

## **Aproximació lexicomètrica a les interferències de base fonològica en els escrits autògrafs de Cristòfor Colom**

Autor: Lluís de Yzaguirre i Maura (Universitat Pompeu Fabra)

Relator: Jordi Marí i Salvador (Centre d'Estudis Colombins)

### 1.- Resum

Aquesta comunicació presentarà el treball efectuat per encàrrec de Discovery Channel (i algun de posterior) al voltant de la possibilitat d'aplicar tècniques actuals de recerca lingüística a la incògnita sobre l'origen de Colom. En el nostre cas, la proposta fou la de mesurar la probabilitat que la llengua materna de Cristòfor Colom fos el català, tasca en què pretenguérem aportar un nou nivell de reflexió: les interferències fonològiques manifestades en les vacil·lacions (orto)gràfiques dels escrits autògrafs del personatge.

### 2.- Hipòtesi fonològica

Partim del fet que no es conserven escrits de Colom en cap de les llengües que es podrien reivindicar com a candidates a ser la seva llengua materna. En cas que algun dia se'n trobin, les tècniques actuals de lingüística forense podrien intentar establir la veracitat d'aquests escrits i la versemblança que hagin estat produïts per un parlant nadiu.

També partim del fet admès generalment que el castellà no és la seva llengua materna i que, doncs, en els seus escrits autògrafs en castellà és versemblant que s'hi manifesti d'alguna manera la seva llengua materna, en forma d'interferències en sentit propi o pel fet de triar solucions genuïnament castellaneres en una freqüència molt superior a l'habitual en castellà.

Diversos estudis anteriors han aportat evidències de coincidències en el pla lèxic entre el castellà usat per Colom i un inventari molt reduït de llengües. No deixa de tenir importància per al nostre plantejament el fet que les llengües en controvèrsia siguin molt poques, com recordarem més endavant.

Doncs, ens vàrem plantejar trobar traces de la seva llengua materna en els seus escrits, però en un pla nou, el de les interferències grafo-fonològiques.

### 3.- Factors concurrents

Abans d'emprendre la recerca que presentem en aquest treball, ens vàrem voler informar sobre els factors concurrents quant a la possible catalanitat de Colom, en el sentit que no ens interessava esmerçar esforços en una causa perduda; dit en altres paraules, si els nostres resultats havien de ser els únics que corroboressin la hipòtesi catalana, potser esperaríem que fossin d'altres amb més dedicació a la qüestió els que fessin el primer pas.

Aquests factors concurrents els vàrem extreure de les converses amb el Sr. Francesc Albardaner del Centre d'Estudis Colombins (que fou l'intermediari de la proposta de recerca per part de Discovery Channel), persona que reuneix el rigor d'una titulació superior científico-tècnica i de la praxi professional dilatada amb una vastíssima erudició històrica i una experiència de recerca en el tema avalada, per exemple, pel fet d'haver estat autoritzat a fer indagacions durant mesos als arxius vaticans.

A més de posar-nos al dia en els resultats obtinguts pels partidaris de la hipòtesi catalana, aquestes converses (luxe intel·lectual) ens varen convèncer que les nostres recerques se sumarien a d'altres fets rellevants aportant-nos un conjunt de dades que es poden agrupar en dos aspectes que contradiuen la hipòtesi genovesa (el seu origen noble i el seu perfil lingüístic) i en els seus motius per a ocultar la seva identitat.

a) el seu origen noble explica:

- les seves noces amb una noble portuguesa
- l'admissió dels seus fills en la Cort
- la seva cal·ligrafia segons el model de la Cancelleria de Barcelona<sup>1</sup>
- el seu coneixement del castellà, dominant en la Cort dels Trastámara<sup>2</sup>

b) del seu perfil lingüístic m'interessa destacar:

- el poc que ha escrit en "italià" és macarrònic
- es comunica amb el seu germà en castellà
- i amb els seus banquers genovesos
- després de 9 anys a Portugal i sense haver trepitjat el regne de Castella, coneix millor el castellà que el portuguès
- en els seus escrits apareix molt lèxic català<sup>3</sup>
- s'inspira en topònims catalans per als seus descobriments<sup>4</sup>

c) els seus motius per a ocultar la seva identitat

El seu interès per ocultar la seva identitat sota un hipotèticament fals origen genovès s'explica per no voler reconèixer els seus anteriors enfrontaments amb el poder reial i al seu torn explica que no usi la seva llengua materna quan escriu al seu propi germà.

