

## RIGOR METODOLÓGICO E INVESTIGACIÓN FILOLÓGICA<sup>1</sup>

COLOMA LLEAL GALCERAN  
*Universitat de Barcelona*  
colomalleal@gmail.com

### RESUMEN:

En las últimas décadas, los avances técnicos han puesto a nuestro alcance abundante documentación. Pero el filólogo debe ser consciente de los peligros de una deficiente utilización de los corpus documentales a fin de evitar la creación de los temibles *fantasmas lexicográficos*. En este artículo se recuerdan las principales fuentes de este tipo de error: fenómenos de etimología popular por parte de los hablantes, confusiones de los copistas, errores de los transcritores en la interpretación de abreviaturas y variantes gráficas de los manuscritos, alteraciones onomásticas introducidas por los cartógrafos por desconocimiento de las modalidades lingüísticas, deficiencias en la localización cronológica de algunos documentos, confusiones en el proceso de lematización y de valoración de los textos..., todos ellos contribuyen, en mayor o menor medida, a falsear o a enmascarar los datos en que el filólogo basa su investigación. De ahí la importancia del rigor filológico en la transmisión y estudio de los textos antiguos.

**PALABRAS CLAVE:** Lingüística de corpus, anotación de corpus, fantasmas lexicográficos.

**ABSTRACT:** In recent decades, technological advances have made extensive documentation available to us. But the philologist must be aware of the dangers of poor use of the documentary corpus in order to avoid creating dreaded ghost words. In this paper we recall the main sources of this type of error: folk etymology phenomena among speakers, copyists' errors, transcribers' errors in the interpretation of some abbreviations and graphic variants of the manuscripts, onomastic changes introduced by cartographers' ignorance of linguistic variants, gaps in the dating of some documents, confusion in the processes of lemmatization and the evaluation of texts... All these sources of error contribute, to a greater or lesser degree, to the distortion or to the masking of the data on which the research of philologists is based. Hence the importance of philological rigour in the transmission and study of ancient texts.

**KEY WORDS:** Corpus linguistics, corpus annotation, ghosts words.

### 1. INTRODUCCIÓN

Decir, a estas alturas, que la filología no puede entenderse sin el estudio paciente de la documentación puede parecer una afirmación redundante. Pero no es, en absoluto, gratuita. Quienes partimos del análisis directo de los documentos podemos seguir investigando con nuevas herramientas y brillantes presentaciones. Ya no nos llenamos de polvo ni nos hemos de levantar para pedir «la lámpara» al conserje para intentar adivinar los fragmentos borrados, por el tiempo o por investigadores anteriores. Hemos dejado de ser ratas de archivo (o, al menos, de serlo a tiempo completo). Pero esto es solo la apariencia. Porque seguimos atados al texto y a sus dificultades.

El lingüista no filólogo seguirá citando a Isidoro como ejemplo de etimologías pintorescas y probablemente se reirá recordando la frase:

*Musio appellatus, quod muribus infestus sit. Hunc vulgus cattum a captura vocant. Alii dicunt, quod catta, id est videt. Nam tanto acute cernit ut fulgore luminis noctis tenebras superet. Vnde a Graeco venit catus, id est ingeniosus (Isidorus Hispalensis, *Etymologiarium sive originum*, liber XII: De animalibus, cap. II: De bestiis, l. 38).*

---

<sup>1</sup> Reelaboración del tema presentado como conferencia plenaria en el *Tercer Congreso Internacional del grupo CHARTA: Tradición e Innovación. Nuevas perspectivas para la edición, la investigación y el estudio de documentos antiguos*. Salamanca, 5-7 de junio de 2013.

El filólogo, en cambio, celebra en este *cattat* el hallazgo de una clara documentación de la asimilación del grupo -PT-, y del cambio semántico de *captare*, ‘tratar de coger, dar caza’ a ‘coger con la mirada, ver’, que anuncia el valor del castellano *catar*.

Atados al texto y, por ello, a los errores de que nos hablaba Kurt Baldinger en el discurso inaugural del Primer Congreso Internacional de Historia de la Lengua Española (Baldinger 1987) y a los que también se refería el maestro Joan Coromines en las primeras páginas de *Tópica hespérica* (aquí, refiriéndose a la onomástica prerromana, pero aplicable a toda labor filológica):

En el terreno de la onomástica prerromana hemos errado todos, unos más, otros menos. Y seguiremos pecando. [...] Yo también he pecado a veces, como todos. [...] Hay que suponer que habré cometido no solo varios errores, sino incluso quizás algún error evitable. Seguramente menos que otros, por haber echado mano de dos recursos: por haberme limitado a nombres de armazón consonántica sólida, larga y compleja (donde es más difícil que se produzca el juego de espejuelos de la homonimia), y por haber buscado previamente documentación medieval (Coromines 1972).

Errores evitables, pero que nos asechan por todas partes convirtiendo los documentos en un mar de minas en que algunos naufragan. De ahí la necesidad de recordar el origen de los principales errores que pueden desorientar al investigador o falsear el resultado de su estudio. Veamos, a continuación, en qué medida y mediante qué procesos, hablantes, copistas, transcritores, cartógrafos, fechadores, lematizadores... nos tienden sus trampas empujándonos hacia el fantasma lexicográfico o hacia el étimo fantasioso (Bastardas 1995, Reeve 1994).

## 2. ERRORES DEL HABLANTE

De todos ellos, el hablante es el menos culpable, por cuanto no hace más que aplicar con rigor y coherencia los mecanismos derivados de su conocimiento de la lengua, llevándolos, a veces, hasta sus últimas consecuencias, o más allá aún. Consciente en su ser más íntimo de la verdad de la afirmación de que en una lengua «*tout se tient*», mucho antes de que la formulara Meillet (1893: 318-19), intenta buscar el entramado de relaciones que unen un término con su entorno, negándose a aceptar la existencia de unidades aisladas (Maiden 2008).

Los filólogos sabemos que, a veces, como consecuencia de la evolución fonética, una palabra se separa del grupo al que inicialmente había pertenecido, pero los hablantes no lo aceptan fácilmente e intentan encontrarle un nuevo grupo. Entonces, como consecuencia de diversos cruces y (re)motivaciones, los bosques se pueden convertir en valles, las casas en hombres y la vida en viajes.

