnología es la escritura. Y un ejemplo de escritura de gran interés para el desarrollo de este artículo es el caso del Lineal B[3]. El Lineal B es un tipo de escritura antigua localizada en la isla de Creta y cuya traducción no se logró hasta los años cincuenta del siglo XX. El mérito suele atribuirse al británico Michael Ventris quien, tras identificar grupos de símbolos que se repetían a lo largo de los textos en las diversas tablillas de arcilla conservadas, tuvo la inspiración de relacionarlos con los nombres de los núcleos urbanos más prominentes de la

1 Illich, I. (1973). *Tools for Conviviality.*

2 Kuhn, T. (1962). *The Structure of the Scientific Revolutions*.

3 Documento de internet disponible en <http://en.wikipedia.org/wiki/Linear_B>.

isla, en el entendimiento de que los topónimos suelen variar poco con el paso del tiempo. Su intuición resultó acertada y poco a poco, los textos empezaron a hacerse comprensibles.

Sin embargo, el trabajo de Ventris no hubiese sido posible sin el extraordinario *tour de force* de otra entusiasta del Lineal B, la norteamericana Alice Kober[4], profesora asociada del departamento de lenguas clásicas en el Brooklyn College de Nueva York que, en torno a las décadas de 1930 y 1940, se dedicó en solitario a clasificar todos los símbolos individuales y combinaciones de símbolos presentes en las tablillas conservadas. Kober murió de cáncer antes de poder descifrar el Lineal B, pero su legado, que llega a sumar unas 180.000 fichas, fue la base que dio a Ventris la clave del éxito.

Kober realizó su trabajo de forma completamente manual, almacenando las fichas en las cajas de cartón en las que compraba el tabaco que la acabó matando. Cualquier intento de recabar la información así consignada debía realizarse de forma igualmente manual, sin ningún tipo de apoyo técnico. Sin embargo, en la época en la que Kober iniciara su titánica labor, existían ya herramientas mecánicas destinadas a lograr acometer tareas relacionadas con la información aún más grandes que la realizada por Kober.

1.2. Lenguaje y Tecnología. El pasado

> *Haced un censo de toda la comunidad israelí, de sus clanes y familia, y haced una lista de todos sus nombres, uno a uno [...] y los de más de 20 años servirán en el ejército [...]*
>
> (Números 1,2,3)

A finales del siglo XIX, Estados Unidos se enfrentaba a un problema nunca antes afrontado por ningún país de la vieja Europa, con la excepción sincrónica, si cabe, de Rusia. El tamaño de su población y la enorme extensión de territorio que abarca el país hacían prácticamente imposible que se pudiese confeccionar un nuevo censo de población en un tiempo inferior al del intervalo estipulado para poder ser empleado de forma práctica en la ordenación de la sociedad norteamericana.

Sin embargo, un joven llamado Herman Hollerith[5] se alzó con una solución que logró resolver el problema. Hollerith diseñó un sistema mecánico capaz de procesar y recuperar información utilizando máquinas tabuladoras y tarjetas perforadas. Al lograr automatizar el proceso de recogida de datos provenientes de una ingente cantidad de cuestionarios, el censo pudo completarse a tiempo y la

---

4 Documento de internet disponible en <http://www.bbc.co.uk/news/magazine-22782620>.

5 Documento de internet disponible en <http://en.wikipedia.org/wiki/Herman_Hollerith>.

empresa de Hollerith, International Time Recording Company, pasaría a ofrecer sus soluciones a muchas otras naciones del planeta.

Entre estas naciones se encontraba Alemania que, a mediados de la década de 1930, afrontaba un tipo de rompecabezas demográfico cuya solución también vino de la mano de Hollerith, cuando su empresa ya respondía al nombre por el que es hoy conocida: International Bussines Machines o IBM, literalmente «la empresa de las soluciones».

La presencia de máquinas tabuladoras automáticas y eléctricas en los campos de concentración nazis está abundantemente documentada[6], así como lo está su empleo como base de trabajo para los planes de exterminio llevados a cabo por el III Reich.

