

Explorando cambios sociales silenciosos: una propuesta a partir de la explotación digital de una gran mina de datos históricos (Cataluña, siglo XVIII).

La razón de ser de este proyecto es la exploración de las posibilidades que ofrecen las nuevas tecnologías para mejorar el análisis histórico de las transformaciones sociales. El proyecto se enmarca en el Programa de Investigación UNearthing SILENt social ChangeE (UNSILENCE), que están desarrollando los miembros del equipo de la UdG, y resume la voluntad de realzar procesos y dar voz a individuos que muchas veces han sido silenciados por la Historia. Partimos de dos constataciones, de procedencia muy distinta, aunque complementarias. La primera, fruto de cierta insatisfacción respecto del estado actual de la investigación histórica, es la necesidad de afrontar el estudio de algunos procesos de cambio social que, por haber merecido hasta ahora poca atención de los historiadores, llamamos silenciosos. Entre ellos queremos destacar, por su relevancia en la historia social, aquellos procesos que, pudiendo tener como protagonistas tanto a grupos sociales humildes como a grupos sociales pudientes, conllevan la transformación de estos mismos grupos sociales. El carácter innovador de nuestro proyecto radica en su capacidad de detectar, capturar e interpretar estos cambios. La factibilidad del proyecto radica en la segunda constatación señalada, fruto de los últimos progresos alcanzados en las técnicas de reconocimiento automático de texto manuscrito y también en las técnicas de explotación y minería de datos. Estos desarrollos nos ofrecen hoy la posibilidad de explorar de un modo nuevo, y menos costoso que los métodos convencionales, algunas fuentes documentales capaces de reconstruir, por la riqueza de su contenido, el dinamismo social en sus múltiples dimensiones. Los resultados obtenidos gracias al sometimiento a la doble renovación metodológica que aquí se propone –nuevas preguntas, nuevas técnicas– de una rica fuente registral, los libros del Registro de Hipotecas de Cataluña, permitirán incidir en algunos debates historiográficos que cuentan con una larga tradición en la historia social y económica, como el papel de las élites en relación al cambio histórico, las raíces del crecimiento económico, u otros más recientes, como los generados en torno a la revolución industrial o a las desigualdades sociales. Pero el mayor potencial del proyecto se revelará en su capacidad de abordar nuevos problemas. Entre ellos, algunos relativos a temas muy actuales, y por lo tanto de interés para el conjunto de los científicos sociales, que hasta ahora apenas han sido abordados por los historiadores, como, por ejemplo, la emergencia o el declive de grupos medios, o la salida de la pobreza de capas importantes de la población.

1. Estado de la cuestión y objetivo principal del proyecto

El objetivo científico del proyecto es la exploración y aplicación de métodos innovadores para la reconstrucción de dinámicas sociales históricas. En concreto, proponemos la utilización de nuevas técnicas informáticas, que culminen, en su estadio final, en la aplicación de las metodologías analíticas de “minería de datos” para capturar procesos de cambio social difíciles de detectar a partir de los métodos convencionales. Defendemos que el tratamiento masivo e integral de los datos (de carácter muy diverso, que han tendido a ser contemplados y analizados parcialmente y aisladamente) acumulados en algunas fuentes documentales históricas especialmente ricas, no sólo puede incidir en la forma de afrontar viejos temas, sino que permite el planteamiento de nuevos problemas y nuevas hipótesis de trabajo, algunos de gran actualidad,

que hasta ahora apenas han sido objeto de atención por parte de los historiadores. Teniendo en cuenta que el objeto de estudio son las dinámicas sociales, la obtención de resultados nuevos –y tal vez inesperados– en este campo, gracias a la renovación metodológica que se propone, fácilmente trascenderá el campo de la historia social y económica y repercutirá en otros campos del conocimiento histórico y de las ciencias sociales interesadas en la comprensión del cambio social.

El proyecto permitirá afrontar y revisar de una manera nueva algunos tópicos bastante anclados en la historiografía y en las ciencias sociales y al mismo tiempo permitirá plantear algunos problemas nuevos, que hasta ahora han merecido poca atención por parte de esa misma historiografía. Entre los primeros, queremos destacar, porque lo comparten diferentes escuelas historiográficas, la atribución de un rol determinante a las élites y a las instituciones como motores del cambio social. Entre los segundos, uno que presenta una indudable actualidad: el estudio del dinamismo social que supuso la emergencia de los grupos sociales que podemos considerar intermedios, lo que significa la superación de la pobreza por parte de grupos humildes, o el empobrecimiento de sectores antes acomodados. Proponemos analizar de un modo nuevo, posible gracias a los avances adquiridos en las tecnologías digitales, todas estas problemáticas –las viejas y las nuevas- a partir de una fuente especialmente rica –los libros de las Contadurías de Hipotecas en Cataluña– cuyas coordenadas temporales nos permiten incidir, además, en otra temática de especial interés para todos los científicos sociales: la crisis del antiguo régimen y los orígenes de la sociedad contemporánea.

Renovación de un viejo tema historiográfico: el papel de las élites y de las instituciones como motor del desarrollo histórico y del crecimiento económico.

- El papel decisivo de las élites constituye un lugar común en las principales corrientes historiográficas. Pero ¿hasta qué punto esta idea no ha sido condicionada por la necesidad de trabajar a partir de unos pocos datos documentales, entre los que sin duda los relativos a las élites se hallan sobrerrepresentados? Un tratamiento integral de las noticias reunidas en los libros de las Contadurías de Hipotecas, relativas al conjunto al conjunto de individuos que actúan en una sociedad- permitirá observar el conjunto de dinámicas sociales y resituar el papel de las élites en el lugar real que ocupan en este conjunto.

- El papel de las instituciones también constituye otro de los lugares comunes de las principales corrientes historiográficas, especialmente en el campo de la historia económica. En especial, el cambio institucional se ha considerado básico y relevante para explicar los orígenes de las sociedades occidentales contemporáneas. Pero ¿hasta qué punto esta visión no ha sido condicionada por la ignorancia sobre el funcionamiento real y cotidiano de muchas de las instituciones? El tratamiento integral que proponemos de los datos contenidos en la documentación generada por las Contadurías de Hipotecas permite observar, en unos años clave de cambio institucional, ya que abarca el período final de antiguo régimen y las primeras décadas de gobiernos liberales en España, el funcionamiento de algunas instituciones –entre ellas, el mismo Registro objeto de estudio- en el día a día.

Observación de nuevas problemáticas y planteamiento de nuevas hipótesis de trabajo: los procesos de empobrecimiento y enriquecimiento de determinados colectivos y los procesos de emergencia de nuevas clases medias.

- Hasta hace algunas décadas, los debates sobre el nivel de vida de las clases populares europeas en la etapa final del antiguo régimen se habían centrado sobre todo en los supuestos procesos de expropiación campesina que habrían conducido a la proletarización. El reciente debate historiográfico en torno a la idea de una "revolución industrial", en cambio, sugiere la necesidad de estudiar el conjunto de oportunidades que, en determinados contextos, podían favorecer un "enriquecimiento" o, simplemente, un crecimiento de los bienes consumidos en los hogares de diversos grupos sociales. No es fácil estudiar el nivel y las condiciones de vida de los grupos más humildes de las sociedades, porque a menudo dejan poco rastro en las fuentes documentales. Nos proponemos obtener esta información a partir del tratamiento masivo de la documentación contenida en los libros de las Contadurías de Hipotecas.