Así, es de todos conocido que los *berrojos*, separados de un arcaico *veru*, se acercaron a *cerrar*, como el *\*codalobo* (< CODA LUPI) perdió la cola y, en cambio, ganó volumen al pasar a *gordolobo*. Y el *mayordombre* de tantos documentos aragoneses nos indica que, al adaptar el catalán *majordom* (< MAIOR DOMUS) los hablantes reinterpretaron el resultado de *domus* como si se tratara de un derivado del catalán *d'hom*. Así también

actuaba la buena mujer de quien nos hablaba hace años Sebastián Mariner, contenta porque su marido, después de haberle dado unos *antiviáticos*, se encontraba mucho mejor. Y de ahí también los múltiples resultados de la ‘niña del ojo’, en catalán, donde *la nineta* se convierte en *l’animeta* (Veny 2000), o el *fel sobreixit*, ‘ictericia’ deja de relacionarse con *sobreixir*, ‘rebosar, derramarse’ para transformarse en el mucho más prosaico *fel sofregit*, ‘sofrito’ (Veny 1990). La misma actitud que ya en periodo clásico había transformado la *malacia*, ‘calma’, en el vulgar *\*bonacia* (> *bonanza*), por considerar errónea la relación del tiempo apacible con *malus*, un tiempo necesariamente bueno y favorable, sobre todo en la mar. Pero *malacia* era un préstamo del griego, derivado de *μαλακός*, ‘blando, suave’, y nada tenía que ver con *malus* ni con *bonus* (DECLC: s.v. *bonanza*), contrariamente a lo que pudieran creer los hablantes.

Y esto es particularmente sensible en el caso de los topónimos (Baldinger 1986), en los que a menudo perviven elementos desaparecidos del resto de la lengua, y de ahí la necesidad de reinterpretarlos. Así, el catalán *Les Poses*, alejado del valor inicial de ‘lugar de descanso’ (< PAUSAS), acaba convirtiéndose en el mucho más comprensible, pero también más enigmático, *Les Esposes*. Y el vascónico *Baso aran* (‘bosque’ + ‘valle’), en el valle de Benasque, tras evolucionar al opaco *Basarán*, con el sentido de *aran* perdido ya en la conciencia de los hablantes de esas áreas pirenaicas, es reinterpretado como *Bacheran*, con un valle reencontrado, aunque con la particular evolución retrofleja de la palatal lateral propia de esta zona. De la misma manera que el catalán *Saltalinya* (< SALTU ILICINA), se aleja de sotos y encinas y acaba beatificado en *Santalinya*, el burgalés *Somotero* (SUMMU ALTARIU) se transforma en *San Otero* (Veny 1990) y el andalusí *Qa<sup>c</sup>ta al-<sup>c</sup>ayn*, (‘lugar de la fuente’) de Murcia, se convierte en *Santa Catalina* (Bramon 2013), hermanados así con los numerosos hagiopónimos de nuestra geografía.

A estos ejemplos, habría que añadir los casos de falsos arabismos, por una parte, y de arabismos oscurecidos, por otra. Así, ante los tan abundantes topónimos con el elemento árabe *ibn*, ‘hijo’ (o su plural *banu*), como *Vinaixa* (*Ibn <sup>c</sup>A’iša*), *Benicarló* (*Ibn Gazlun*), *Binicalaf* (*Ibn Khalaf*), o *Binissalem* (*Ibn Salam*), las antiguas *Pinna Cadiella* y *Pinna Riscu* de los documentos medievales se han arabizado en *Benicadell* y *Binarrisc*, respectivamente, al tiempo que, quizás para compensar, el arábico *Banu <sup>c</sup>Adila* se transformó en *Penàguila* (Bramon 2013, Barceló 1991).

Ante fenómenos de este tipo, solo el análisis de la documentación antigua puede evitar que postulemos étimos fantasmagóricos, mucho más risibles que el *cattus* isidoriano.

### 3. ERRORES DE LOS COPISTAS

Un segundo grupo de errores, quizás los más peligrosos, se los debemos a los copistas, que escribían como buenamente podían las palabras que desconocían y que, ellos también, intentaban motivar. Por ello, un respeto reverencial a la fidelidad al texto puede, una vez más, generar nuevos fantasmas (Sánchez Prieto, 1996).

Claro está que, cuando en un texto moralizador se alude a una obra titulada *Gestis angelorum*, entre la conservación, sin más, del término, o la corrección sin comentario alguno, cabe una tercera vía que permita identificar la obra pero, al mismo tiempo, poner de manifiesto la perplejidad del copista, porque ello, probablemente, también forma

parte del entorno de la obra<sup>2</sup> (Castillo Lluch 2006). De una obra que nos habla de la historia eclesiástica del pueblo de los anglos (*Gentis Anglorum*), y no de las hazañas de los ángeles:

- (1) Item cuenta Beda en el libro de **Gestis angelorum**: que en el tiempo de Constantino el Menor. Un caullero fallecio cerca del año del nacimiento de nuestro señor .lxxxvj. el qual restituydo despues a la vida: (*DiCCA-XV: Cordial de las cuatro cosas postrimeras*, 049v [1494]).

De la misma manera que habrá que restituir su nombre a un pobre *Semuel* de Miranda de Ebro, convertido en *senil* sin darse cuenta<sup>3</sup>, o agrupar los dos componentes de un sorprendente *San Son*, si no queremos crear un nuevo miembro del santoral:

- (2) El angel empulla al monte Gargano consagrando aquel: el angel en Roma mostrando se çima el castillo de Santangel y d-aquj tomo el nonbre d-aquel: el angel denunciando la conçeçion de **san Son** y sant Juhan Batista en el templo a Zaquarjas su padre: e por no dar fe a las palabras d-aquel enmudeçio fasta el tiempo fue complido. (*DiCCA-XV: Triste Deleytaçion*, 097v [1458-67]).

En cualquier caso, la constatación de la existencia de estos errores no debería empujarnos al extremo opuesto, donde, llevados por una idealizada concepción del autor como emisor perfecto, sin errores ni vacilaciones, intentáramos rehacer el texto expurgándolo de todo aquello que no coincida con el esquema forjado previamente: el texto, entonces, entendido no como documento a partir del cual podamos deducir las características de la lengua, sino como modelo idealizado de nuestra propia concepción de la lengua.