En definitiva, la misma empresa que inició las aplicaciones comerciales del cálculo automatizado para problemas demográficos es también la empresa que inició una aplicación a otros problemas de cálculo no menos complejos, que se revelaron apremiantes en el fragor de la II Guerra Mundial. La necesidad de encontrar soluciones a complejos problemas de balística asociados a los primeros sistemas de disparo automático, o las aplicaciones del cálculo automático para descifrar los muchos códigos con los que los diferentes bandos de la contienda cifraban sus comunicaciones, son algunos ejemplos que ilustran la vertiginosa velocidad (política) que se imprimió a las TIC hasta llegar al presente.

Su finalidad a día de hoy, sin embargo, sigue siendo la ilustrada por la cita que abre este apartado y atestiguada en el apartado anterior: el objetivo de las tecnologías de la información y de la comunicación es identificar y mantener un registro de aquellos elementos que son esenciales para una comunidad.

A nadie debe sorprender que el descifrado del Lineal B diese como resultado una relación de ciudades, con sus recursos y su localización. El descifrado resultó ser, en definitiva, una constatación de una suposición plausible más que una verdadera revelación histórica.

---

6 Documento de internet disponible en <http://en.wikipedia.org/wiki/IBM_and_the_Holocaust>.

2. HUMANIDADES DIGITALES. EL DILEMA DE MR. PROPER

Para la construcción de un estado o una nación no solo es necesario disponer de un sustrato cultural común, como una lengua. Es también necesario conocer qué recursos hay, su disponibilidad y su ubicación. En su obra *La Ciudad en la Historia*[7], Lewis Mumford escribe:

> Ejercer el poder en todas sus formas era la esencia de la civilización […]. Las comunidades urbanas, entregadas de lleno a la nueva expansión del poder, perdieron este sentido de los límites: el culto del poder se regodeaba en su misma ostentación sin límites.

Hoy en día, las TIC siguen sirviendo el mismo propósito para el que fueron creadas y lo hacen dentro del mismo marco político en el que fueron promocionadas. En esencia, las TIC permiten:

- enumerar recursos,
- determinar las relaciones entre estos y
- localizar dichos recursos.

En tanto que usuarios, la mayor parte de las personas que emplean los servicios suministrados por las corporaciones que dominan el mercado de las TIC acceden a participar de estas características y, por tanto, su actividad se convierte, asimismo, en recursos.

Pero, a diferencia de los ejemplos históricos propuestos, bien sea el censo en los Estados Unidos, el empleo de las tabuladoras automáticas en la Alemania nazi o el censo que se ordena hacer a Moisés de las tribus israelíes en su peregrinación a la Tierra Prometida, hoy día las TIC presentan una diferencia cualitativa. Las TIC han mimetizado, a través de representaciones eminentemente audiovisuales, pero también táctiles, muchas de las características interactivas de los entornos lingüísticos, sociales, culturales y naturales, hasta el punto de que presentan una versión alternativa verosímil en la que se ha vertido una gran parte de la experiencia vital de los seres humanos. En definitiva, han permitido transferir la esencia del concepto de *comunidad*, y han dado pie a la continuación de aquellas actividades que han caracterizado la propia evolución de los grupos humanos, de sus culturas, de sus relaciones y de su lengua. En pocas palabras: quién hace qué, cómo y con quién[8].

---

7 Mumford, L. (1966). «La cristalización de la ciudad», en *La Ciudad en la Historia* (Vol. 8, tomo 1). Destino..

8 Dunbar, R. (1996). «Grooming, Gossiping and the Evolution of Language».

Por tanto, si definimos *Humanidades* como «aquello que hacen los humanos», queda claro que las redes y las comunidades en ellas creadas representan un espacio de estudio de infinitas posibilidades para las disciplinas humanísticas y, concretamente, para aquella característica que define a la humanidad: su lenguaje.

Para ilustrar el grado de importancia que la Red y sus espacios tienen para las humanidades, me gustaría extender con dos citas la analogía establecida entre el concepto de *ciudad* descrito por Mumford en el capítulo «La cristalización de la ciudad» de su obra *La Ciudad en la Historia* y el concepto de *comunidad* indicado más arriba:

> Así, la más preciosa invención colectiva de la civilización, la ciudad, a la que sólo precede el lenguaje en la transmisión de la cultura, se convirtió desde el principio en el receptáculo de destructoras fuerzas internas, orientadas hacia el constante exterminio.
>
> Tenemos la prueba de este súbito sentido de exaltación en las dimensiones cada vez mayores de las grandes pirámides; del mismo modo que tenemos su representación mitológica en la historia de la ambiciosa Torre de Babel, a la que puso fin la comunicación imposible que una excesiva extensión del territorio lingüístico y de la cultura puede haber producido una y otra vez.