1 Ho corroborà el Dr. Gabriel Roura i ho manifestà al programa de Discovery Channel

2 La influència del castellà a les terres catalanes és un fet antic, que comença a adquirir un caràcter parainstitucional al llarg del segle XIV, amb l'adveniment de la nova dinastia castellana dels Trastámara i la instal·lació a la cort de cercles intel·lectuals castellans, amb el prestigi que se'n deriva per a la llengua catalana. Vegeu <http://www.racocatala.com/~cat1714/documents/histcat.pdf>

3 vegeu Caius Parellada (Parellada 1987) i (Parellada 1992)

4 vegeu Nito Verdera (Verdera 2000)

El resum de tots aquests fets concurrents és que en cas de quedar descartada la hipòtesi genovesa (que bàsicament s'autodescarta pel fet que els seus punts més forts han estat desautoritzats pel pas del temps, essent el fet més destacat el descobriment de la relació de persones embarcades en la flota en què suposadament Colom arribà a Portugal, relació en la qual Colom no consta), la hipòtesi més plausible és la catalana, amb independència del que puguin aportar les nostres recerques.

#### 4.- Principis científics

Després de la contextualització històrica de la qüestió, cal fer-ne una contextualització en referència als principis científics concernits. Ens fixarem sobre la importància que la interferència té en els estudis lingüístics, en un exemple prototípic de l'àmbit de la lingüística forense pel que fa als mínims quantitius dels materials analitzats i en dos principis del mètode científic que tindran importància en el nostre raonament posterior.

##### 4.1.- sobre la interferència

Tal com recull el reputadíssim professor David Crystal,

Properties of the L1 are thought to exercise an influence on the course of L2 learning: learners 'transfer' sounds, structures, and usages from one language to the other. A widely used typology distinguishes two kinds of transfer. Similarities between the two languages cause 'positive transfer': it proves acceptable to use the L1 habits in the L2 setting (e.g. the assumption that the subject goes before the verb satisfactorily transfers from English to French). Differences cause 'negative transfer', generally known as 'interference': the L1 habits cause errors in the L2 (e.g. the same assumption about subject-verb order does not satisfactorily transfer into Welsh<sup>1</sup>

Moreover, not all errors are equally systematic, disruptive, or unacceptable. Errors of vocabulary, for example, are less general and predictable than errors of grammar, but they are usually more disruptive of communication. Some errors, indeed, become so acceptable that they do not disappear: they become 'fossilized' – tolerated by learners (insofar as they are conscious of them) because they do not cause major problems of communication (e.g. the pronunciation errors that constitute a foreign accent)<sup>2</sup>

Hem triat aquests mots per l'autoritat acadèmica del seu autor i per la claredat amb que estableixen el caràcter profundament psicològic (i difícilment perceptible per un individu qualsevol del segle XV) de la influència que la llengua materna (tècnicament, L1) exerceix sobre les altres llengües que un individu incorpora al llarg de la vida (tècnicament, L2), especialment amb un aprenentatge exclusivament pragmàtic.

##### 4.2.- sobre la quantitat de material necessari

Aquesta és una qüestió que a priori pot generar prevencions, atesa la migradesa dels materials històricament conservats. Ens centrarem en un dels casos arquetípics de la lingüística forense, en què es dona per provat que una confessió (que va servir d'evidència per dictar una pena de mort, que fou executada) va ser manipulada per la policia. Aquest cas va contribuir a la supressió de la pena de mort al Regne Unit i es considera el punt de partida de la lingüística forense:

1 (Crystal 1987:372).

2 (Crystal 1987:372-3).

de mort al Regne Unit i es considera el punt de partida de la lingüística forense:

The actual phrase Forensic Linguistics was not used until 1968 when a linguistics professor by the name of Jan Svartvik recorded its first mention in a now famous analysis of statements given to police officers at Notting Hill Police Station in 1953. Timothy John Evans was accused of the murder of his wife and baby at 10 Rillington Place, Notting Hill, London, England, tried at the Central Criminal Court of England and Wales (The 'Old Bailey') and hanged at Pentonville Prison. In the 1960's the statements he had allegedly given to police following his arrest, troubled several people, including a well-known journalist by the name of Ludovic Kennedy, and Svartvik was commissioned to analyse the statements. Svartvik was one of the earliest linguists involved in corpus studies, which is the systematic analysis of language through the collection and study of large bodies (hence *corpus*, pl. *corpora*) of language, and therefore he was able to approach the task of analysing the Evans' statements in a methodical manner. He quickly realised that the statements contained two styles and he set about quantifying the differences, ultimately demonstrating that they were, in fact, an educated written style and a marked spoken style. Along with other evidence collected in the course of many different threads of investigation, the findings of Svartvik showed that Evans could not, as had been claimed at his trial, have dictated the statements attributed to him.<sup>1</sup>