#### 4. ERRORES DE LOS TRANSCRIPTORES

En tercer lugar, quienes transcriben los documentos no están exentos de incurrir en error. Ponen a nuestro alcance un sinfín de textos, pero en ocasiones el inevitable cansancio les ha impedido ‘leerlos’ realmente, y de ahí algunas formas disparatadas, como las ‘*medias finas*’ (en lugar de *bassins*, ‘bacines’) denunciadas por Baldinger (Baldinger 1965), resultantes de un doble error: confusión de <S> larga con <F>, y errónea separación de las dos sílabas de una misma palabra como consecuencia del salto de línea:

- (3) *troys petis baf*  
*fins dor*

En este juego de errores no puede sorprendernos la cantidad de *indios* que pueblan algunos documentos, en contextos que no parecen justificar el error:

<sup>2</sup> De lo que Cesare Segre denominaba el *diasistema* del texto, es decir, de la mezcla de rasgos propios del sistema del autor más los rasgos propios del sistema del copista (o de los sistemas de los copistas) (Segre 1976: 285).

<sup>3</sup> Según la comunicación presentada por Clara Grande López, de la Universidad de Salamanca, en el *Tercer Congreso Internacional de CHARTA*. Salamanca, junio de 2013.

- (4) Deffendemos que ningún christiano, nin **indio**, nin otro omne ninguno non sea osado de comprar (CORDE: Alfonso X: *Fuero Real*, 1251-1255).
- (5) LEY VIIIª. Por toda petición, si **indio** o cristiano, fasta .I. morauedi, iure el cristiano sin cruz e el **indio** sin la tora e si la petición fuere de .I. morauedi arriba...  
Todo omne que agraz uendiere ante que las uinnas sean uendimiadas, peche .I. morauedi o sea cristiano o **indio** o moro. E esta calonna aya el querelloso con los mayordomos del conçello. (CORDE: *Fuero de Plasencia*. c 1300).
- (6) [...] por acto publico, recebido e testificado por hun **indio**, llamado Yaco de la Rabiça, notario, vezino de la dicha ciudad de Tudela (CORDE: Documento notarial, 1492).

Y si, en estos casos, no resulta difícil detectar el error, porque la confusión se establece entre términos de significado conocido, las dificultades se multiplican cuando, como consecuencia de la lectura errónea, se postula la existencia de un término nuevo. Entonces, si los lexicógrafos aceptan directamente alguno de ellos, sin sospechar al menos ante su carácter de *hapax legomena*, permiten que numerosos fantasmas se paseen impunemente por los diccionarios (Colón 1969). Muchos de ellos, debidos a un desarrollo erróneo de algunas abreviaturas. Como el misterioso *subera*, a partir de la grafía *subã* (Baltar Veloso 1986), donde la lineta no reemplaza la sílaba *-er-*, sino el conjunto de las sílabas interiores: *stanti*.

O las frecuentes confusiones de *por* y *pre*, una de las cuales, a pesar del evidente contrasentido, engañó a Moll. Así, en la edición de la *Carta de sant Bernat a sa germana*, traducción que del *Liber de modo bene vivendi ad sororem* hizo el valenciano Antoni Canals, atribuyéndolo a Bernardo de Claravall (Peris 1997), encontramos dos veces la forma *preguacio*:

- (7) a. Deus corregeix los elets e aquells que els veu errar per lur **preguació**  
b. Fou plagat Llatzer pobre per sa **preguacio**.

lo cual justifica la entrada correspondiente en el DCVB (s.v. *pregació*):

- (8) **PREGACIÓ**. f. ant. Pregària. *Corregeix los elets e aquells que els veu errar per lur preguació*. Canals Carta, c. 64.  
ETIM.: pres del llatí **precatiōne**, mat. sign.

Y, a través de este, el fantasma reaparece en el DECLC de Coromines, que lo presenta en la lista de derivados de *pregar*. Y sin embargo, las frases de Canals, y principalmente la segunda, solo adquieren sentido si desarrollamos *por* en lugar de *pre* y leemos *porgació*, so pena de atribuir una incoherencia maligna a la divinidad<sup>4</sup>.

En otros casos, el error es debido a confusión de letras, como el *baschea* que engañó a Carpentier<sup>5</sup> cuando en la edición de 1766 lo añadió inútilmente al Du Cange, donde, naturalmente, la palabra correcta, *baselica*, ya constaba:

<sup>4</sup> Quizás por ello, Moll omitió prudentemente en su cita el sujeto de la frase.

<sup>5</sup> Por confusión de <eli> con <ch>.

- (9) **BASCHEA**, Basilica, ecclesia. Charta ann. 845. in Append. ad Marcam Hispan. col. 781: *Et in pago Redense, in locum ubi dicitur Saltum, vindo in ibidem villa Donacanum cum ipsa Baschea, qui ibidem fundata est in honore S. Felicis. Vide Basilica.*

O las confusiones de las grafías de *n/u/v*, de *ni/m*, de *in/vi*, etc., con lecturas como las que comentábamos acerca de los indios, que han generado fantasmas lexicográficos, como el *balanste* (< BALAUSTIUM) que engañó a Wartburg (*FEW* 21, 89a), o como los citados a propósito de la transcripción del texto de Canals (Peris 1997), cuando *munificar* es relacionado con la familia de *munificiència* y *munificent* en el *DCVB* (s.v. *munificar*) a partir, precisamente de estas documentaciones engañosas, donde solo la lectura *vivificar* haría transparente el sentido.

- (10) **MUNIFICAR**. v. tr. Donar muníficament; cast. *prodigar*. *Diu Tholomeu en lo pròlech del Almagest que nuyll temps mor aquell qui ha munificada la sciència*. Canals Carta, pròl. *La caritat que cuyl o aiusta e munifica tots los membres de la Sgleya*, Canals Carta, c. 13.

Etim.: pres del llatí **munificare**, mat. sign.

O la lectura como <a> del enlace <ci>, que propicia la aparición de términos misteriosos como *absina* y *falaera* que, afortunadamente, despertaron la duda en nuestro lexicógrafo: acertó al considerar un error de lectura el primero, pero no resolvió el caso en el segundo donde, sin embargo, la exigencia de *falcier*, (grafía antigua por *falsier*, ‘falsario’) era mucho más clara.

- (11) **Absina**. f. ant. (tal volta cal llegir *absinci*?): *E la mell ensems ab la manna son al gust absina e fel amargós*. Canals Carta, p. 416. (*DCVB*: s.v. *absina*).

- (12) **FALAER, -ERA**. adj. ant. per *falaguer*? *Les persones falaeres e mentideres fan que encare que diguen veritat no sien creguts*. Canals Carta, c. 32. (*DCVB*: s.v. *falaer, -era*).