- En una línea parecida, en los últimos años, algunos debates recientes generados en torno a la emergencia de nuevas clases medias en algunos países hasta hace poco considerados del Tercer Mundo, o el declive de las clases medias en los países industrializados, a consecuencia de la crisis actual, invitan a plantear el tema de los procesos de enriquecimiento o empobrecimiento de algunos grupos sociales que hasta hace poco la historiografía apenas había contemplado. Nuestra fuente y la metodología concreta propuesta ofrecen la posibilidad de analizar para un largo período de tiempo numerosas trayectorias individuales y colectivas, relativas a los distintos grupos sociales. Este último factor confiere especial relevancia al proyecto, ya que las únicas fuentes que hasta ahora han permitido llevar a cabo ejercicios similares, los archivos privados patrimoniales, suelen hallarse restringidos a los grupos sociales bien situados en la escala social.

2. Una fuente documental especialmente idónea para el tratamiento propuesto: los libros de las Contadurías de Hipotecas de Cataluña.

Las Contadurías de Hipotecas fueron creadas mediante una Real Pragmática publicada en 31 de enero de 1768 y estuvieron en funcionamiento hasta que la ley Hipotecaria de 8 de febrero 1862 las sustituyó por el moderno Registro de la Propiedad. Sus fondos abarcan casi un siglo. Hay tres razones que hacen esta fuente, que el equipo investigador conoce bastante bien, como se verá, particularmente atractiva para el trabajo que proponemos:

1) abraza documentación notarial, para una zona de gran tradición con las escrituras notariales, de tipo muy diverso: escrituras de compraventa de tierras, contratos de arrendamiento y de acceso a la tierra vía enfiteusis, contratos asociados a operaciones de crédito, como censales y obligaciones, contratos asociados a la reproducción social y familiar, como capítulos matrimoniales, contratos de sociedades de comercio e industria, etc.; de hecho, la singularidad que hace especialmente rica esta fuente, en comparación con otras similares de otros países y de la misma fuente en otros territorios del estado español, deriva en buena parte del interés de los colegios de notarios de Barcelona y Girona, desde su creación, en 1768, por asegurar su buen funcionamiento. Esta acción contrasta enormemente, por ejemplo, con lo ocurrido en otras zonas de España o en países vecinos, como en Francia, donde el interés por el control de la información por parte de los notarios de París conllevó que algunas iniciativas parecidas fueran abortadas.

2) reúne, de modo centralizado, toda la información relativa a un territorio relativamente extenso, como son los distritos hipotecarios. El hecho de trabajar sobre áreas regionales permite trascender el nivel local y formular hipótesis de trabajo susceptibles de ser validadas en función de las diversas realidades socio-económicas que pueden darse dentro del territorio.

3) se trata de una fuente que cubre un período que abarca cambios institucionales importantes, ya que comprende el paso de una sociedad de antiguo régimen a una sociedad liberal. Por esta razón, permite estudiar con cierto detalle la incidencia real de determinadas medidas legales en los cambios socioeconómicos. Una tesis bastante extendida en la historiografía española es la de que la mayoría de los cambios fueron fruto de las medidas liberales. La observación de importantes cambios en las formas de contratación agraria, en las formas de constitución de sociedades, en las etiquetas sociales o en el valor de las dotes, para poner algunos ejemplos, durante el antiguo régimen, subraya la relevancia de las transformaciones que precedieron a las reformas liberales y que se produjeron de forma bastante independiente a ellas.

3. Composición e idoneidad del equipo de trabajo.

El proyecto que se presenta requiere la colaboración entre historiadores sociales e ingenieros informáticos especializados en la transcripción automática de manuscritos y en el manejo de grandes cantidades de datos. El equipo de trabajo se halla constituido por 4 historiadores, el director del Archivo Histórico de Girona y 3 ingenieros informáticos. Los 4 historiadores son miembros del Centre de Recerca d'Història Rural de la UdG, que cuenta con una importante trayectoria en el ámbito de la historia social y económica. La investigadora principal del proyecto, Rosa Congost, es la actual coordinadora del Centre, catedrática de historia económica de la UdG, presidenta de la European Rural History Society (EURHO) e investigadora ICREA Academia. Rosa Ros y Enric Sagner son profesores titulares de historia económica de la UdG y Ricard García Orallo es investigador a tiempo completo y profesor asociado de la UdG. A este núcleo de historiadores debe añadirse Joan Ferrer, director del Archivo Histórico de Girona.

Los aspectos técnicos del proyecto se desarrollarán bajo la supervisión y responsabilidad del equipo de ingenieros especializados en la transcripción automática de manuscritos y en minería de datos. Tres de ellos son miembros del PRHLT (Pattern Recognition and Human Language Technology) de la Universidad Politécnica de Valencia. Por su parte, Jordi Regincós, experto en el tratamiento de bases de datos, es profesor titular de Lenguajes y sistemas informáticos en la Escuela Politécnica de la UdG.

4. Metodología de la investigación: una propuesta innovadora formulada desde la experiencia del equipo investigador

Lo que confiere originalidad a este proyecto es la propuesta de una nueva forma de explotar la fuente que se ha presentado. A pesar de su riqueza documental, lo cierto es que ha sido poco utilizada para la investigación histórica. Ello es debido precisamente al difícil manejo del conjunto de documentación contenida en los libros, que hasta ahora, sólo ha podido ser explotada de una forma parcial. Esta forzosa parcialidad, que requería una selección previa del material a examinar, en cierta manera condicionaba los resultados obtenidos. De este modo no se aprovechaba la que consideramos la principal ventaja de la fuente, que constituye el objetivo principal de este proyecto y que podemos resumir así: estudiar el conjunto de las transformaciones sociales globales sin apriorismos. Hemos querido ayudarnos de un gráfico

(imagen 1) para explicar, a la vez, la experiencia adquirida previamente, por parte de los miembros del equipo investigador procedentes de la UdG, en el tratamiento de la fuente, y la radicalidad de la nueva propuesta, posible gracias a la colaboración, desde hace un año, con el equipo de investigadores de la Universidad Politécnica de Valencia, que cuenta con una larga experiencia en el reconocimiento óptico de textos manuscritos. Como se ha indicado, dos de sus miembros se han integrado formalmente en este proyecto.

Experiencias previas de los miembros de la UdG del equipo investigador en la exploración de la fuente.