Ambas citas reflejan sendas características de la Red. En primer lugar, la Red es un constructo que incluye una inmensa pluralidad en la que hay verdadero espacio para la libertad y para el pleno disfrute de los derechos civiles, incluido el derecho a reivindicarlos si estos sufren menoscabo. En segundo lugar, esa misma pluralidad es germen de una riqueza social, cultural y lingüística que apenas sí podemos empezar a entender.

De esta manera, podemos afirmar que en la Red existen ciudades que, muy a su pesar, aún coinciden con la clásica definición de «asentamiento mayor cerrado»[9] en las que los *ciudadanos* son *clientes* de aquellos intereses que configuran sus preferencias. Pero, al mismo tiempo, la Red está formada por comunidades libres con una gran fuerza creativa, y es en estas comunidades donde se gestan los usos aún no conocidos de la Red como entorno social, cultural y lingüístico.

Por tanto, para los profesionales de la lengua, como garantes de la transmisión de una cultura, existe una responsabilidad que va más allá de criticar el uso o el abuso que, con respecto a una norma (adecuada o apropiada), se pueda hacer de una lengua. El dilema de Mr. Proper se diluye en su propia inconsistencia. Los aspectos políticos que generaron la existencia de la Red crean una ética digital en la que nuestra actividad queda registrada, es observable y es, por tanto, susceptible de ser estudiada.

---

9 Kolb, F. (1992). «La Ciudad en la Antigüedad». Gredos.

3. LAS NUEVAS OPORTUNIDADES

3.1. Mr. Proper prisionero

Conocer el origen así como la ética intrínseca de la Red no es motivo para rechazar las oportunidades que presenta. El fenómeno digital más bien alienta la responsabilidad que, como transmisores de una lengua y una cultura, los profesionales de las humanidades hemos de mantener en la esfera de las redes. Si bien se ha afirmado anteriormente que la tecnología es política, no es menos cierto que ninguna política tiene la hegemonía sobre el futuro, ni siquiera a corto plazo.

Siendo así, la mejor estrategia ante el dilema de Mr. Proper, que resulta ser una mal velada analogía del dilema del prisionero[10], no es la de caer en la tentación de *delatar* un uso correcto o incorrecto del lenguaje en la Red, sino más bien entender que la cooperación entre ambas es la mejor solución, aunque tal estrategia pase por un obligado periodo de silencio. En este ínterin, podemos aprovechar para observar, registrar, analizar y, quién sabe, contribuir en el futuro a alumbrar la identidad de nuevas comunidades aún desconocidas.

Los clanes, las aldeas y las familias siguen siendo las unidades sociales sobre las que aplicar las interesantes opciones que ofrecen, entre otras, las disciplinas de la lingüística computacional, el *machine learning* o la inteligencia colectiva. Citando de nuevo a Mumford:

> Es irónico —pero también es un consuelo— que las ciudades hayan sobrevivido reiteradamente a los imperios militares que, en apariencia, las destruyeron para siempre. Damasco, Bagdad, Jerusalén y Atenas siguen en los mismos solares que inicialmente ocupaban, vivas, aunque poco más que fragmentos de sus antiguos cimientos queden a la vista.

O, dicho de otra forma, con palabras de «El nombre de la Rosa», de Umberto Eco: «De la rosa solo nos queda el nombre».

Y es gracias a los nombres que se han podido reconstruir lenguas que una vez sirvieron para organizar personas y comunidades.

Sin embargo, será siempre necesario el duro trabajo de recogida, clasificación y análisis que, en el futuro, permita alumbrar nuevas identidades y nuevas comunidades, preferentemente antes de que el tiempo las extinga. Esta es, hoy por hoy, la gran tarea que aguardan los profesionales de las Humanidades en la Era Digital.