No quisiera, sin embargo, que estos ejemplos se entendieran como una crítica despiadada de aquellos a quienes consideramos nuestros maestros. Como decía Coromines, todos hemos pecado muchas veces, quizás demasiadas<sup>6</sup>. Pero nunca estarán de más los esfuerzos para evitarlo. Así, en el proceso de elaboración del *DiCCA-XV* (GHCL y Lleal 2013) algunos de estos temibles fantasmas asomaron en más de una ocasión:

a) En la transcripción del *Arte cisoria* teníamos la forma *çidierueda* que se dejó prácticamente igual en el proceso de lematización automática (con solo el paso de la *cedilla* a *ce*):

- (13) a. *salan en dos toçinos enteros sola mente aparten del espinazo al luengo e las çidieruedas de-las costillas et los lomjillos e sus pies e su menudo onde (DiCCA-XV: Arte Cisoria-044v);*

<sup>6</sup> Tantos, que un equipo de la Universidad de Nancy, del grupo del Atilf / CNRS, ha podido desarrollar una base de datos con las palabras fantasma presentes en el diccionario de Godefroy y en el *FEW* (puede consultarse en <http://www.atilf.fr/MotsFantomes>).

b. tajar de-la pieça del toçino el cuero este fazia el cortador sus **çidieruedas** faze-se vn tajo en-la parte de fuera de tajadas anchas fasta (*DiCCA-XV: Arte Cisoria-045r*).

Posteriormente, cuando ya no trabajábamos con manuscritos sino con formas lematizadas, el elemento *rueda* del significante nos despistó. Solo cuando intentamos buscarle un étimo caímos en la cuenta de que no nos enfrentábamos a unas *ruedas*, sino que se trataba de las *cediérvedas*.

b) Teníamos también unas *anguxidades* que solo acertábamos a relacionar con el catalán *angoixa*, lo cual no acababa de cuadrar con el contenido a no ser que forzáramos mucho la situación:

(14) en-la carne fistulando la. por todo postemaçiones e finchaduras. postillas. sanies e **anguxidades** en el son falladas. dolor de tripas costipaçon de vientre passion de estomago perdimiento (*DiCCA-XV: Tratado de la Lepra-133r*).

Pero cuando volvimos al manuscrito y vimos la posibilidad de leer una <U> en lugar de una <N>, dimos con la forma y el sentido correctos: se trataba de las *auguxidades*, claramente relacionado con el catalán *augositat*, alteración de *aguositat*.

Ciertamente, todos pecamos. Pero para evitar desastres mayores, hay que revisar y revisar. Y volver siempre al manuscrito. Sin esperar a que alguno de estos fantasmas «*fulgore luminis noctis tenebras superet*» y cobre vida real, como les ocurrió al *cenit* alfonsí y al *cartílagu epiglotal*.<sup>7</sup>

## 5. ERRORES DE LOS CARTÓGRAFOS

Un nuevo grupo de errores se lo debemos a los cartógrafos, que a veces actúan como lo haría un hablante, cuando, en cambio, están trabajando como analistas de una variedad que no siempre coincide con la suya propia. Y, como consecuencia de ello, pueden fijar unas formas que nada tienen que ver ni con la realidad ni con la lengua.

a) Por desconocimiento del entorno: este sería el caso de algunos falsos topónimos, como los ribagorzanos *Las Antigas* en lugar de *Las Santigas*, aplicado no a una zona con construcciones antiguas, sino a un zarzal (Terrado 2009), que es lo que significa *santiga* (<SENTICEM). O el desconcertante *Racó de les Cedres* (*sic*) donde no ha habido nunca cedros (y menos aún en femenino) sino hiedras, que es como se debería haber cartografiado: *Racó de les Hedres*. En las misma zona donde *Les Alteres*, zona de cerros y peñascos, son

<sup>7</sup> *Cenit*, del árabe *samt ar-ra's*, 'la zona de la cabeza', que habría dado *zemt* (*arras*), y de ahí por error de lectura, *zenit* (*DCECH*, s. v. *cenit*).

El *cartílagu epiglotal* o *laríngeo* es el nombre con que durante la Edad Media solía denominarse el apéndice xifoides, cartílago con que termina el esternón, que nada tiene que ver con la laringe. En el *Canon* de Avicena es designado con la palabra árabe *aljanÿyar*, 'daga ancha', denominación metafórica por la forma del cartílago. Pero Gerardo de Cremona, el traductor de Avicena al latín, quizás atribuyéndolo a un error del copista, lo interpretó a partir de la palabra *alhanÿyar*, 'laringe', por considerarla más acorde con el entorno médico. Y así, el cartílago en forma de daga se convirtió en *epiglotal* o *laríngeo* (Gutiérrez Rodilla 2006).

transformadas en *Las Salteras* por los cartógrafos, aunque no haya ni un solo salto (o soto) en los alrededores<sup>8</sup>.

b) Por desconocimiento (o menosprecio) de la lengua en zonas bilingües, particularmente en lo que Enrique Gargallo (2013) ha denominado *lengua de frontera*. Entonces, *Les Feixanes*<sup>9</sup> (del catalán *feixa*, ‘faja de terreno abrupto, risco’) pueden llegar a convertirse en *Los Faisanes*, denominación tan absurda como el nombre que pretenden dar a la lengua de sus hablantes; de la misma manera que el *Val de Cueyus* (del portugués *coelho*, ‘conejo’) puede convertirse en *Valdecuellos*. Actitud denunciada en el poema *Solulóquiu*, escrito en portugués oliventino por el escritor extremeño Servando Rodríguez Franco, de San Jorge de Alor:

U que forom *As Arêas*  
agora sãu *Las Áreas*:  
terras muntu granulosas  
derom ceim metrus quadradus

U que foi *U Ariêru*  
é oji u *El Arriero*.  
Foi solu de munta arêa  
queim conduz bestas de carga.

U que foi *Vila Veya*  
oji diz-si *Villa Bella*:  
foi lugar de fartus anus  
i gañó despois beleza.

U que foi *Val de Cueyus*  
oj'ê poreim *Valdecuellos*:  
foi prulíferu ruedor  
reduzidu oj'a pescoçu.

U que foi *A Assêssêra*  
Oji diz-se *La Sesera*:  
Lá foi solu de salguêrus  
u qu'em miolêra acabó.

[...]

Nomis da miña boa terra  
tudu é i ja nã é.

Como consecuencia de ello, cualquier estudio onomástico deberá analizar detenidamente la documentación antigua a fin de evitar desastres de este tipo.

<sup>8</sup> En algunos de estos casos, es probable que la fonética propia del área ribagorzana, con realización de la sibilante final como sorda en los enlaces con la vocal de la palabra siguiente, igual que en castellano, y no como sonora según la norma del catalán, haya favorecido el error.