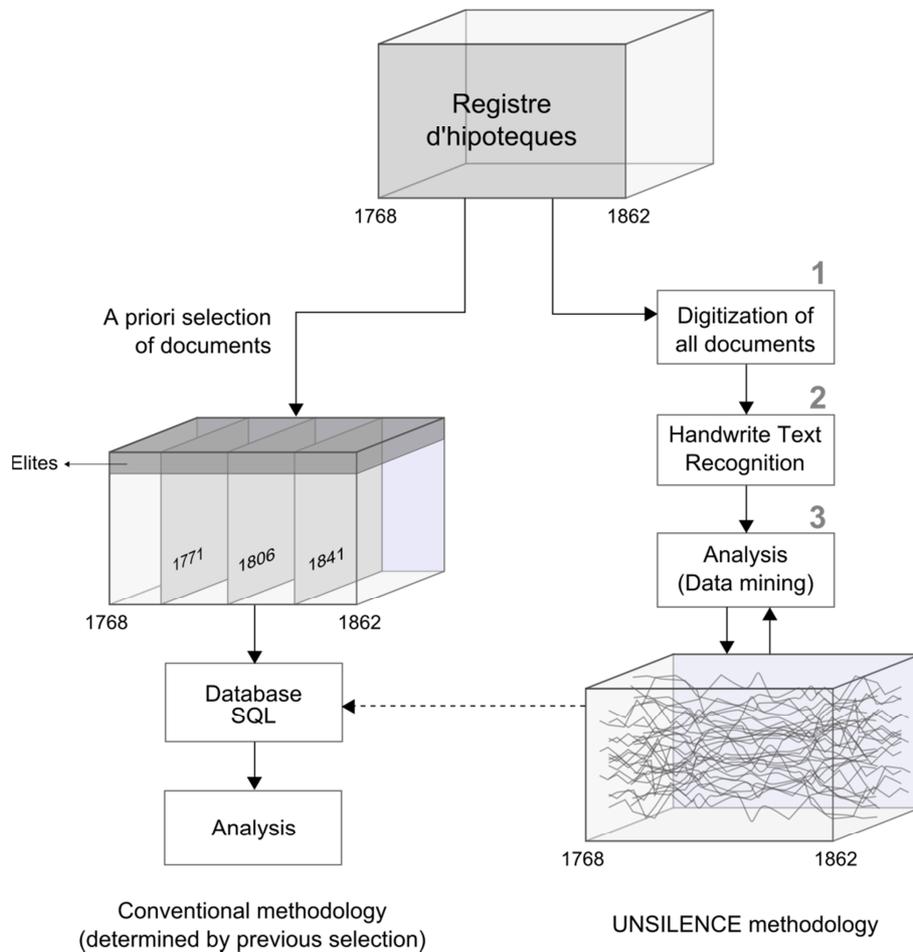
Hasta ahora, habíamos concebido dos formas de trabajar los datos, que implicaban siempre una selección previa de materiales, y que se hallan representadas en la parte izquierda de la imagen 1. Ambas formas ya han sido experimentadas por los miembros del equipo investigador de la UdG en los libros de la Contaduría de Hipotecas de Girona. La primera, que la investigadora principal llevó a cabo en 1990, en su tesis doctoral, consistió en seleccionar, para su estudio, aquellos contratos en los que las élites eran protagonistas. Por ejemplo: escrituras de compraventas de fincas de un valor superior a las 1.000 libras o escrituras de contratos matrimoniales con dotes superiores a esta misma cantidad. Pero esta manera de actuar, si bien nos puede proporcionar referencias interesantes, difícilmente podrá hacernos replantear nuestra opinión sobre la importancia de las élites en la historia. La segunda, que la misma investigadora principal propuso en 2006, en el marco de una experiencia que fue premiada como innovadora por el Ministerio de Economía y Competitividad, y que llamaremos EXPLORA, consistía en vaciar exhaustivamente todas las escrituras correspondientes a tres años concretos: 1771, 1806 y 1841.

Los resultados fueron muy espectaculares, ya que pusieron de manifiesto un protagonismo inesperado de los llamados “trabajadores” en la documentación, es decir, de los grupos humildes, tanto en las compraventas como en las dotes. El proyecto EXPLORA no sólo permitió ver las posibilidades que ofrecía el tratamiento masivo de los datos, sino que ponía de manifiesto el sesgo que podía significar trabajar el dinamismo social sólo con los datos procedentes de las élites, es decir, de los grupos dominantes. Ahora bien, al tratarse de resultados relativos a períodos cortos, de un año, el problema era el de interpretar correctamente los cambios sociales detectados entre ellos. ¿Habían surgido de golpe, o se habían producido lentamente? ¿Habían sido producto de algunos acontecimientos históricos? Por ejemplo: los cambios que observábamos en 1841 respecto de 1806, ¿había que atribuirlos a la revolución liberal? Para responder de un modo adecuado a estas preguntas, tendríamos que analizar más años. Pero la elección de los años podía contaminar una vez más los resultados. La única solución sería realizar el vaciado exhaustivo, año tras año, de todos los datos. Pero esto estaba fuera de nuestro alcance.

Para hacernos una idea, durante el proyecto EXPLORA, fue necesario ocupar a seis estudiantes durante dos años trabajando 15 horas semanales. El vaciado de los 95 años que cubren los libros de las Contadurías de Hipotecas, llevaría a unos 60 años con el mismo trabajo o, si pensamos que podemos realizar este trabajo en los cinco años que dura el proyecto, tendríamos que contratar cada año 76 estudiantes para llevar a cabo el vaciado exhaustivo de las fuentes. Tan solo éste trabajo representaría un coste de 5400 euros x 76 estudiantes x 5 años: 2.052.000 euros. Y sólo estaríamos hablando del distrito hipotecario de Girona. Al finalizar el trabajo, dispondríamos de una enorme base de datos de tipo relacional y estaríamos en condiciones de

realizar trabajos muy interesantes sobre, por ejemplo, el mercado de la tierra, el mercado del dinero o los contratos matrimoniales. Pero continuaríamos limitados, como sucede con frecuencia, por la rigidez de una estructura que, entre otros inconvenientes, dificulta el tratamiento de información compleja y desincentiva, por los enormes costes que supone, la redefinición de los campos para adaptarlos a los posibles nuevos interrogantes sugeridos por los resultados obtenidos.

Imagen 1



Lo que queremos llevar a cabo, representado en la parte derecha de la imagen 1, es radicalmente distinto de lo realizado hasta ahora. En los próximos dos años, con la ayuda de este proyecto, nos proponemos acceder a todos los datos contenidos en los libros del Registro de Hipotecas de Girona, para el período 1768-1800. Y, además, pretendemos tenerlos disponibles para trabajarlos de un modo nuevo, más flexible y con capacidad para formular nuevas preguntas y nuevas hipótesis de trabajo. El paso que nos proponemos dar, pues, gracias a las nuevas técnicas de digitalización, reconocimiento automático de manuscritos, y minería de datos, constituye un paso de gigante desde el punto analítico, ya que permite resolver los dos principales problemas que impedían aprovechar la principal ventaja de nuestra fuente: 1) formular preguntas sobre el cambio social que no condicionen de antemano sus respuestas; y 2) en el caso, previsible, que las respuestas obtenidas exijan nuevas preguntas, estar en disposición de responderlas sin necesidad de volver al punto inicial.

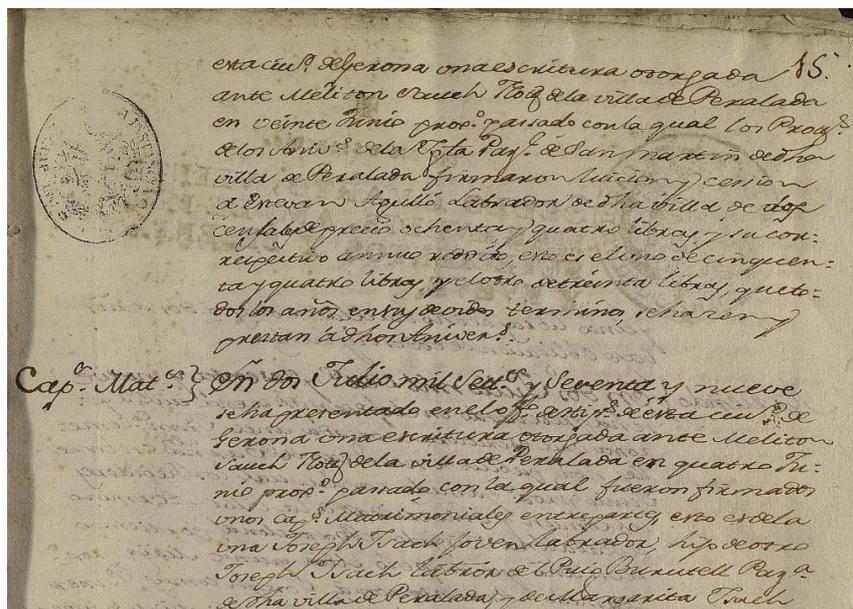
Experiencias previas de los miembros del equipo investigador de la UPV en el reconocimiento de textos manuscritos y, en el último año, en los libros del Registro de Hipotecas de Girona.