---

10 Documento de internet disponible en <https://en.wikipedia.org/wiki/Prisoner's_dilemma>.

3.2. Más allá de Google

3.2.1. Python no muerde

Superar el grado de *usuario* en la esfera digital es relativamente fácil para cualquier persona que disponga de interés y de unos mínimos conocimientos generales de los entornos digitales. La web, en su mayoría, está programada en código abierto, es decir, la estructura de los programas que configuran la arquitectura de sus servicios son hasta cierto punto accesibles y modificables, y ello permite, entre otras opciones, acceder a una ingente cantidad de datos digitales susceptibles de procesarse de múltiples formas.

Sin embargo, para el lector no especializado, las preguntas surgen a continuación de forma inmediata: ¿qué tipo de datos?, ¿cómo acceder a ellos?, ¿cómo manipularlos? y, sobre todo, ¿con qué propósito?

En mi experiencia como consultor para la creación de programas de formación del profesorado en Valencia, he afrontado estas mismas preguntas en diversas ocasiones, resultando que, por lo general, la mayoría de las personas que deben usar TIC por motivos profesionales rara vez entienden, no ya cómo se ejecutan los programas, sino qué pueden hacer con ellos.

Mi misión ha sido precisamente presentar estas opciones con el objetivo de alumbrar nuevas alternativas en el cotidiano devenir de la actividad educativa y lograr que, entre otras ventajas, se produzca un vínculo más fuerte y proactivo del usuario con los sistemas que emplea, contribuyendo a aumentar el sentido de propiedad sobre el trabajo realizado. Este es un cambio cualitativo que ha producido numerosos beneficios en las comunidades educativas en las que he intervenido, y que ha favorecido la mejor integración de las herramientas TIC que, como todo el mundo sabe, son imprescindibles para el éxito de la educación. La totalidad del ámbito de la comunidad educativa es también, aunque nadie lo sepa, imprescindible para el éxito de la educación.

El ejemplo de esta experiencia nos da la pista para empezar a responder a las preguntas apuntadas con anterioridad. Cualquier comunidad educativa dispone, hoy por hoy, de un entorno digital que refleja, con mayor o menor grado de precisión, la propia estructura y funciones de la comunidad a la que sirve: página web de información general, *links* de esta web a información de diverso tipo, incluyendo calendarios, eventos, organización, etc., y, si seguimos buscando, descubrimos aún más niveles de información, muchos de ellos de acceso restringido, que incluyen información personal o registros de la actividad de alumnos y profesores. Y toda esta información se actualiza, es decir, se modifica e incrementa a diario.

Por consiguiente, respondiendo a la primera pregunta, podemos decir que el tipo de datos disponibles abarca todo aquello que la propia política del Centro educativo acepte como susceptible de *colgarse* en su web. En esencia, datos de texto, sonido, imagen y vídeo. Si bien todos ellos son de interés, son los datos de texto (y posiblemente sonido) los más convenientes para las aplicaciones que aquí se presentan.

En cuanto a la segunda pregunta, *¿Cómo acceder a ellos?*, existen diversas técnicas, pero desde la perspectiva de este artículo la única vía válida de acceso a los datos es solicitar permiso para ello. Sin embargo, ninguna solicitud llegará a buen puerto si no existe un motivo debidamente justificado para granjearlo. Por tanto, hemos de dejar momentáneamente la tercera pregunta para saltar directamente a la cuarta.

Es posible incluso que haya que dar más de un salto, o mejor dicho, que haya que dar dos tipos diferentes de salto, uno de orden (de la segunda pregunta a la cuarta) y otro, doble, de Fe:

- Fe-A: los datos se pueden manipular de múltiples formas;
- Fe-B: la forma de hacerlo depende totalmente de la respuesta a la cuarta pregunta.

Para alentar la respuesta a esta última pregunta, en primer lugar es absolutamente necesario convertir el valor intangible Fe-B en el valor ligeramente más tangible Fe-A antes de lograr su transformación en un valor sólido y tangible como el hierro (Fe). Y para ello será imprescindible hablar de serpientes y computación o, más exactamente, hablar con serpientes que saben de computación. Pero no teman, no se trata de la Serpiente del Paraíso y, ya que estamos, ni siquiera muerde. Es más, ni siquiera es una serpiente, pero da la casualidad de que su nombre es Python, circunstancia que, en ocasiones, puede dar lugar a cierta confusión.

