

Desenvolupament d'una eina correctora d'Eles Geminades per a documents PDF

Marc Antoni Malagarriga i Picas

2 de juny de 2006

Resum

Com si d'una reintroducció d'espècies protegides es tractés, el cas de la *Ela Geminada* ens posa en safata un projecte de reconstrucció tipogràfica en ple segle XXI. El que venim a explicar en aquesta ponència és només una petita solució que forma part d'un seguit d'objectius que han de tornar a posar al seu lloc, en termes d'eficiència i correcció, un caràcter de l'escriptura que, si continuem badant, pot córrer el risc de quedar empassat definitivament dins l'anomenada *esquerda digital*.

1. El 2013 farà 100 anys

Des que fa gairebé un segle l'*Institut d'Estudis Catalans* va establir el caràcter de la *Ela Geminada* (LL) com a normatiu per a la representació d'un determinat so de la nostra llengua (les «Normes Ortogràfiques» de la llengua catalana del 1913^[1]), la correcció en la composició d'aquest dígraf s'ha anat degradant de forma inversament proporcional a la flexibilitat de les tecnologies utilitzades per a la seva representació.

Cronològicament la *Ela Geminada* només ha conservat la seva correcció i puresa compositiva^[2] dins els sistemes clàssics de composició amb tipografia de metall: els tipus mòbils de muntatge manual peça a peça^[figura 1] i el sistema basats en la *Linotípia*^[figura 2], que mecanitzava la composició de línies en plom.



Figura 1: diferents detalls d'un tipus mòbil de metall pertanyents a l'antiga *Editorial Barcino* ^[3]

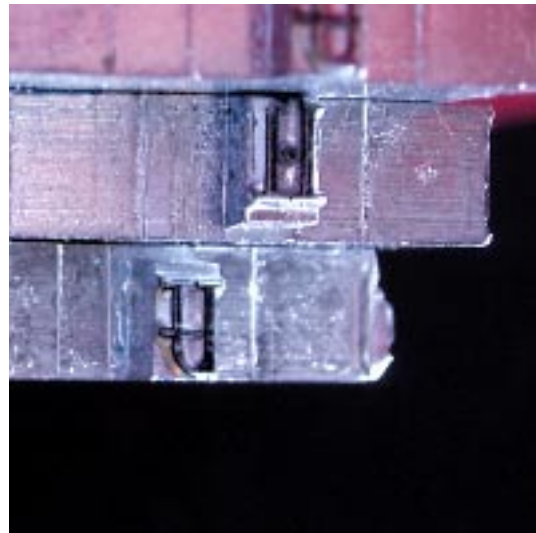


Figura 2: diferents detalls d'un caràcter compost i fos amb plom de la linotípia d'en *Lluís Crespo* [9]

L'evolució tecnològica de la composició de textos evolucionà, a partir dels anys cinquanta, cap als sistemes de *fotocomposició* i aquí es desintegrà per primer cop aquest dígraf d'una manera generalitzada (hores d'ara no hem trobat cap testimoni de màquines o material imprès per mitjà de la *fotocomposició* que utilitzés la *Ela Geminada* com un sol caràcter).

La màquina d'escriure elèctrica ens tornà a recuperar aquest dígraf en la dècada dels vuitanta, encara que d'una forma mutilada (L·) per les limitacions de la tecnologia de les margarides (*Daisywheel*). El mataroní *Joan Majó*, aleshores ministre d'Indústria i Energia de l'Estat espanyol, promulgà el 1985 un reial decret[4] on quedaven homologats per primer cop els teclats de dispositius electrònics i s'obligava als fabricants a incloure determinats caràcters, com és el cas de la *Ela Geminada*. Marques com *Olivetti*, *Panasonic* o *Tryumph-Adler* [figura 3], entre d'altres, fabricaven, i fabriquen encara, màquines d'escriure amb les tecles [L·] i [Ç] per al català i [Ñ] per a l'espanyol dins el mateix teclat.

La ràpida expansió de l'ordinador personal i els sistemes de composició 100% digitals a partir de finals dels anys vuitanta provoquen el tomb tecnològic definitiu, que situa el problema que ens ocupa en l'escenari actual. Es crea, encetada la dècada dels noranta, l'organisme *Unicode*[5], que vetlla i regula en l'àmbit mundial totes les codificacions dels caràcters de l'escriptura.