En 2006, cuando la investigadora principal solicitó, y obtuvo, el proyecto EXPLORA, no era posible pensar que la fuente podía someterse a un reconocimiento automático de texto manuscrito. Si ahora podemos plantearlo, es porque en los últimos tiempos se ha avanzado significativamente en este terreno y porque, para la preparación de este proyecto, se ha contado con la colaboración de los investigadores del proyecto READ (<https://read.transkribus.eu/>) del equipo PRHLT (Pattern Recognition and Human Language Technology) de la Universidad Politécnica de Valencia, que se han manifestado dispuestos a integrarse en el equipo de trabajo. Consciente que ha habido muchos intentos fallidos en operaciones parecidas, quisiéramos destacar, en primer lugar, la principal ventaja que ofrece nuestra fuente frente a otras: se trata de una fuente escrita durante períodos largos por la misma persona. Para llevar a cabo estas pruebas ha sido muy útil la digitalización ya realizada de catorce volúmenes de la Contaduría de Hipotecas de Girona por parte del Archivo Histórico de Girona, que ha permitido unos primeros ejercicios.

La tecnología más común de transcripción automática de texto, desarrollada en las últimas décadas, se basa en el reconocimiento óptico de caracteres (OCR). Esta tecnología ofrece buenos resultados para los textos impresos, pero no para los textos manuscritos a causa de la dificultad de segmentar de manera fiable el texto en caracteres. La tecnología requerida debe ser capaz de reconocer los elementos de la escritura manuscrita como un todo, sin ninguna segmentación (HTR, *Handwritten Text Recognition*). HTR es un campo relativamente nuevo dentro del ámbito de la visión computerizada. El método que se ha desarrollado a través del proyecto READ (*Recognition and Enrichment of Archival Documents*) se basa en la conversión de la imagen en grafos de palabras y en un proceso de transcripción interactiva mediante la aplicación CATTI (*Computer-Assisted transcription of Text Images*) en que usuario y sistema trabajan de forma asistida hasta que se obtienen transcripciones suficientemente precisas. El sistema utiliza modelos ocultos de Markov (MOM) y N-gramas.

Las pruebas realizadas hasta ahora son altamente esperanzadoras. Para ilustrar la potencialidad de esta tecnología, reproducimos un fragmento seleccionado al azar del primer lote de imágenes (batch 1, 47 imágenes procedentes de las primeras páginas del libro de 1769) que se sometió a transcripción tanto manual como automática. Esta prueba no incorpora el aprendizaje adquirido por el sistema tras la corrección de los lotes posteriores. La imagen corresponde al final de una escritura de redención de un censal y el inicio de una escritura de capítulos matrimoniales. Se compara la transcripción manual de dicho fragmento y su equivalente automático, indicando en rojo los errores cometidos en la HTR.

Imagen 2

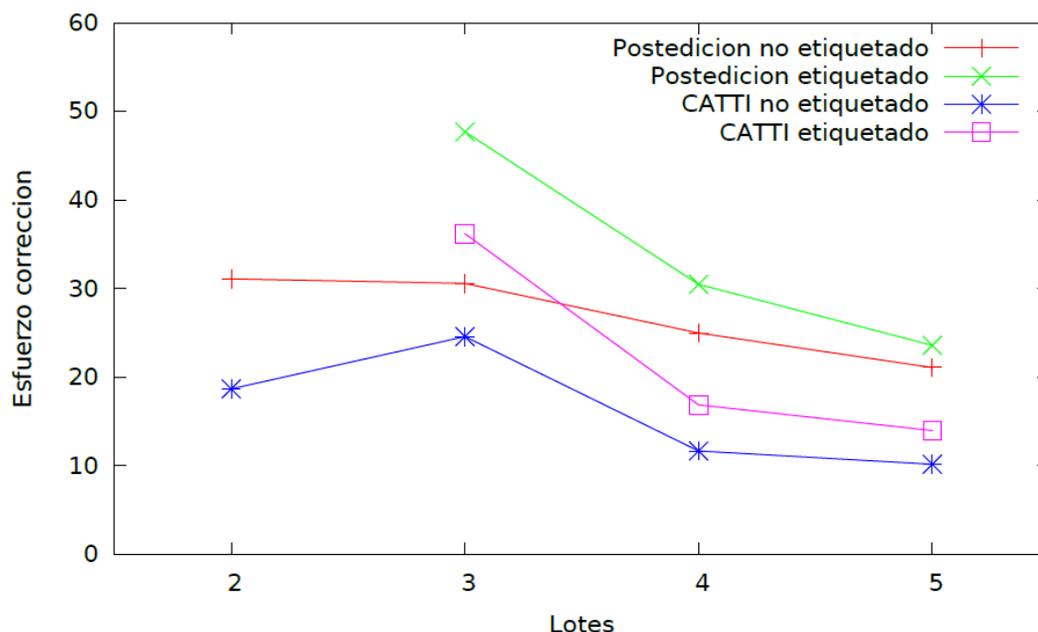


Transcripció manual	Transcripció automàtica
15	y
esta ciud. de Gerona una escritura otorgada ante Meliton Sauch Nott de la villa de Peralada en veinte Junio proxo. passado con la qual los Protts. de los Anivs. de la Igla Parql. de San Martin de dha villa de Peralada firmaron luicion y cession a Estevan Agulló Labrador de dha villa de dos censales de precio ochenta y quatro libras y su cor: respectivo annuo reddito, esto es el uno de cinquen= ta y quatro libras, y el otro de treinta libras, que to= dos los años en sus devidos terminos, se hazen y prestan á dhos Anivers.	esta ciu. de Gerona una escritura otorgada ante Meliton Sauch Nott de la villa de Peralada en veinte Junio proxo. passado con la qual los Bou dhos Anivs. de la Igla Parql. de San Martin de dho villa de Peralada firmaron luicion y Cessión de estevan Miguel Labrador de dha villa de trigo censal de precio ochenta y quatro libras y su cor= respectivo annuo reddito, esto es el auto de cinquen: ta y quatro libras y el otro de treinta libras suel: do los años mismos unidos termino se hazen a presentan de dhos Anivs.
Caps. Mats. En dos Julio mil Setts. y Sesenta y nueve se ha presentado en el Offo. de Hips. de esta ciud. de Gerona una escritura otorgada ante Meliton Sauch Nott de la villa de Peralada en quatro Ju= nio proxo. passado con la qual fueron firmados unos caps. Matrimoniales entre partes, esto es de la una Joseph Isach joven labrador, hijo de otro Joseph Isach labor del Puig Burutell Paqa. de dha villa de Peralada, y de Margarita Isach	Cenl. Mats. En dos Julio mil Setts. y Sesenta y nueve se ha presentado en el Offo. de Hips. de esta ciu. de Gerona una escritura otorgada ante Meliton qual Nott de la villa de Peralada en quatro Ju: nio proxo. passado con la qual fueron firmados unos caps. Matrimoniales entre partes esto es de la una Joseph Battle joven labrador hijo de otro Joseph Barris labor del Puig Palafrugell Paqa. de dha villa de Peralada, y de Margarita Balle

Tras unos meses de explotación del sistema CATTI, los resultados han mejorado sustancialmente y cada vez es menor el esfuerzo que debe realizarse para corregir la transcripción sugerida por el sistema. Tomando como punto de partida el primer lote (batch 1) transcrito tanto manualmente como automáticamente, se han procesado cuatro lotes más, de 50 páginas cada uno, en los que se ha empleado CATTI para transcribir, computando las mejoras de esfuerzo con respecto a haber hecho postedición. En total, se tienen 5 lotes; en los tres primeros lotes se debió hacer un etiquetado (tagging) manual de los términos, mientras que los dos últimos lotes ya incluyen el

etiquetado en el proceso CATTI. Para el procesamiento de cada lote se han empleado los datos transcritos de los lotes previos para generar los modelos correspondientes. Como resultado, el error de postedición va disminuyendo en cada lote, tanto teniendo en cuenta el etiquetado como sin considerarlo. Sin embargo, incluso con esta disminución, la transcripción usando CATTI sigue representando una reducción de esfuerzo considerable; en los últimos lotes, sin tener en cuenta etiquetado, la reducción de esfuerzo es superior al 50%, y teniendo en cuenta el etiquetado es superior al 40%. Las gráficas anexas muestran la evolución de dichos factores.