<sup>9</sup> Topónimo menor del término de Finestres, junto al embalse de Canelles, en Huesca.



## 6. ERRORES DE LOS FECHADORES

En ocasiones, los errores posibles se relacionan con el proceso de clasificar y fechar un corpus documental. El investigador apresurado puede entonces llegar a conclusiones disparatadas. Generalmente, porque en algunos corpus informatizados no se discrimina entre el texto propiamente dicho y lo que pertenece a las regestas, introducciones o estudios críticos posteriores.

Así, alguien podría creer, atendiendo solo al gráfico, que *administración* aparece ya en el siglo XIII. Claro que su ausencia en el periodo inmediatamente posterior debería ser ya significativa.

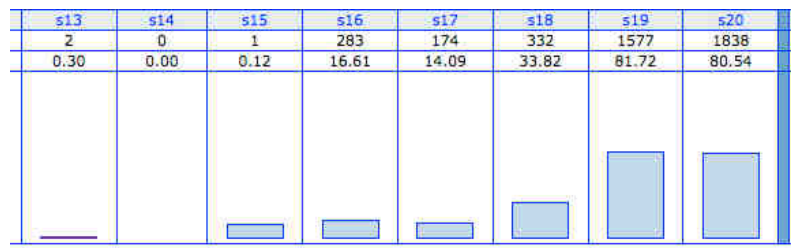


Figura 1: Distribución de *administración* en el *Corpus del español*.

Pero, a pesar de que el título de la obra (*Memoria sobre el Fuero de Madrid del año 1202*) engañó a los fechadores del corpus, el texto en que se encuentran las dos ocurrencias documentadas solo puede engañar a alguien que no esté muy familiarizado con la lengua medieval (aunque desconozca que se trata de un texto de Antonio de Cavanilles, autor del siglo XVIII):

- (15) a. para conocer los usos y costumbres, la **administración** y policía de este pueblo...
- b. Cuatro eran los alcaldes encargados de la **administración** de justicia, cuando...

Y algo parecido ocurre si intentamos rastrear el origen del artículo indefinido, sorprendentemente frecuente en los siglos XI y XII si nos limitamos a contar los ejemplos, sin leerlos, en unos listados en que conviven formas medievales con las regestas del transcriptor moderno:

```

lanzo Ezonis la cuarta parte de Ezar y unas casas situadas en Sangarrén y Huesca. Anónimo I ** 1099
Leire y los vecinos de Equiroz sobre unas tierras que Sancho el Mayor donó a san Salvador ** 1099
103 L Pedro I concede a don Galindo unas casas y heredades, sitas en Huesca. Anónimo In ** 1099
103 L Pedro I dona a Fortún Garcés unas casas que Mahomet Iben Gazer había poseído en Hu ** 1098
o I dona a su mayordomo Sancho Garcés unas casas en Huesca y una viña. Anónimo In nomine D ** 1098
la iglesia de Santiago de Compostela unas casas en Huesca. Anónimo In Dei nomine et eius ** 1098
103 L Pedro I concede al abad Jimeno unas casas sitas en Huesca, unas viñas y un huerto. A ** 1098
ad Jimeno unas casas sitas en Huesca, unas viñas y un huerto. Anónimo In nomine Domini nos ** 1098
ro I da a Lope Fortuñones, zalmedina, unas casas y campos en Huesca. Anónimo Ego Petrus Sa ** 1098
ncede al obispo don Pedro de Pamplona unas casas en Huesca por la salvación de su alma y la ** 1097
I otorga al monasterio de Montearagón unas tierras en Almudévar, una torre situada en Frumi ** 1096
sibi data in suo bapismo; et in Iaka unas kasas et illos valneos et unoquoque anno mille s ** 1086
ula cuncta. Similiter autem dono tibi unas casas, cum earum hereditate in Uno Castello, qua ** 1086
reina doña Berta dona a Iñigo Banzones unas casas, con sus heredades, en Sangarrén. Anónimo ** 1105
ire y el senior Fortún Iñiguez, sobre unas viñas sitas en Lerda, tras la intervención de Pe ** 1094
13 L Pedro I da a su tallador Burfange unas casas situadas en Bolea. Anónimo In Dei nomine ** 1104
n Victorián la iglesia de San Pedro y unas casas. Anónimo In nomine Domini. Hec est carta v ** 1094

```

Figura 2: *Unas* en documentación del siglo XI del *CORDE*.

O en aquellos casos en que el autor del corpus parte de una copia o de una edición posterior, pero le atribuye la fecha del original. Así, por ejemplo, cuando se introduce la edición del texto de las *Siete Partidas* glosada por Alonso Díaz de Montalvo y editada en Sevilla en 1491, pero se le atribuye la fecha de redacción de las *Partidas*. De donde se podría suponer (a pesar de la presencia de una alusión al rey Juan II en la última frase) que un término como *adición* se documenta ya en el siglo XIII (con 49 ocurrencias):

SECCIÓN	s13	s14	s15
OCURENCIAS	49	0	2
POR MILL	7.30	0.00	0.25

e ser ençerrada en vn monesterio que faga penitencia por toda su vida. // **Adición.** / La ley. xxj. libro. j. titulo. iij. delas ordenanças  
 > quela parte quel alguazil omerino deua auer que sea para los dichos muros // **Adición.** / En la ley. xxiiij. ley del dicho titulo fecha poreal rey don  
 azer cosa en que caya en irregularidad & por quello ay an deprender. // **Adición.** / En la ley. vj. libro. 77 titulo. xvij. del fuero  
 comunal del rey & de toda la tierra & de desulamiento de daño // **Adición.** / En la ley. xvij. titulo. iij. libro. j. delos ordenamientos  
 biuen con los legos que fagan el fuero & las costumbres dellos. // **Adición.** / Pon aqui las leyes. xiiij. xv. titulo. iij. libro primero ordenamientos.  
 ombres otrosi no deue ninguno posar en las casas delos clerigos sin plazer dellos // **Adición.** / En los ordenamientos reales en la ley primera titulo terçero dize  
 is o les apremiare afazer: otra seruidumbre contra voluntad delos perlados diogesanos. // **Adición.** / En la. viij. ley del libro. ij. titulo. xxj. delos  
 a el vendedor de fazer sana la cosa a aquel quela comprase del. // **Adición.** / Uey el titulo. xix. libro. iij. del fuero y  
 drar & sanar aquella cosa ante quien faze la demanda el lego. // **Adición.** / El rey don Juhan segundo en valladolid ordeno que los clerigos capellanes del rey

Figura 3: *Adición* atribuido al siglo XIII en el *Corpus del español*.