S'aproven també les primeres normatives europees d'homologació de teclats i novament la *Ela Geminada* queda inexplicablement soterrada i en una situació de surrealisme tecnològic (tret d'uns primers models que va treure *Olivetti*, el caràcter mutilat [L·] no s'ha vist mai més en un teclat d'ordinador).



Figura 3: teclat d'una màquina d'escriure elèctrica, *Tryumph-Adler*, de margarida (per gentilesa d'*Ofimat*)

2. Surrealisme tecnològic per a l'únic caràcter que ens fa diferents

La *Ela Geminada* és l'únic signe que identifica tipogràficament la llengua catalana de la resta d'idiomes d'aquest planeta. Malgrat aquesta singularitat identitària, com podreu veure tot seguit, l'escenari actual d'aquest dígraf és ple de contradiccions:

A. *Unicode* ens té homologat aquest caràcter^[6] en la seva versió mutilada —suposem que gràcies al famós decret de 1985 pensat per a les màquines d'escriure— d'aquesta forma:

- Caràcter [L·] (anomenat internament *Ldot*) amb codi hexadecimal <013F> (decimal 319)
- Caràcter [l·] (anomenat internament *ldot*) amb codi hexadecimal <0140> (decimal 319)

B. Ningú no ha explicat ni a *Unicode* ni als tipògrafs de les principals empreses fabricants de tipografia com cal dissenyar una *Ela Geminada* per tal que compleixi les pautes estructurals que ja complien quan es fonien en plom. Resultat d'això és que el 100% de les *Ldot* i *ldot* que hi ha actualment en la majoria de les tipografies *OpenType* que tenim carregades a l'ordinador estan mal fetes^[figura 4].

C. Tampoc, però, no tenim cap possibilitat raonable d'utilitzar aquest caràcter en un ordinador comercialitzat a Catalunya, perquè la [L·] ni apareix com a tecla directa, ni sota cap mena de combinació. El joc de caràcters que l'acull: *Latin Extended-A*^[7], tampoc no ve configurat per defecte en els principals sistemes operatius. Daltra banda, qui coneix actualment que existeixen *Ldot* i *ldot* ?

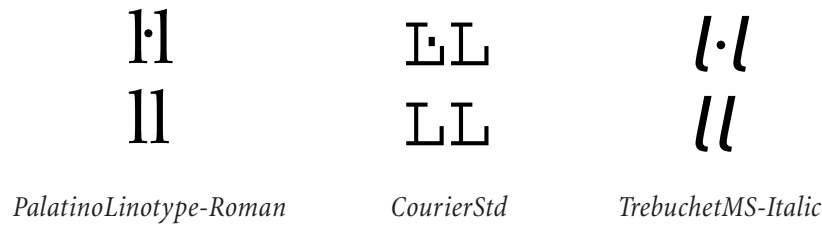


Figura 4: exemples compostos amb els caràcters *ldot* i *Ldot*. Ho podreu comprovar fent saltar el cursor per sobre la lletra (1 salt = 1 caràcter). Compareu l'ample de les dues eles amb el de la *Ela Geminada*.

D. Les actuals normatives europees de configuració de teclats per als estats membres ignoren la *Ela Geminada* com a caràcter a tenir en compte en els diferents aparells (ordinadors, PDA, telèfons mòbils, etc.) que treballin també en la nostra llengua. Ja ningú no es recorda del reial decret de 1985.

E. Actualment només podem compondre raonablement aquest dígraf prement tres tecles: *Ela+Punt Volat+Ela*. Aquest fet ens duu a una situació actual de màxim desgavell, on l'error és pertot arreu:

- Mai no la podrem compondre bé (amples i centralitats incorrectes)[figura 5].
- Encara hi ha molta gent que fa la falta d'ortografia d'escriure-la amb un *Punt Normal* o amb un *Guionet*.
- Dins un document electrònic (PDF, HTML) o a Internet, les paraules escrites amb *Ela Geminada* moltes vegades no les podrem trobar amb el cercador, a causa de les inestabilitats de codificació interna que a voltes es produeixen amb el *Punt Volat*. Això suposant que la paraula que busquem no s'hagi escrit amb un *Guionet* o amb un *Punt Normal*.