Imagen 3



En resumen, los test iniciales realizados en la transcripción automática de textos manuscritos aseguran el hecho de que, al finalizar el proyecto, se pueda acceder, para el distrito hipotecario de Girona, a una cantidad ingente de datos contenidos de un modo nuevo, más flexible y con capacidad para formular nuevas preguntas y nuevas hipótesis de trabajo. Habremos avanzado mucho, pues, en la constitución de un laboratorio idóneo para “provocar experimentos” capaces de poner a prueba no sólo, como pedía Marc Bloch, ideas que hoy se hallan muy arraigadas en la historiografía, sino también las ideas nuevas que puedan ir surgiendo en el futuro. En la voluntad de participar en la elaboración de estas nuevas ideas del futuro reside, de hecho, la mayor ambición del proyecto.

5. Bibliografía comentada

Sobre los debates citados en torno a la “revolución industrial”, por una parte, y el papel de las instituciones, por la otra, podemos citar entre las principales obras clásicas:

VRIES, Jan de (2009) *La Revolución industrial: consumo y economía doméstica desde 1650 hasta el presente*, Barcelona : Crítica.

NORTH, Douglas C. (2005): *Understanding the process of economic change*, Princeton N.J: University Press, Princeton.

OSTROM, Elinor (1990): *Governing the commons. The evolution of institutions for collective action*, Cambridge: Cambridge University Press.

Para una discusión de las tesis de De Vries en relación a Cataluña:

MARFANY, Julie (2012), *Land, Proto-industry and Population in Catalonia, c.1680-1829. An Alternative Transition to Capitalism?*, Farnham: Ashgate.

Por lo que respecta a análisis críticos del institucionalismo relacionados con los temas abordados en el proyecto, podemos citar, entre otros:

BÉAUR, Gerard, Jean Michel CHEVET, Philip SCHOFIELD y María Teresa PÉREZ-PICAZO(eds.) (2013), *Property rights, land markets and economic growth in the European Countryside (13th-20th Centuries)*, Turhout: Brepols Publisher.

CONGOST, Rosa, Jorge GELMAN y Rui SANTOS (eds.) (2016), *Property Rights in Land: Issues in Social, Economic and Global History*, Oxford: Routledge.

Una revisión de las distintas aproximaciones al tema de la pobreza y de las clases medias en la literatura sobre los actuales países en desarrollo en:

BANERJEE, Abhijit y Esther DULFO (2007): “The Economic Lives of the Poor”, *Journal of Economic Perspectives*, Vol. 21, n.1, pp.141-168.

RAVAILLON, Martin (2009): “The Developing’s World Bulging (but Vulnerable) Middle Class”, *World Development*, 38, n.4, pp. 445-454.

Por lo que respecta a la especificidad de los libros de la Contaduría de Hipotecas de Cataluña, en relación al conjunto de España, que la convierten en una fuente especialmente apta para el análisis global de una sociedad:

LÓPEZ, Marina y Mercè TÀTJER (1984): “Observaciones sobre la historia de los Oficios de Hipotecas en Cataluña (1768-1861””, *Revista Crítica de Derecho Inmobiliario*, 560 (enero-febrero), pp.131-150.

CONGOST, Rosa (1990): “Una font poc utilitzada: els llibres del Registre d’hipoteques”, *Estudis d’Història Agrària*, núm. 8, pp.201-234.

VILLALÓN, Sebastià (2008): “Els problemes de la manca d’informació en una societat d’Antic Règim. Els notaris davant la creació del registre d’hipoteques”, en CONGOST, Rosa (ed.); *Dels capbreus al registre de la propietat. Drets, títols i usos de la informació a Catalunya (segles XIV-XX)*, Girona: Documenta Universitaria, pp. 241-274.

Pueden hallarse referencias al fracaso de iniciativas parecidas en Europa y, en concreto, en Francia:

HOFFMAN, Philip T., Gilles POSTEL-VINAY y Jean Laurent ROSENTHAL (1998): “What do Notaries do? Overcoming Assimetric Information in Financial Marrkets: The Case of Paris, 1751”, *Journal of Institutional and Theoretical Economics*, Vol. 154, núm. 3, pp.499-530.

POSTEL-VINAY, Gilles (1998) *La Terre et l’argent: l’agriculture et le crédit en France du XVIIIe au debut du XXe siècle*, Paris : Albin Michel

En relación a los retos que suponen las nuevas tecnologías de la información y el análisis de Big Data para los estudios históricos, así como el reconocimiento óptico de textos manuscritos, es interesante seguir con atención las actividades de la Alliance of Digital Humanities

Organizations (ADHO) desde 1998. Entre las revistas más recientes se halla *Big Data and Society*, fundada en 2014. También podemos citar, entre otras muchas, estas aportaciones:

COHEN, D. J. y R. ROSENZWEIG (2006) *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*, Philadelphia: University of Pennsylvania Press.

MANNING, Patrick (2013) *Big Data in History*, London: Palgrave

WELLER, Tony (2013) *History in the Digital Age*, Oxford: Routledge

Finalmente, sobre los avances recientes en el reconocimiento óptico de textos manuscritos, especialmente cuando han sido escritos por una misma mano, puede consultarse la web del proyecto TranScriptorium, financiado por la Unión Europea: <http://transcriptorium.eu/>

6. Plan de trabajo y calendario

Para explorar nuevas vías de detección de cambios sociales importantes, difíciles de detectar por los métodos convencionales, lo que proponemos, es radicalmente distinto de lo realizado hasta ahora pero muy parecido a lo que Bloch definía como trabajo científico: provocar experimentos capaces de trastocar nuestras posiciones de partida. En el cronograma se visualizan las cuatro fases principales del proyecto: 1. Fase de digitalización de los libros del Registro general de escrituras de las Contadurías de Hipotecas; 2. Fase de reconocimiento de imágenes, en que las imágenes digitales serán sometidas a un proceso de reconocimiento de texto manuscrito que permita obtener una versión textual de las palabras contenidas en cada imagen; 3. Fase de preprocesamiento analítico; y 4. Fase de análisis de datos. La base creada que, al final del proyecto, puede contener aproximadamente unos 200 millones de palabras, podrá y deberá ser explotada de distintas formas y mediante distintas herramientas.

6.1. Fase de digitalización de los libros de la Contaduría de Hipotecas. En concreto, el proyecto va a centrarse en el análisis del partido hipotecario de Girona. En conjunto, el área estudiada abarca una extensión de 2.900 km², con unos 123.000 de habitantes en 1887. Se prevé aprovechar la digitalización ya realizada por el Archivo Histórico de Girona, que alcanza los libros de 1768 a 1780, lo que justifica la selección de este distrito. Las tareas propiamente de digitalización comprenderían el período 1781-1800 y permitirán añadir unas 20.000 imágenes. La obtención de las imágenes digitales deberá realizarse en las instalaciones del Arxiu Històric de Girona, que alberga la fuente, y se prevé externalizar dicha operación para encargarla a una empresa especializada que dispongan de la infraestructura de escaneo adecuada. Se prevé que dicha empresa entregue cada imagen en tres estándares de calidad (de alta calidad para su uso en el proceso de transcripción, de calidad estándar para un uso interno y de menor calidad y resolución para su publicación en la web). El estándar de calidad alta será de imágenes sin compresión (TIFF), a 300 ppp y color 24 bits, la calidad media la definimos como imágenes comprimidas (jpeg), a 150 ppp y color 24 bits; y, finalmente, el tercer formato tendrá mayor compresión (jpeg) y una resolución de 70 ppp.