3.2.2. Un lenguaje para aclararnos todos y librarnos de las tinieblas

La interesante historia del lenguaje de programación Python puede leerse en su correspondiente entrada de Wikipedia[11]. Para nuestros propósitos es suficiente indicar que su código es abierto, que cuenta con una increíble popularidad y apoyo entre la comunidad de programadores vinculados a internet, que es relativamente rápido de aprender y que cuenta con una gran cantidad de aplicaciones y librerías disponibles para su uso directo. Quizás la característica más curiosa sea que, con unas nociones básicas de este lenguaje, se puede leer un programa escrito en Python y llegar a hacerse una idea de lo que el programa hace.

---

[11] Documento de internet disponible en <http://en.wikipedia.org/wiki/Python_(programming_language)>.

Además de estas ventajas, hay dos razones fundamentales por las que merece la pena conocer Python.

Una de ellas es que existen una gran cantidad de contenido en la Red que está programado en Python y es accesible, previa solicitud (y con ello cumplimos el criterio estipulado en la pregunta número dos), a través de las API (Interfaz de Programación de Aplicaciones, en su traducción al español). Existe una web llamada <www.programmableweb.com> que ofrece directorios actualizados de sitios web que cuentan con estos accesos. Sin embargo, bastará con indicar que Facebook, Twitter, Google o diversas webs de blogs disponen de ellos para que estos sitios web perciban la relevancia de Python como herramienta de trabajo.

La segunda razón fundamental es que Python es el lenguaje de programación en el que se ha creado la aplicación NLTK (Natural Language Toolkit, en español Kit de herramientas de lenguaje natural). Esta aplicación, accesible de forma gratuita a través de la web del mismo nombre[12], cuenta con una extensa documentación que permite aprender, paso a paso, no solo los aspectos básicos de la programación en lenguaje Python, sino las diversas aplicaciones del área de la lingüística computacional que pueden ejecutarse sobre cualquier conjunto de datos.

Por tanto, una lectura, incluso somera, de los contenidos y posibilidades de NLTK puede bastar para hallar una respuesta a la cuarta pregunta con que se cerraba el apartado anterior. Las posibilidades abarcan, entre otras, desde la mera cuantificación de palabras hasta el análisis sintáctico o la clasificación automática de textos.

De entre las múltiples opciones que NLTK ofrece habrá alguna que nos inspire. De esta forma, conseguimos dar el salto cualitativo desde el valor intangible Fe-B al valor ligeramente más tangible Fe-A.

4. CONCLUSIÓN

A lo largo de este artículo hemos revisado la importancia que las redes digitales tienen para el desarrollo de una lengua. Se ha hecho especial hincapié en la necesidad de estudiar estos fenómenos lingüísticos desde el ámbito de las humanidades y, más particularmente, desde el campo de la enseñanza ELE. Se ha mostrado también que, por la propia naturaleza de la información digital, existen en nuestros entornos más cercanos cantidades ingentes de datos cuyo análisis puede aportar un enorme valor añadido a la enseñanza de una lengua, pero se ha mostrado asimismo la absoluta necesidad de un propósito en la investigación. Finalmente, se han presentado algunas de las alternativas más asequibles para introducirse en el campo de la computación lingüística de la mano de Python y NLTK, herramientas que pueden proporcionar no solo la capacidad analítica,

12 Documento de internet disponible en <http://nltk.org/>.

sino también principios generales sobre los que diseñar programas de investigación de carácter local.

En conclusión, retomaré la analogía establecida entre la *ciudad* de Mumford y el concepto de *comunidad*. En su discurso, Mumford identifica la *aldea* como el germen generador de la fuerza creativa que nutre la ciudad, pero es la misma cualidad autodestructiva de la *ciudad* la que pone en peligro la existencia misma de la *aldea*:

> Pero no bien permitamos que la aldea desaparezca, este antiguo factor de seguridad se desvanecerá. La humanidad todavía tiene que reconocer este peligro y eludirlo.

Y en las manos de los profesionales de las Humanidades están las herramientas capaces de conjurar este peligro.