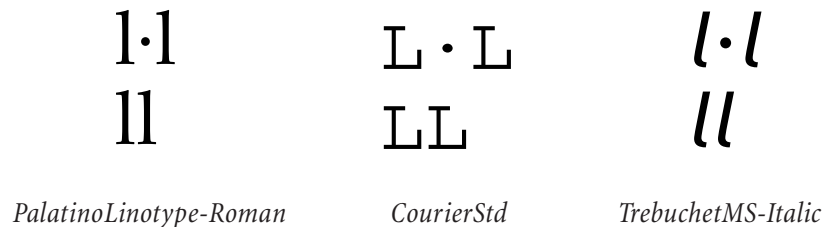


Figura 5: exemples compostos amb els caràcters *l + · + l*. Ho podreu comprovar fent saltar el cursor per sobre la lletra (1 salt = 1 caràcter). Compareu l'ample de les dues eles amb el de la *Ela Geminada*.

3. Si Pompeu Fabra i Joan Coromines aixequessin el cap!

El tretze de juny de mil nou-cents vint-i-tres, *Pompeu Fabra* escrivia la seva *Conversa Filològica Tres-Cents Noranta-U*, i deia: “*Hem notat que molts, en escriure un mot en l doble, acostumen a deixar entre les dues l un espai capaç de contenir una m*” i deia més endavant: “*I els impressors espaien també les dues l continuant a realitzar la doble l amb dues l separades per un punt volat. ¿Quan es decidiran, almenys els que imprimeixen molt en català, a adquirir el signe ll*”. Han passat més de vuitanta anys d'aquests comentaris del mestre Fabra[2] i, tot haver evolucionat del plom al píxel, són plenament vigents[8].

Joan Coromines va morir el 1997 i la seva obra magna de toponímia dels Països Catalans, l'*Onomasticon Cataloniae*, va ser composta per expressa voluntat seva, fins la darrera setmana, amb *Linotípia* (composició mecànica amb plom). A causa de la ultracorrecció d'aquest sistema de composició tipogràfica i sobretot a l'impecable professionalitat dels linotipistes[9], *Coromines* no volia saber res de la composició electrònica actual. Malgrat això, *Curial Edicions Catalanes* —editora de l'*Onomasticon*— va cometre un nyap imperdonable tant en l'àmbit editorial com, sobretot, pel que fa al respecte al mestre: d'una forma intercalada, dins els vuit volums de l'*Onomasticon*, aquesta obra va ser acabada de compondre (un 38% d'aquesta obra) amb eines digitals[10]. Com us podeu imaginar, el resultat d'això salta a la vista[figura 6] i és precisament la *Ela Geminada* un dels elements que delaten aquesta barreja tan poc professional. Compareu, si no, dins del tom VII[figura 7], la paraula *illustre* (composta en linotípia a la pàgina 207, reng 46) amb la paraula *illustre* (composta en digital a la pàgina 441, reng 43).