Se proporcionará al Arxiu Històric de Girona una copia de todas las imágenes obtenidas con los volúmenes procedentes de sus fondos documentales para que sean puestas a disposición pública. El archivo se encargará de realizar todas las operaciones necesarias para ofrecer la documentación a través del servicio de archivos en línea de la red de archivos públicos catalanes (<http://arxiusenlinia.cultura.gencat.cat/>).

6.2. Fase de transcripción automática. Las imágenes digitales serán sometidas a un proceso HTR (Handwritten Text Recognition) basado en el conjunto de herramientas desarrolladas por el proyecto READ y el grupo PRHLT de la UPV. En el último año (mayo 2016-julio 2017) hemos estado realizando distintas pruebas y ensayos con la finalidad de poner a punto el sistema informático específicamente para su uso con la documentación del Registro de Hipotecas. Aunque quedan algunos meses de trabajo, es probable que cuando se resuelva esta convocatoria la fase de entrenamiento del sistema esté ya finalizada y pueda empezarse la transcripción automática de los libros de la Contaduría de Hipotecas de forma sistemática.

La documentación manuscrita, y específicamente la documentación histórica, presenta una problemática específica por distintos motivos: irregularidades en el estilo de escritura, fluctuación de las líneas de texto, presencia de elementos ornamentales (por ej. papel sellado), deterioro del soporte (manchas de humedad, hongos) o de la tinta... El sistema desarrollado en los últimos años es capaz de detectar, en primer lugar, los elementos que componen cada imagen para discriminar las zonas de texto de las no textuales. Para ello, las imágenes deben ser preprocesadas para eliminar los bordes, las manchas y otros elementos que puedan considerarse ruido, así como para corregir las distorsiones generadas en la captura, binarización y normalización. A continuación, se debe proceder a un proceso de segmentación de la imagen para identificar las áreas que contienen texto y, dentro de ellas, las líneas que la estructuran. Esta etapa se subdivide en dos pasos: en primer lugar, la detección y delimitación de los bloques de texto; después la identificación de las líneas de texto (*baselines*). Se trata de una fase crucial, dado que el éxito del reconocimiento de texto manuscrito reposa fuertemente en la *accuracy* conseguida en el proceso de segmentación de líneas. Actualmente se está entrenando al sistema con imágenes de la Contaduría de Hipotecas de Girona para que sea capaz de identificar automáticamente los bloques y líneas de texto de estos libros.

La siguiente etapa del proceso es propiamente la etapa HTR. El sistema implementado por el equipo PRHLT de la Universidad Politécnica de Valencia hasta ahora se ha basado en modelos ocultos de Markov y N-grams, y en la consideración de grafos de palabras (Word graphs) en vez de caracteres aislados. Las pruebas con este sistema están más avanzadas y dan resultados muy satisfactorios, aunque el equipo PRHLT se están planteando un cambio de tecnología para aplicar modelos de redes neuronales que aún podrían mejorar los resultados de transcripción. El resultado obtenido es un archivo de texto en ASCII (texto plano) que puede enriquecerse añadiendo automáticamente etiquetas xml que permiten mejorarlo, delimitarlo y estructurarlo. Actualmente hemos ensayado con resultados muy satisfactorios la generación automática de etiquetas xml para distinguir los distintos tipos de bloque de texto (tipología, texto principal, nota marginal, número de página) y para remarcar algunos elementos recurrentes de especial valor analítico (antropónimos, topónimos, profesiones y tipología documental). También hemos aplicado el mismo procedimiento para conseguir desarrollar automáticamente los términos que aparecen abreviados en el documento original.

Antes de alcanzar un estadio de automatización el sistema requiere una etapa de aprendizaje tanto en lo que se refiere a la identificación de las zonas de texto y de las líneas en que estas se organizan, como en lo referido al reconocimiento del texto. Para la prueba inicial (con un lote de 50 imágenes) se utilizó la herramienta *TransKribus*, creada por el proyecto READ (<https://transkribus.eu/Transkribus/>), que aún se halla en fase de desarrollo. Mediante *TransKribus* se introduce manualmente la transcripción de una muestra de los documentos con

la finalidad que el sistema aprenda y que pueda generar una pauta o *Ground Truth* (GT) con la cual testear los resultados de la transcripción automática. A partir de esta primera pauta, el sistema informático ya es capaz de transcribir automáticamente las imágenes, aunque la calidad de los resultados es mejorable y por ello debe seguirse entrenando al sistema. Para aprovechar el aprendizaje realizado, en vez de continuar introduciendo transcripciones manuales que sirvan de pauta de aprendizaje y testeo, muy costosas en tiempo, los pasos siguientes se basan en la corrección de los resultados de la transcripción realizada automáticamente por el sistema. Para ello se ha utilizado la aplicación CATTI (*Computer-Assisted Transcription of Text Images*), desarrollada por el grupo PRHLT de la UPV. Esta herramienta permite un proceso interactivo de corrección de los errores de la transcripción automática. Cuando el usuario detecta un error de transcripción, CATTI sugiere otras opciones posibles y el usuario puede escoger la correcta. Estas correcciones son incorporadas al sistema para acrecentar su nivel de información y aprendizaje. Cuando finalice la fase de aprendizaje del sistema, se espera obtener un resultado por encima del 90% de acierto, teniendo en cuenta que buena parte de lo que se computa como error con frecuencia no compromete la correcta comprensión del texto.

Este proceso de aprendizaje del sistema no sólo deberá llevarse a cabo en una fase inicial, ya que es probablemente que la calidad de los resultados disminuya cada vez que, en la documentación, se produzca un cambio de registrador o de estilo de escritura. Cuando suceda, deberemos volver a interactuar con el sistema para restablecer la calidad de la transcripción.

6.3. Fase de preprocesamiento analítico. El corpus textual que pretendemos generar, correspondiente a los libros de la Contaduría de Hipotecas de Girona entre 1768 y 1800, puede contener, aproximadamente unos 13 millones de palabras. Para su explotación se utilizarán herramientas de minería de datos y de textos que permitirán contabilizar frecuencias, analizar concordancias, etc. Nuestra apuesta, sin embargo, es intentar dotar al texto plano de una estructura mínima que permita su procesado mediante el lenguaje de bases de datos (SQL). El etiquetado automático en xml del texto transcrito (identificando las regiones del texto, la tipología documental, los antropónimos, los topónimos y los oficios) facilitará su transformación en una base de datos de tipo relacional, convirtiendo las etiquetas en campos. Posteriormente se definirán, para cada tipología documental, una serie de reglas estructurales y formales que permitirán la identificación automática de los actores que intervienen en cada documento (por ejemplo, el notario, el comprador y el vendedor en cada compraventa) y otros elementos relevantes (la fecha de registro, la fecha y lugar del acta notarial original, el precio...).