Com diu David Crystal, respecte a l'estudi del Dr. Svartvik,

In an analysis of the Evans statements, amounting to nearly 5000 words, it proved possible to show that the language contained many conflicting stylistic features...<sup>2</sup>

Amb aquest apartat hem volgut prevenir qualssevol opinions adverses al treball fet per raó de l'extensió dels materials analitzats, o el que és el mateix, hem volgut aclarir perquè vàrem acceptar l'encàrrec sense reserves quant a l'aplicabilitat de mètodes lexicomètrics als escrits colombins.

#### 4.3.- sobre la tangibilitat i validesa de les proves

Per tal que una teoria es consideri vàlida no cal que porti proves tangibles o irrefutables sinó que cal que permeti explicar de manera més plausible tots els fets coneguts:

Also, scientific progress is actually based on faith. When any new, radical ideas are thought up, a theory is worked out and developed, by faith, with no experimental backup. Then the theory must be accepted enough to be tested experimentally - sometimes costing millions of pounds - purely on faith that the idea is worth investigating. Many commonly held theories are still unproven, and are just what people think fit the facts. We stand by these theories by faith until something better comes along.<sup>3</sup>

Doncs, ens proposem determinar quina era la llengua materna de Colom, la que explica les interferències grafo-fonològiques en els seus escrits, de manera que contribueixi a "fit the facts". Això tindrà acceptabilitat metodològica si plantegem un procediment per falsar les nostres pròpies aportacions; és a dir, què cal fer per demostrar el contrari i per relativitzar-ne les conclusions:

1 John Olsson [http://www.thetext.co.uk/docs/what\\_is.doc](http://www.thetext.co.uk/docs/what_is.doc)

2 (Crystal 1987:69).

3 <http://plato.wadham.ox.ac.uk/~cu/HTML/articles.html> (christian union at wadham college, Oxford University)

In place of naïve falsification, Popper envisioned science as evolving by the successive rejection of falsified theories, rather than falsified statements. Falsified theories are to be replaced by theories which can account for the phenomena which falsified the prior theory, that is, with greater explanatory power.<sup>1</sup>

#### 4.4.- la navalla d'Occam

El darrer element de reflexió que aportarem fa referència a l'acceptabilitat de certes argumentacions sota els principis del mètode científic:

La navalla d'Occam (o navalla d'Ockham, o principi d'economia) fa referència a un tipus de raonament basat en una premissa molt simple: la solució més senzilla és probablement la correcta. Concepte ideat per Guillem d'Occam. Aquesta regla ha tingut una importància cabdal en el desenvolupament posterior de la ciència. S'utilitza fonamentalment com complement de les lleis de la lògica, amb la finalitat d'evitar el pensament màgic. Segons aquest principi, sempre que es trobin diverses explicacions a un fenomen, s'ha d'escollir la més senzilla que ho expliqui per complet.<sup>2</sup>

Aquest principi, denominat originalment “frustra fit per plura quod potest fieri per pauciora”, ens ha resultat útil en valorar alguns dels que hem anomenat factors concurrents:

#### **Per què Colom escriu en castellà als seus banquers genovesos?**

Perquè creia en la capacitat màgica de les cartes de transmutarse d'idioma en creuar una frontera

Perquè era un exhibicionista intel·lectual i volia lluir el seu domini (?) del castellà

Perquè no tenia altre idioma en comú amb els seus banquers genovesos

#### 5.- Variables triades i motius

Hem adoptat quatre variables fonològiques en què el català de l'època (més precisament, algun dels seus dialectes) difereix del castellà coetani.

Aquestes variables són

1. confusió de 'a' i 'e' en síl·laba àtona
2. confusió de 'o' i 'u' en síl·laba àtona
3. confusió generalitzada de 'b' amb 'v'
4. confusió generalitzada de 'ce/ci' amb 's' i les seves variants gràfiques

Les triem pel seu caràcter diferenciador respecte a d'altres llengües i perquè les considerem operatives per a l'època de Colom. Reproduïm fotogràficament les fonts per tal com en algun article de molt dubtosa qualitat científica hem trobat atribuïdes a Badia i Margarit afirmacions contràries al que aquí es pot llegir.