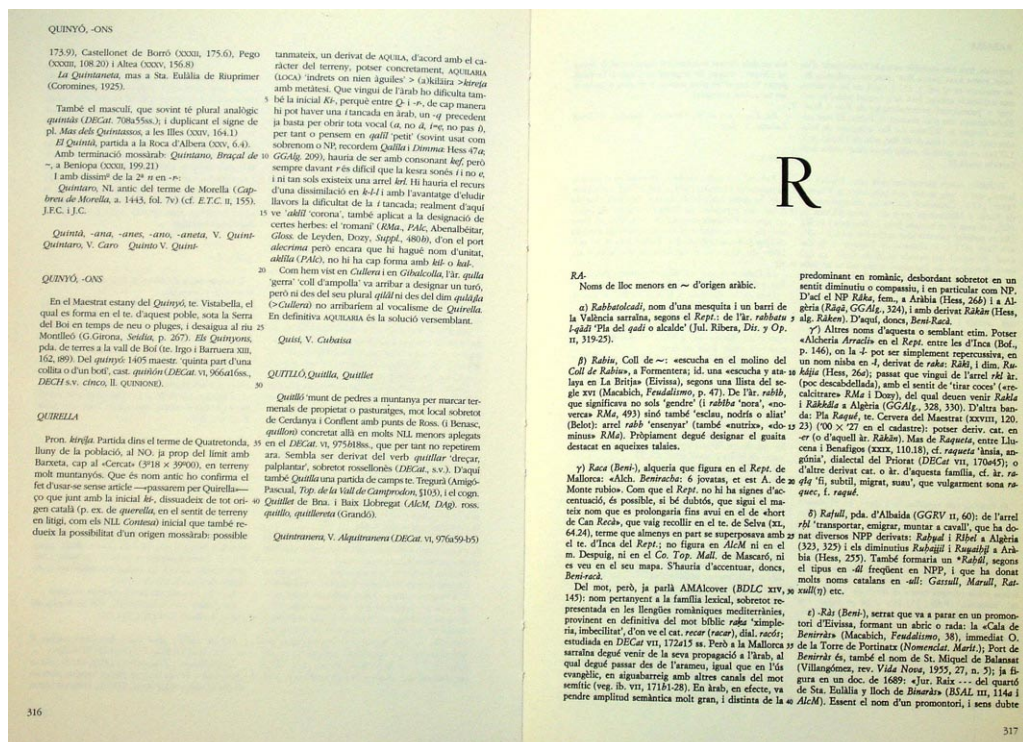


Figura 6: pàgines 316 i 317 del Tom VI, la de l'esquerra composta en digital i la de la dreta en linotípia

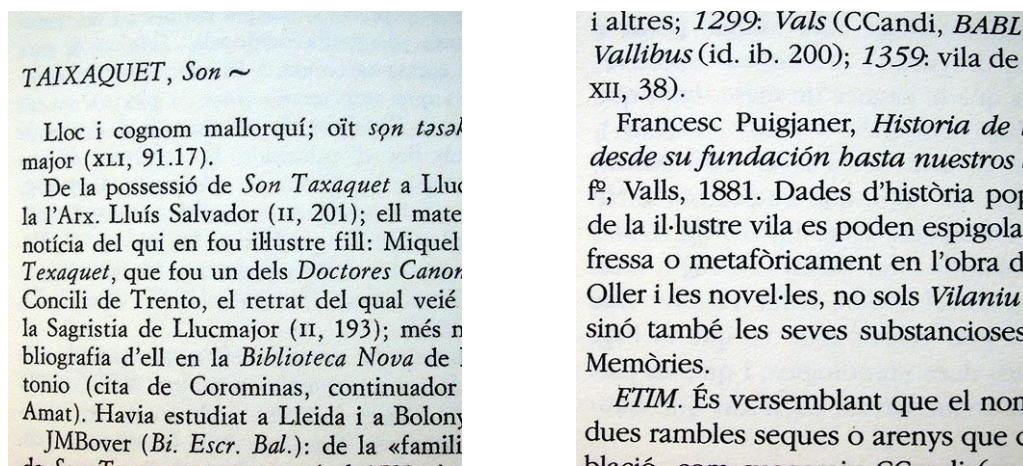


Figura 7: la paraula *illustre* composta en linotípia (a l'esquerra) i en digital (a la dreta)

Una colla de professionals —que treballem la llengua, la tipografia, els sistemes d'informació i edició— amb el suport d'estaments com el *Departament de Disseny i Imatge*^[11] de la *Facultat de Belles Arts* de la *Universitat de Barcelona*, ens hem proposat normalitzar l'ús de la *Ela Geminada* en quatre aspectes fonamentals:

1. Per a una correcta composició.

Només hi ha una manera d'aconseguir que la *Ela Geminada* quedi sempre ben composta: que torni a formar part d'un sol caràcter, una sola peça, una sola tecla.

2. Per a una correcta indexació i cerca.

Mai no es troba allò que es busca quan els caràcters o les seves codificacions no es corresponen amb el que podem prémer des del teclat, o els criteris del cercador (que hauria de poder contemplar excepcions i un cert llinard d'error) no són prou flexibles. El cas de la *Ela Geminada* pateix actualment problemes de tota mena en aquest sentit.