Algo parecido ocurre si el investigador intenta documentar un fenómeno fonético-morfológico como es el de la evolución de las desinencias de segunda persona del plural, y se encuentra con que, contra todo pronóstico, está ampliamente documentado en el siglo XIII. Un análisis más detenido le muestra que todas las ocurrencias pertenecen al mismo documento: *Los Siete sabios de Roma*, editado en 1510, traducción castellana de la versión latina previa, a su vez adaptación del *Sendeban o Libro de los engaños e los assayamientos de las mugeres*, este sí del siglo XIII.

matar con mis manos quando me veo tan desdichada! Dixo el Emperador: -No **diçáis** esso, mas ave una poca de paciencia que presto avrá y estava ende enojado. E dixo el cavallero: -Señor, si no os **enojáis** tomaremos un poco de más plazer. Respondió ella: -Fágamosle el plazer mi tierra e ruégovos que acompañéis a mi muger hasta la nave y que la **queráis** informar cómo ha de amar a mí sobre todos los hombres y todo e veréis si en este poco tiempo avrá fin e si vuestro hijo no **perdonáis** e le matáis por las palabras de vuestra muger, peor os acacera si en este poco tiempo avrá fin e si vuestro hijo no **perdonáis** e le **matáis** por las palabras de vuestra muger; peor os acacera que aquel ca dezir algunos secretos míos, por tanto os ruego por un solo Dios que me **abráis**. Entonce alla abrió y Él entró e díxole: -Señora, yo demand cuerpo del ladrón que estava ahorcado; por ende os pido por merced que me **queráis** aconsejar. Dixo ella: -Yo he por cierto muy gran compaña con Él a escusarse de lo que le difamavan. Dixo el Emperador: -Vosotros **echáis** esta fama. Pluguiesse a Dios que yo oyesse hablar a mi hijo la dulcedumbre de las mugeres porque remediéis al que avéis mortalmente herido, porque no **seáis** causa de su muerte. E dixo la Infanta: yo tengo un amigo empongoñado e leproso e si vosotros vos ofrecéis de sanallo no **haçáis** estima de oro ni de plata ni de bienes temporales Alteza e por tanto vos rogamos que vos plega de nos dexar hazer e lo **consintáis** por lo que a nuestra honra toca. Dixo entonce el rey: -¿N de la vida e yo os diré la verdad. Respondió el rey: -No **temáis** cosa alguna. Dixo el cavallero: -Señor, nosotros ovimos un hijo muy sabio llenos y cayeron en el suelo y Él levantólos con mucho amor diziendo: -No **temáis**, mas alegraos e tomad plazer que por esto ningún mal a e no lo hizieron. Yo vos consejo que mientras estéis en vuestro poder y **estáis** sano que le destruyáis porque no vos eche la gente maldición

Figura 4: *Los Siete sabios de Roma*, atribuido al siglo XIII en el *Corpus del español*.

También aquí, solo el investigador apresurado puede caer en la trampa, porque la presencia de *vosotros* (líneas 8 y 10), *nosotros* (l. 12), *os* (l. 2, 4, 5, 6, 7, 12), *alegraos* (l. 13), y (conjunción, l. 2), la asimilación *sanallo* (l. 10), *hablar* (l. 8), *hazer, hijo...* podrían haberle dado la clave cronológica.

## 7. ERRORES DE LOS LEMATIZADORES

Otra fuente de fantasmas lexicográficos parte del proceso de lematización. Así, aunque según las reglas morfológicas del castellano las formas verbales terminadas en *-ece*, *-esca*, *-escen* y similares se lematizan automáticamente en el corpus del *DiCCA-XV* como pertenecientes a verbos incoativos del tipo *amanecer*, en la revisión posterior hemos tenido en cuenta que en textos del área aragonesa a menudo estos morfemas actúan de acuerdo con el paradigma verbal del catalán, de manera que en determinados verbos de la tercera conjugación el incremento incoativo se asocia a los morfemas flexivos de tiempo, persona y número y no al morfema radical. Por ello, el desconocimiento de la gramática del texto puede llevar a la generación de formas aberrantes de infinitivo, del tipo:

- (16) \**atribuecer* (en lugar de *atribuir*, a pesar de la presencia de *atribuece*)
- \**complecer* (en lugar de *complir*, a pesar de la presencia de *complezca*)
- \**constituecer* (en lugar de *constituir*, a pesar de la presencia de *constitueixco*)
- \**contribuecer* (en lugar de *contribuir*, a pesar de la presencia de *contribuezcan*)
- \**diferecer* (en lugar de *diferir*, a pesar de la presencia de *diferece*)
- \**digerecer* (en lugar de *digerir*, a pesar de la presencia de *digerescas*)
- \**instituecer* (en lugar de *instituir*, a pesar de la presencia de *institueixco*)
- \**punecer* (en lugar de *punir*, a pesar de la presencia de *punece*)
- \**revertecer* (en lugar de *revertir*, a pesar de la presencia de *revertesca*)
- \**seguecer* (en lugar de *seguir*, a pesar de la presencia de *seguesca, seguescen*)
- \**transferecer* (en lugar de *transferir*, a pesar de la presencia de *transferesce*)

Habrà que ver, en cada caso, cómo se realizan las formas de primera y segunda persona del plural, (\**atribuecemos / atribuimos*) o los participios correspondientes (\**constituecido / constituido*), frente a *ablandecemos* o *comparecido*, que sí suponen un radical verbal con incremento incoativo, para evitar que muchos de estos falsos infinitivos incoativos se paseen por nuestros vocabularios medievales.

## 8. ERRORES DE LOS INVESTIGADORES

Una última fuente de errores se relaciona con algunos de los rasgos avanzados en apartados anteriores y parte de la actitud precipitada de algunos usuarios de los corpus que, sin entender el sentido de la filología, están más interesados a veces en la obtención de datos estadísticos que en el análisis crítico de los documentos.