6.4. Fase de análisis de datos. Aunque los resultados de la transcripción y la base que de ellos se derive permitirán analizar muchas cuestiones y de diversa índole, durante los dos años de ejecución del proyecto se intentará avanzar en tres líneas de análisis histórico:

Detección de cambios sociales ocurridos durante el período. Algunos pueden hacerse visibles a partir de la simple observación de los datos contenidos en la fuente. Por ejemplo: aparición o desaparición de oficios o etiquetas sociales; aparición o desaparición de diferentes modalidades de contratos. Otros requerirán un trabajo más minucioso: por ejemplo, cambios en la evolución de las dotes, o en las cláusulas de contratos que revelen iniciativas o capacidad de negociación por parte de los contratantes, independientemente de su etiqueta social. Ciertos cambios pueden

revelar iniciativas de sectores poco visibles de la población, por ejemplo, las cláusulas relativas a las viudas en los contratos matrimoniales.

Captura de los momentos y lugares decisivos en que se produjeron algunos de los cambios sociales detectados. Esta captura permitirá ver la importancia que tuvieron en el desencadenamiento de algunas dinámicas sociales determinadas coyunturas, como guerras, carestías, o, determinadas disposiciones legislativas. A la inversa, este ejercicio también permitirá comprobar si y hasta qué punto algunas disposiciones legislativas o algunas coyunturas específicas tuvieron los efectos que se les suponen.

Seguimiento de trayectorias particulares de individuos, familias o instituciones, interrelacionando los diversos tipos de actuaciones y operaciones que éstos escrituraban ante notario. El análisis del universo de trayectorias individuales permitirá avanzar substancialmente en la comprensión de las dinámicas sociales, los procesos de empobrecimiento o de enriquecimiento y la emergencia de nuevos grupos sociales. El tratamiento al que habitualmente se somete la documentación notarial o registral sólo permite el análisis aislado de los diferentes tipos de operaciones, sin que sea posible estudiar la actuación global de los individuos o familias. Una aproximación de este tipo hasta ahora sólo es posible si se dispone de archivos privados patrimoniales, que no acostumbran a ser abundantes en los casos de los grupos humildes. Por esto, nuestra aportación en este terreno será también novedosa.

Podemos entrever las enormes posibilidades de las tres líneas de trabajo seleccionadas, a partir de dos ejemplos, uno relativo a un tema clásico, como es el de los derechos de propiedad, otro, relativo a un tema nuevo, como es el del surgimiento de nuevas etiquetas. Tanto en un caso como en otro, los resultados que obtengamos pueden tener una incidencia en muchos trabajos realizados hasta ahora en torno a temáticas relacionadas con el cambio social. Por ejemplo, muchos trabajos sobre conflictos sociales han tomado como referencia las etiquetas de los individuos y muchos trabajos sobre desigualdades sociales han partido de estadísticas relacionadas con la propiedad de la tierra. Podemos seguir con el mismo ejemplo, insistiendo ahora en las posibilidades que ofrece el hecho de trabajar sobre un área geográfica tan extensa como la que abarca el distrito hipotecario de Girona y, en el futuro, con distintas áreas geográficas. Las etiquetas sociales, ¿eran las mismas en todas las áreas en el punto de partida? ¿evolucionaron del mismo modo? Y si no lo hicieron ¿significa ello que los individuos que las llevaban no experimentaron ninguna evolución? El mismo tipo de preguntas pueden hacerse respecto a las distintas modalidades de contrato de acceso a la tierra, en relación a los derechos de propiedad, y, de hecho, en un número incalculable de temáticas sobre las que la fuente proporciona información significativa.

Cronograma

	captura imágenes digitales				Transcripción automática		Preprocesamiento analítico			análisis de datos			difusión del proyecto			
	Selección del material de archivo	contratación de la empresa que se encargará de la digitalización	proceso de digitalización	almacenado y transferencia de archivos de imagen	Execució i control del procés de transcripció automàtica	Correcció dels resultats del procés de transcripció	Definició de las estructuras de cada tipología documental	Diseño de la base de datos y de los algoritmos de identificación	Ejecución del preprocesamiento	Detección de cambios sociales	Captura momentos de ruptura	reconstrucción trayectorias	campana de difusión	construcción y mantenimiento web	campana de colaboración ciudadana	comunicaciones a congresos y artículos académicos
mes 1																
mes 2																
mes 3																
mes 4																
mes 5																
mes 6																
mes 7																
mes 8																
mes 9																
mes 10																
mes 11																
mes 12																
mes 13																
mes 14																
mes 15																
mes 16																
mes 17																
mes 18																
mes 19																
mes 20																
mes 21																
mes 22																
mes 23																
mes 24																

El primer año del proyecto se dedicará preferentemente a la primera y a la segunda fase (captura de imágenes y transcripción automática), aunque a medida que se disponga del corpus documental transcrito podrá avanzarse en las tareas de preprocesamiento. Al finalizar el primer año del proyecto, prevemos estar en condiciones de iniciar la ejecución de las tareas de procesamiento, a las cuales se dedicará buena parte del segundo año. Durante este segundo año también se concentrarán las tareas de carácter analítico, dado que es necesario disponer de la mayor parte de los datos para iniciarla. A lo largo del desarrollo del proyecto también se prevén distintas actividades de difusión –que se detallan en el punto siguiente- tanto del propio proyecto como de sus resultados científicos.

7. Plan de difusión y explotación de resultados.

Es necesario distinguir, en relación a su proyección social, entre los resultados de carácter más técnico del proyecto y aquellos que tienen que ver con el corpus documental y la información proporcionada.

Respecto a los resultados técnicos, tenemos muy presente que actualmente ya hay un importante número de portales y bibliotecas digitales que ofrecen la posibilidad de bucear en el contenido de la documentación histórica, pero con una restricción clave: que se trate de letra impresa. Como ya se ha dicho, las tecnologías OCR, que reconocen carácter a carácter, sólo son operativas en textos impresos, pero no son efectivas en documentación manuscrita. Sin embargo, la mayor parte de la documentación histórica es manuscrita y, en consecuencia, hasta ahora ha quedado al margen de los avances tecnológicos que han permitido poner a disposición pública (on-line) documentación del pasado. El impacto básico de la investigación que aquí se propone, y no nos parece menor, es que ayudará a desarrollar unos instrumentos que facilitaran este tipo de acceso, hasta ahora limitado a los textos impresos, también a la documentación manuscrita. Evidentemente el proyecto que se presenta sólo se refiere a un corpus documental muy específico (los libros del Registro de Hipotecas), pero su aplicabilidad puede extenderse a cualquier otro tipo de texto manuscrito.

En un sentido genérico, la posibilidad de ofrecer una transcripción del contenido de los documentos históricos tiene dos efectos muy positivos:

- a. Reduce las barreras y los costes de acceso a la documentación histórica por parte de colectivos sociales que no tienen una preparación específica. Si bien no elimina los problemas de interpretación conceptual y léxica asociados al vocabulario de cada época (y en nuestro caso, a la lengua específica de los notarios y registradores), contribuye a hacer más accesible el acceso a la documentación histórica en tanto que reduce las dificultades que a menudo presentan la caligrafía y las abreviaturas.
- b. Permite el manejo masivo de la documentación histórica a partir de búsquedas selectivas y la aplicación de técnicas de minería de datos. Este aspecto ofrece el máximo interés para investigadores académicos, doctorandos, investigadores locales, estudiantes universitarios o incluso estudiantes de secundaria, que podrán disponer de un potente instrumento de búsqueda de datos históricos.