3. Per a un ús directe, eficient i per evitar la falta d'ortografia.

Tenir una tecla directa per la *Ela Geminada* voldrà dir: passar a prémer una tecla en lloc de tres, i això guanya importància quan parlem d'estris petits com PDA o telèfons mòbils. D'altra banda, tallem d'arrel la falta d'ortografia provocada pel punt normal o el guionet.

4. Per divulgar la singularitat d'aquest caràcter únic al món.

Les accions que es duren a terme a *Unicode* i a congressos de tipografia d'abast mundial com, *ATypI*, juntament amb la futura <http://www.LL.cat>, ajudaran a posar la *Ela Geminada* al mapa.

Per assolir la plena normalització que hem exposat, aquest projecte ha de fer realitat tres objectius ordenadament:

En primer lloc

Demandar a *Unicode* la rectificació del caràcter “l.” (*ldot*), codi <0140>, per “ll”, amb el mateix codi, i del caràcter “L.” (*Ldot*), codi <013F>, per “LL”, també amb el mateix codi. L'únic que caldrà variar serà el nom intern de caràcter, que pel fet de ser usat només en català no caldrà pas que s'expressi en anglès.

En segon lloc

Confecció d'un manual d'ús per a tipògrafs i dissenyadors, on es donarien les pautes bàsiques per a construir la *Ela Geminada* sota qualsevol estil, mantenint sempre un acord de mínims tipogràfics obligatoris.

En tercer lloc

Homologació d'un teclat català (usant el joc de caràcters *Latin Extended-A*^[7]) on la *Ela Geminada* aparegui novament com una tecla directa (tal com esdevé en les màquines d'escriure de margarida). Aquest també és un tema polític, en el qual potser caldrà rectificar la normativa europea en aquest àmbit.

Cal esmentar molt especialment la feina feta per la gent de la <http://www.fundació.cat>, gràcies a la seva històrica aportació dels quals la *Ela Geminada* és al mapa dels dominis d'Internet^[11]. Des de fa uns mesos ja podem

registrar noms web amb *ll*. Els protocols d'Internet encara són molt encarcarats i no accepten gaires refinaments ortogràfics; llavors, pel que fa al nostre dígraf, de moment cal compondre'l com ja sabem: *Ela+Punt Volat+Ela*.

4. Una primera mesura palliativa en programari lliure

El desenvolupament d'una eina correctora d'Eles Geminades per a documents PDF forma part només d'una solució parcial per resoldre provisionalment els punts 1 i 2 esmentats en l'apartat 3. En una primera anàlisi aquest programari el preveiem com segueix:

Llenguatge de programació

PostScript Level 3

Intèrpret

Ghostscript, en qualsevol de les seves distribucions (GPL, AFPL) a partir de la versió 8.#

Flux bàsic de treball

El programa PostScript (*PS Applet*), un cop executat dins Ghostscript, donarà d'alta un nucli de redefinicions corresponents a tota l'estructura d'operadors del format PDF. Tot seguit executarà literalment dins el seu propi context el fitxer PDF, on cal reparar les *Eles Geminades*. Fruit d'aquesta execució s'esdevindrà un resultat en tres etapes: anàlisi de les *Eles Geminades* existents al document, reparació cas a cas i, finalment, reescriptura del document PDF modificat.

Radiografia de l'anàlisi de les Eles Geminades

- Filtrarem les combinacions següents en caixa alta i caixa baixa: *l+periodcentered+l* | *l+period+l* | *l+bullet+l* | *l+hyphen+l* | *l+minus+l* | *l+endash+l* | *l+dot+l*
- Si descobrim nous casos no contemplats ampliarem la taula d'excepcions, si es deuen a una lògica raonable. En cas contrari (certes codificacions privades, on és impossible d'endevinar el caràcter sense analitzar la seva forma), ens plantejaríem utilitzar un petit diccionari de paraules que s'escriuen amb *Ela Geminada* com a darrer comprovant (el nombre de mots a consultar és relativament petit).
- Descartaríem de la correcció les *Eles Geminades* amb separació sillàbica a final de línia.
- El format final de la reparació sempre serà el següent: *Ela+Punt Volat+Ela*. Llavors, arribats aquí, se'ns presenten almenys dues possibilitats de treball: utilitzar (sempre que sigui possible) el *Punt Volat* de la *Font* en què és composta la paraula a reparar, o utilitzar el *Punt Volat* de les *Fonts MultipleMaster* que tots els visualitzadors *AdobeReader* duen incorporats. Per tal de donar major compatibilitat al PDF en altres visualitzadors (*Xpdf*, *GhostView*, *MacOSX Preview*, etc.) caldria incrustar aquesta tipografia.