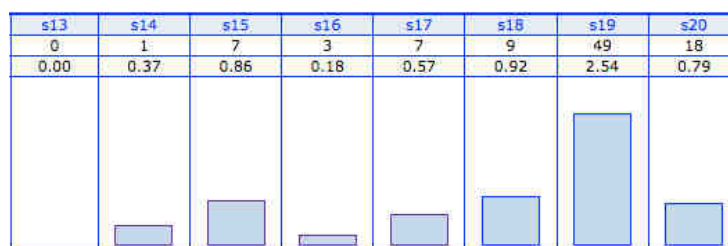
Así, uno puede encontrar en un estudio léxico la utilización indistinta de datos procedentes de usos metalingüísticos junto con datos de uso efectivo del término, sin

discriminar lo que, en el texto, pertenece a otro nivel. Este sería el caso de la documentación de *aroma* en el siglo xv (en la traducción del *Tratado de cirugía*, de Guido de Cauliaco, y en el *Vocabulario* de Antonio de Nebrija):

```
bra. media. del olibano. estoraque. vsnee. calami aroma. espice. absinen. anna .onza.media. confecional ** 1493
lado. insignis .e. egregius Especia de especiero. aroma .atis. species Especiero de especias. aromatopo ** 1495
Adonis, y que amó a Venus, porque este género de aroma es muy caliente; y porque el puerco mató a Adon ** 1580
gualmente flagrantes y preciosos, cuyo fruto, que aroma eterno exhala, más tiene de remedio que de gala ** 1597
```

Figura 5: *Aroma* en el *CORDE*.

o de *régimen* en los siglos xiv y xv, siempre usado como forma latina (en el *Libro de los estados*, de Don Juan Manuel (1327-32) y en el *Universal vocabulario*, de Alonso de Palencia (1490)):



sennor general que deue auer mandamiento sobre todos & **regimen** en latin quiere dezir cosa que deue seer (1327-32) moderadores dicuntur: sed quibus conceditur cuiusuis rei **regimen** quemadmodum veteres voluerunt dici posse. (1490) Regulum occidisset a Carthaginensibus captus est. // Regumen. & **regimen** idem est scilicet regendi modus. (1490) Regulo fue despues preso delos cartagineses. // Regumen & **regimen** lo mesmo significan. es a saber manera de (1490)

Figura 6. *Régimen* en el *Corpus del español*.

Es evidente que si lleváramos al extremo este tipo de actuación, como señalaba Sánchez Lancis a propósito del uso irreflexivo de los corpus (Sanchez Lancis 2009: 169), a partir de la inclusión de referencias de las gramáticas clásicas en un corpus digitalizado, podríamos postular la pervivencia de la terminación verbal {-ades} en palabras paroxítonas todavía en el siglo xvii, y en palabras proparoxítonas en el xviii, o de los participios en {-udo} incluso en el siglo xix, sin atender a la marca como «uso antiguo» con que se presentan:

- (17) La segunda persona plural de los dos presentes i futuro indicativo la usavan los antiguos estendida en *ades, edes, ides: amades, amedes, temedes, temades, consumides, consumades*, como se usa en los imperfetos *amavades, amarades, temiades, consumiades*. Pero de muchos años aca se cortan i contraen por sincopa en el uso comun de todos. (*CORDE*, Gonzalo Correas, *Arte de la lengua española castellana*, 1625).
- (18) Primera conjugacion. Uso Antiguo. Uso Moderno. *amades. amais. amábades. amabais. amástedes. amasteis. amarédes. amaréis. amedes. ameis. amárades. amarais. amariades. amariais. amásedes. amaseis. amáredes. amáreis*. (*Corpus del español*, Real Academia Española, *Gramática de la lengua castellana*, 1771).
- (19) He omitido en el Catálogo que precede los participios irregulares, como *adujo (aducido), rendudo (rendido), sabudo (sabido), tovido (tenido)*, etc. (*CORDE*, *Del*

*arcaísmo y el neologismo*. P.F. Monlau, *¿Cuándo se debe considerar fijada una lengua?*, 1863).

Y, finalmente, frente al respeto crítico de las enseñanzas de quienes nos han precedido, en ocasiones tendemos a aceptar a ciegas las soluciones que nos proponen, llevados por una excesiva confianza en el prestigio de la letra escrita, en una actitud parecida a la que indujo a Friedrich Diez a aceptar sin objeciones el étimo de *requin*<sup>10</sup> propuesto por el *Dictionnaire de Trévoux*, y aceptado a continuación por la Académie Française en la quinta edición de su *Dictionnaire* (aunque eliminado a partir de la sexta edición):

(20) REQUIEM, OU REQUIN. (Car bien des gens écrivent ainfi.) *Chien de mer*, f.m. C'est un gros poiffon de mèr qui dévore les hommes, qui est ainfi nommé, parce que quand on en est mordu il n'y a rien autre chose à faire qu'à chanter le *requiem*. (*Dictionnaire universel françois et latin*. Trévoux, 1704).

(21) REQUIN. subst. masc. Gros poisson de mer très-vorace, qu'on appelle autrement, *Chien de mer*. Les Matelots Normands, pour désigner combien il est redoutable, l'appelèrent *Requiem*, qu'ils prononçoient, par corruption, *Requin*. Le mot est venu d'eux. (*Dictionnaire de l'Académie Française*, 5ª edición, París, 1798).

(22) REQUIN fr. *eine art haifische, auch chien de mer, wegen seiner gefährlichkeit von den normannischen matrosen requiem (scelenmesse) genannt, entsteht in requin (Acad. franç.)*. *Das Dict. de Trévoux schreibt requiem*. (F. Diez: *Etymologisches Wörterbuch der romanischen Sprachen*, 1887)<sup>11</sup>

en una de esas tan peligrosas (y frecuentes) cadenas de transmisión lexicográfica.

## 9. FINAL

Por todo ello, aunque no nos cansaremos de agradecer la publicación de amplios corpus documentales que facilitan nuestra labor investigadora, este agradecimiento ha de ir acompañado de un profundo sentido crítico que nos permita evitar la propagación de incorrecciones que empañen la estimulante tarea de desentrañar los secretos que aún nos reservan los documentos antiguos.

Porque, efectivamente, y a pesar de los numerosos errores, propios y ajenos, que nos pueden desviar, los textos, bien estudiados, nos reservan sorpresas insospechadas. Porque son como los buenos amigos, a los que no se puede traicionar. Y, a cambio, nos proporcionan ese placer íntimo de que nos hablaba hace unos años el profesor Joan Veny, después de haber resuelto un problema textual (Veny 1979): «*Tinc la sensació d'haver rescatat una àmfora mil·lenària que el fang i les correnties dels segles havien soterrat al fons del mar*». Y que ha repetido casi treinta años después (Veny 2007): «*l la*

<sup>10</sup> De origen incierto, probablemente del normando *quin* (< CANEM), con un prefijo intensificador, según el *Dictionnaire étymologique de la langue française* (Bloch y Wartburg 2002).