En segundo lugar, el corpus documental que se propone utilizar (los libros de la Oficina de Hipotecas del distrito hipotecario de Girona, entre 1768 y 1800) constituye un conjunto documental muy rico para conocer en profundidad las relaciones sociales, los flujos de información, los matrimonios, los mercados de tierra y de dinero y muchos otros aspectos de la vida cotidiana de gran parte de la población (de todos aquellos que acudían a las notarías en un período en el que la figura del notario, a causa del bajo nivel de alfabetización, tenía un rol intermediario muy superior al actual). Consideramos que hay cuatro colectivos interesados en la utilización de este corpus:

- a. Investigadores académicos, doctorandos y estudiantes de los últimos cursos de Grado y de Máster que estén interesados en las humanidades digitales. Para ello, se procurará integrar las metodologías utilizadas en este proyecto en el plan de estudios del Máster en Investigación en Humanidades de la Universitat de Girona, específicamente en la asignatura *Noves tecnologies aplicades a les ciències humanes*.
- b. Investigadores académicos, doctorandos y estudiantes de los últimos cursos de Grado y de Máster, y también investigadores locales y estudiantes de bachillerato, interesados en realizar investigaciones, con fuentes originales, sobre la segunda mitad del siglo XVIII o sobre localidades, instituciones, individuos o grupos sociales concretos, topónimos,

etc. En su conjunto, el corpus generado puede ser consultado para la realización de investigaciones de carácter muy diverso: biográficas, de historia económica y social, de historia del arte, etc.

- c. Genealogistas interesados en el seguimiento de apellidos y familias concretas. El colectivo de personas –amateurs o profesionales- interesadas en la genealogía es muy amplio y tiene un perfil claramente internacional, como ponen de relieve los potentes portales dedicados a la reconstrucción genealógica (como My Heritage) o el espectacular FamilySearch creado por la Iglesia de los mormones a partir de documentación procedente de todos los continentes.
- d. Por último, es necesario considerar el interés que este proyecto ofrece a las Instituciones que custodian documentación histórica y a los profesionales que las gestionan. Archivos y archiveros, además de preservar la documentación histórica y velar para que las condiciones para su conservación y localización sean óptimas, también están interesados en la divulgación de este fondo mediante la plataforma Arxiu en línea del departament de Cultura de la Generalitat de Catalunya (<http://arxiusenlinia.cultura.gencat.cat/ArxiuEnLinia/>). Concretamente, el Archivo Histórico de Girona, miembro colaborador de este proyecto, ya ofrece, entre otros documentos, la consulta online de los tres primeros años (1768-1770) de la serie de libros del Registro General de la Contaduría de Hipotecas de Girona.

Plan específico de difusión de resultados:

1. Ampliación de los fondos documentales digitalizados por el Arxiu Històric de Girona. Se cederá al Arxiu Històric de Girona copia de las imágenes digitales obtenidas. Las imágenes se ofrecerán en régimen de acceso libre y abierto a la web, a través de la plataforma Arxiu en línea. El número de beneficiarios puede ser muy elevado, si tenemos en cuenta que, ya en estos momentos, la diversidad, calidad y cantidad de los fondos documentales conservados y gestionados por el Arxiu Històric de Girona (más de 10.000 metros lineales y con documentos a partir del siglo XI), lo sitúan como un centro de referencia a nivel nacional e internacional. Los servicios de atención al usuario generan un impacto notable en la comunidad científica y en la ciudadanía. Los datos que disponemos, relativos a 2016, muestran que 15.714 usuarios virtuales exploraron sus bases de datos en línea, consultando remotamente 10.553 documentos en formato digital. El portal web de l'Arxiu Històric de Girona, también en 2016, recibió 3.853 visitas; sus perfiles de redes sociales, Facebook y Twitter, contaban con 208 y 1.726 seguidores, respectivamente, y recibieron 2.627 y 22.417 visitas.
2. Campaña de difusión del proyecto y de los instrumentos creados para consultar la documentación registral dirigida a centros de estudios locales y comarcales, a estudiantes del grado de Historia y a investigadores. Presentación del proyecto y los instrumentos creados en jornadas dirigidas a archiveros (500 beneficiarios potenciales).
3. Creación de una página web que proporcione la información necesaria sobre las distintas fases del proyecto: el proceso de transcripción y etiquetado, su alcance cronológico y geográfico, las características de la Contaduría de Hipotecas, las tipologías documentales... También se explicará cómo acceder a los diferentes

instrumentos de consulta (transcripción etiquetada en texto plano, bases de datos...) y se establecerán las condiciones de este acceso. Igualmente, se publicitarán los resultados obtenidos, tanto los generados por el equipo del proyecto como los elaborados por usuarios externos que hayan accedido a los datos (1.000 beneficiarios potenciales).

4. Difusión del proyecto a través de distintas redes sociales: redes de archiveros, portales de genealogía e historia, etc. Generación de noticias periódicas vinculadas al desarrollo del proyecto y a sus resultados. (10.000 beneficiarios potenciales).
5. Realización de una campaña de ciencia ciudadana en la que se buscará la colaboración externa de estudiantes, genealogistas e investigadores amateurs para invitarlos a explotar los datos acumulados reconstruyendo historias individuales y familiares. Los resultados se almacenarán y de publicarán a través de la misma web del proyecto (50 participantes).
6. Difusión de los resultados en publicaciones científicas y académicas. A corto plazo (durante los dos años del proyecto) se prevé presentar los primeros resultados en tres congresos internacionales, que han de dar lugar, al menos, a tres publicaciones en revistas académicas.

8. Presupuesto

Como se ha ido explicando, la realización del proyecto requiere la colaboración continuada con el Archivo Histórico de Girona, para la fase de digitalización de los libros del Registro de Hipotecas, ya iniciada, y con el equipo PRHLT de la Universitat Politècnica del País Valencià, para la gestión del sistema de reconocimiento automático de los textos manuscritos digitalizados. Se ha solicitado cofinanciación en otra convocatoria (RECERCAIXA 2017).

El pla Bruniquer (para la digitalización de documentación histórica en los archivos) de la Generalitat de Catalunya ha permitido el escaneo de los libros correspondientes al período 1768-1781. Sin embargo, para desarrollar el proyecto necesitaríamos digitalizar otros veinte volúmenes. Dadas las características de la fuente, la concesión de la otra ayuda solicitada nos permitiría avanzar considerablemente (otros veinte años) en el volumen de la documentación a analizar, con lo que los resultados del proyecto podrían ser más espectaculares.

La colaboración con el equipo valenciano es fundamental en la segunda fase, correspondiente a la de transcripción automática de textos. En parte, esta fase se halla asegurada por el convenio de colaboración entre el CRHR y el equipo del PRHLT de la Universidad Politécnica de Valencia, en el seno del proyecto europeo READ. El grupo de la UdG es MOU partner en dicho proyecto europeo. Además, la ayuda ICREA Academia 2013 permite cofinanciar algunas tareas básicas asociadas durante los dos próximos años. La concesión del proyecto que se solicita nos permitiría contratar a personal de soporte para acelerar significativamente esta fase y supervisar sus resultados.

Finalmente, una parte significativa de los gastos de personal se dedicarán principalmente a las fases tercera y cuarta, consistentes en las tareas de preprocesamiento analítico y de explotación de los datos.

Funcionamiento	
Escaneo de 20 volúmenes de la Contaduría de Hipotecas de Girona (correspondiente a veinte años), mediante la contratación de los servicios de una empresa especializada.	20.000,00
Construcción y mantenimiento de una página web del proyecto.	3.000,00
Gastos de personal	
Contratación de un investigador postdoctoral a tiempo parcial para gestionar el sistema de reconocimiento automático de textos manuscritos	19.000,00
Becas de colaboración (predoctorales) a estudiantes para supervisión de los resultados de la transcripción automática y para tareas de preprocesamiento analítico y análisis de datos correspondientes a la Fases 3 y 4 del proyecto.	19.000,00
Otros	
Gastos de viaje (Girona-Valencia)	900,00
Canon Universitat de Girona (21%)	12.999,00
SUMA	74.899,00