- La reparació sempre anirà en dos sentits: un al nivell compositiu, que afectarà possiblement tota la línia de text on es trobi la paraula amb la Ela Geminada, i un altre al nivell de codificació, per assegurar que l'*Encoding* de la tipografia que compongui la *ll* segueixi una taula estàndard tipus *MacRoman*, *ToUnicode* o *WinAnsi*. És l'única manera d'assegurar-ne la correcta indexació i cerca.
- Al nivell compositiu i dins el context de la *ll* es podran seguir diverses estratègies a l'hora de decidir l'ample de les dues eles, el centrat horitzontal/vertical i la dimensió del *Punt Volat*. Aquests comportaments anirien lligats sempre segons el pes i la categoria estilística de la tipografia utilitzada (p.e: espaiat fix, script, negreta, etc.).

Eines desenvolupades i en producció real en les quals es fonamenta

L'equip de desenvolupament de <http://www.femfum.com> té creats des de fa temps els mòduls de redefinició d'operadors estructurals del format PDF, així com els motors de reescriptura del fitxer de sortida. Compta amb dues eines reals i en producció que es recolzen en aquesta estratègia de treball:

- Un extractor de textos Ascii per a documents PDF (de nom *VeRBiVoR*) que pot treballar selectivament segons criteris (posicions de pàgina, cos, color i estil de lletra, etc.) i que, per exemple, és el que diàriament està utilitzant l'*Institut d'Edicions de la Diputació de Barcelona* en l'extracció d'anuncis del *BOP*.
- Una eina d'anàlisi i separació de colors de fitxers PDF, per a preimpresió (de nom *CaLi2CoPi*), que s'utilitza com a control de qualitat per a grans volums de producció d'originals en factories com *Tetrapack*.

Calendari previst per al desenvolupament

- A finals de setembre hauríem de tenir el nucli PS de l'aplicatiu completament operatiu i hauríem d'estar a punt de començar a fer proves documentals de forma exhaustiva.
- A finals de novembre hauríem de tenir el nucli PS integrat en un mòdul web (PHP o Java) per poder oferir aquest servei *Online*, de forma completament gratuïta.
- Abans no s'acabés l'any s'hauria de presentar en societat aquest servei.

Crida a trobar col·laboració en les àrees que no tenim resoltes

- *Hosting* per a la futura <http://www.LL.cat>
- Integració PHP o Java per a l'ús del motor PS+Ghostscript de forma remota des d'Internet
- Disseny i construcció d'una interfície d'usuari per al servei *Online*
- Lligams amb altres projectes o programaris existents
- Ampliació del seu ús a altres problemàtiques de correcció (per exemple, l'apòstrof)

5. Breu cloenda

Us en recordeu, d'aquella nena, de nom Galla, que no va poder ser inscrita al *Registre Civil* de forma correcta per culpa de les incompatibilitats de la *ll* i un sistema informàtic acabat d'estrenar ?

Queda clar que tenim un problema i que s'haurà d'actuar amb diligència per aconseguir la plena normalització d'aquest dígraf. La creació d'aquesta eina correctora és un primer pas voluntariós en aquest sentit.