<sup>11</sup> Que traduce casi literalmente la entrada de l'Académie: «*Una especie de tiburón, o chien de mer, llamado requiem (misa de difuntos) por los marineros normandos debido a su peligrosidad, desfigurado en requin (Académie Française). El Dictionnaire de Trévoux escribe requiem*».



*satisfacció íntima, el voluptuós acontentament d'una descoberta etimològica és com trobar un tros de paradís».*

Con un trozo minúsculo de paraíso por parte de cada uno de los investigadores podríamos obtener una buena muestra de lo que puede llegar a ser la felicidad. Al menos, para un filólogo.

#### REFERENCIAS BIBLIOGRÁFICAS

- BALDINGER, Kurt (1965): «La pesadilla de los etimólogos», *Revista de Filología Española*, XLVIII, pp. 95-104.
- BALDINGER, Kurt (1986): «Etimología popular y onomástica», *Lexis*, X, 1, pp. 1-24.
- BALDINGER, Kurt (1987): «Esplendor y miseria de la filología», en Manuel Ariza, Antonio Salvador y Antonio Viudas (eds.), *Actas del I Congreso Internacional de Historia de la Lengua Española*, vol I. Madrid: Arco/Libros, pp. 19-44.
- BALTAR VELOSO, Ramón (1986): «SVBERA: una palabra latina que nunca existió», *Compostellanum*, 31, pp. 271-273.
- BARCELÓ, Carme (1991): «Penàliga, un topònim àrab», *A sol post*, pp. 3, 39-44.
- BASTARDAS, Joan (1995): «Mots fantasmes en el llatí medieval de Catalunya», en *La llengua catalana mil anys enrere*. Barcelona: Curial, pp. 179-191.
- BLOCH, Oscar y Walter von WARTBURG (2002): *Dictionnaire étymologique de la langue française*. París: PUF.
- BRAMON, Dolors (2013): «Homophonic Transformation of Toponyms: some Exemples from Andalusi Arabic and a New Instance from Pla de l'Estany», *Summa*, 1, pp. 1-12.
- CASTILLO LLUCH, Mónica (2006): «La impostura lingüística: intervención de copistas, editores y gramáticos en los textos medievales», *Cahiers de linguistique hispanique médiévale*, 29, pp. 497-508.
- CdE = DAVIES, Mark (2002): *Corpus del español*. <http://www.corpusdelespanol.org>.
- COLÓN, Germà (1969): «Valor del testimonio aislado en lexicología», *Travaux de Linguistique et de Littérature*, VII, 1, pp. 161-168.
- CORDE = REAL ACADEMIA ESPAÑOLA, *Corpus Diacrónico del español*. <http://corpus.rae.es/cordenet.html>.
- COROMINES, Joan (1972): *Tópica hespérica*, 2 vols. Madrid: Gredos.
- DCVB = ALCOVER, Antoni M. y Francesc de B. MOLL (1926-35; 1949-62): *Diccionari català-valencià-balear*. Palma: Moll, 1968<sup>2</sup>.
- DECLC = COROMINES, Joan (1980): *Diccionari etimològic i complementari de la llengua catalana*. Barcelona: Curial.
- DECH = COROMINES, Joan y José Antonio PASCUAL (1980-91): *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos (versión electrónica de 2012).
- DICCA-XV = GHCL y Coloma LLEAL (dirs.) (2013): *Diccionario del castellano del siglo XV en la Corona de Aragón*. <http://stel.ub.edu/diccaxv>.
- DU CANGE, Charles du Fresne, seigneur (1678): *Glossarium mediae et infimae latinitatis*, édition augmentée. Niort: L. Favre, 1883-87. <http://ducange.enc.sorbonne.fr>.
- GARGALLO, José Enrique (2013): *Poesía de frontera: Segona trobada Iberoamericana*. Barcelona, 27 de mayo de 2013.

- GUTIÉRREZ RODILLA, Bertha M. (2006): «La impronta de la traducción sobre el discurso científico: toda una historia», *Puntoycoma*, 106. <http://ec.europa.eu/translation/bulletins/puntoycoma/106> [Consulta: mayo 2013].
- MAIDEN, Martin (2008): «Lexical nonsense and morphological nonsense: on the real importance of ‘folk etymology’ and related phenomena for linguistics», en Thórhallur Eythórsen (ed.), *Gramatical Change and Linguistic Theory*. Estrasburgo: John Benjamins, pp. 307-328.
- MEILLET, Antoine (1893): «Les lois du langage. I», *Revue Internationale de Sociologie*, I, pp. 311-321.
- PERIS, Antoni (1997): «Mots fantasmés en el català medieval», *Llengua i literatura*, 8, pp. 44-56.
- REEVE, Michael D. (1994): «Errori in autografi», en Pablo Chiesa y Lucia Pinelli (eds.), *Gli autografi medievali. Problemi paleografici e filologici. Atti del convegno di studio*. Spoleto: Fondazione Ezio Franceschini / Centro italiano di studi sull'alto medioevo, pp. 37-60.
- SÁNCHEZ LANCIS, Carlos (2009): «Corpus diacrónicos y periodización en español», *Cahiers d'études hispaniques médiévales*, 32, pp. 159-180.
- SÁNCHEZ-PRieto BORJA, Pedro (1996): «Problemas lingüísticos en la edición de textos medievales», *Incipit*, 16, pp. 19-54.
- SEGRE, Cesare (1976): «Critique textuelle, théorie des ensembles et diasystème», *Bulletin de la classe des lettres et des sciences morales et politiques*, 62, pp. 279-292.
- TERRADO, Javier (2009): «Léxico patrimonial y metacedeusis en los nombres de lugar», *Archivo de Filología Aragonesa*, 65, pp. 215-234.
- VENY, Joan (1979): «De la bèl·lua al tauró: supervivents catalans del llatí PISTRIX», *Randa*, 9, pp. 51-62.
- VENY, Joan (1990): «Cap a una tipologia de l'etimologia popular», en *Jornadas de Filología. Homenaje al profesor Francisco Marsá*. Barcelona: Servei de Publicacions Universitat de Barcelona, pp. 137-152.
- VENY, Joan (2000): «De la nineta a l'ànima de l'ull», en *Jornades de la Secció Filològica de l'IEC a Elx*. Barcelona / Elx: IEC / Ajuntament d'ELX, pp. 83-92.
- VENY, Joan (2007): «Coromines, colós de l'etimologia», *Zeitschrift für Katalanistik*, 20, pp. 3-19.