Referències

- [1] Mila Segarra | *Història de l'ortografia catalana* Barcelona: Empúries 1985, p. 292-293, 305, 313, 318 i 325.
- [2] Pompeu Fabra | *Converses Filològiques Volum II* (c) Fundació Enciclopèdia Catalana, 1984 per cessió de Carolina i Dolors Fabra (c) EDHASA, 1984 | ISBN:84-350-5111-0 Transcripció electrònica de les converses 323 i 391, on es parla exclusivament de la Ela Geminada... http://femfum.com/PDF/ElaGeminada/PF_CF_323_391.pdf
- [3] Per gentilesa de Jordi de Casacuberta (nebot de Josep M. de Casacuberta, lingüista insigne i fundador de l'Editorial Barcino), vàrem poder localitzar aquestes peces d'Ela Geminada de metall (dels anys 1920/30).
- [4] *Real decret 2707/1985* | Boletín Oficial del Estado, núm. 64, 15.03.1986, p. 9800-9847.
- [5] <http://www.unicode.org>
- [6] <http://www.unicode.org/Public/UNIDATA/NamesList.txt>
- [7] <http://www.unicode.org/charts/PDF/U0100.pdf>
- [8] En aquest document hem tingut la cura de rectificar manualment l'espaiat de totes les Eles Geminades que hi surten.
- [9] Vull donar les gràcies a *Lluís Crespo*, el linotipista de Joan Coromines, que compongué més del 80% de l'*Onomasticon Cataloniae* (ens referim a la part que va ser composta amb linotípia) i que ens ha documentat, amb el seu testimoni i amb proves reals, tots aquests fets i totes les mostres que s'exposen en aquesta ponència.
- [10] Les aproximadament 3.700 pàgines que formen els vuit toms de l'*Onomasticon Cataloniae* presenten aquest lamentable panorama compositiu:
- Tom I
compost 100% en linotípia
 - Tom II
compost 100% en linotípia
 - Tom III
compost 100% en linotípia
 - Tom IV
compost un 85% en linotípia i un 15% en digital
 - Tom V
compost 100% en digital
 - Tom VI
compost un 34% en linotípia i un 66% en digital
 - Tom VII
compost un 79% en linotípia i un 21% en digital
 - Tom VIII
compost 100% en digital
- [11] El *Departament de Disseny i Imatge* de la *Facultat de Belles Arts* de la *Universitat de Barcelona* disposa del fons de punxons i matrius que no estan en producció a la foneria tipogràfica *Bauer-Neufville de Barcelona*, que inclou els fons de la *Fundición Tipográfica de Madrid*, la *Fonderie Typographique Française*, *Lettergieterij* d'Amsterdam, la *ER* del Líban i la foneria *Ludwig&Mayer* d'Alemanya. Els fons compostos per punxons i contrapunxons,

Referències

matrius i catàlegs, configuren una de les col·leccions més importants del món, en paral·lel a l'*Imprimerie Nationale Française*, el Museu *Plantin-Moretus* i el Museu *Bodoni*.

The image shows a page from a Bodoni type case (pòlissa). At the top, it reads 'Cpo. 14 N.º 1074 Bodoni Caja N.º 463 7337 234'. The main part of the page is a grid with columns for letters and printer's marks. The letters listed are: a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, and various ligatures like 'aa', 'bb', 'cc', 'dd', 'ee', 'ff', 'gg', 'hh', 'ii', 'jj', 'kk', 'll', 'mm', 'nn', 'oo', 'pp', 'qq', 'rr', 'ss', 'tt', 'uu', 'vv', 'ww', 'xx', 'yy', 'zz'. The printer's marks include: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, and various symbols like '·', 'ˆ', '˜', '¨', 'ˇ', '˘', '˙', '˚', '˛', '˜', '˘', '˙', '˚', '˛'. The number '11' is written in the right margin next to the 'll' row. At the bottom, it says 'Materia: 182'.

Aquesta imatge correspon a una pòlissa de Bodoni, es a dir, la nota detallada de la quantitat i classe de lletra que un impressor demana a una foneria o que aquesta li serveix. Podeu veure-hi el caràcter *ll* escrit a mà al final de la plana. Aquest és un dels molts llibres de pòlisses que pertanyen al fons d'aquesta universitat.

[12]

<http://www.icann.org>

* A Internet també podreu trobar aquest document a:

http://femfum.com/PDF/ElaGeminada/5JPL_PonenciaElaGeminada.pdf

* La correcció lingüística d'aquesta ponència ha anat a càrrec d'Antoni Romero i Dedéu