<sup>1</sup> <http://en.wikipedia.org/wiki/Falsifiable>

<sup>2</sup> <http://ca.wikipedia.org/wiki/Occam>

*a, e, o, u*, a les quals fan correspondre quatre sons diferents.<sup>1</sup> A darreries del segle XV era general la confusió de *a-e* i *o-u* àtones en el dialecte central, com es pot comprovar perquè d'aquesta època són les normes que condemnen “*vaguer per veguer*”, “*pera per pere*”, “*massete per masseta*”, “*ànimas per ànimes, e semblants*”, i “*porgatori per purgatori*”, “*culgar per colgar*”.<sup>2</sup>

-----  
Badia i Margarit, *Gramàtica Històrica*, pàg. 120

II] Fricatives.—A. Làbio-dental *v*. 1) Sembla que l'articulació làbio-dental de la *v* llatina es conservà a la Romània durant segles; així veiem que la llengua catalana antiga distingeix acuradament la *v* de la *B* bilabial. Tanmateix, a darreries de la mateixa Edat Mitjana es confondrien els dos sons en el bilabial modern, exemples: VENTRE > *ventre bēntre*, VINU > *vi bí*, VOCE > *veu béu*. \*2) La substitució de la làbio-dental antiga *v* per la bilabial *b* no ha estat universal en català: a) conserven avui l'articulació làbio-dental: el dialecte balear,<sup>3</sup>

Badia i Margarit, *Gramàtica Històrica*, pàg. 178

D. Sibilant sorda *C* (+ *E, I*). La *C* (+ *E, I*) havia avançat el seu punt d'articulació ja en llatí vulgar, arribant a pronunciar-se africada pre-palatal i assibilada. 1) Aquesta primitiva africada sibilant *š* (com *ts*) es mantingué molt de temps en català medieval;<sup>4</sup> en el segle XV s'havia simplificat en sibilant fricativa (*s*) que és l'articulació de la llengua moderna: CAELU > *cel* ant. *šél*, mod. *sél*; CIVITATE > *ciutat siutát*. 2) En l'etapa primitiva africada i pel caràcter una mica pala-

Badia i Margarit, *Gramàtica Històrica*, pàg. 180

Podem contrastar aquestes afirmacions amb la *Gramàtica española* de Juan Alsina Franch i José Manuel Blecu:

La confluencia de *b* y *v* en el castellano en /b/ realizado como [b] ~ [β] es muy complejo (...) La confluencia cubre una amplia zona del sudoeste románico en la que hay que incluir, además del castellano, todo el gallego y el portugués del Norte, el leonés, el aragonés y gran parte del catalán, el gascón y una zona lingüística del Mediodía francés; (...) La historia de la confusión desde el punto de vista de las descripciones de los gramáticos clásicos fue redactada por A. Alonso, *De la pronunciación medieval a la moderna en español*, Gredos, 1955, pp 25-48. Según las noticias examinadas, "Hasta después de 1550, la confusión era propia del Norte de España; de Burgos dicen la mayor parte de los testimonios", *op. cit.* p. 25; (...) La confluencia existiría ya a fines del siglo XV, aunque -como advierte Lapesa- el uso general distinguiría los dos fonemas. (pàg. 310)

(...) lo que hoy llamamos *seseo* y *ceceo* son variedades que resultan de un complejo

(...) lo que hoy llamamos *seseo* y *ceceo* son variedades que resultan de un complejo conjunto de cambios sistemáticos ocurridos en los siglos XV y XVI, irradiados desde Sevilla al resto de Andalucía, Canarias e Hispanoamérica. (pàg. 344)

La importància d'aquestes variables rau en el fet que com a grup no coincideixen ni amb el portuguès, ni amb el castellà ni amb l'occità. La divergència és encara més contundent (i més pertinent) si comparem amb les variants protoitalianes coetànies. Tot això a partir de treballs que no poden haver estat influïts per les necessitats d'aquesta recerca car han estat publicats amb molta anterioritat. En el cas del castellà, la confusió b-v només es troba en zones de substrat basc (llengua que no les diferencia) o adstrat asturleonès o aragonès i la confusió /s/ vs /θ/ s'expandeix justament amb la colonització d'Amèrica.

## 6.- Materials analitzats

A proposta dels Srs. Arenas i Albardaner, del *Centre d'estudis colombins* d'Òmnium Cultural, el material triat varen ser les cartes autògrafes de Cristòfor Colom segons edició de Consuelo Varela<sup>1</sup>. Es tracta d'un conjunt d'escrius que cobreixen el període entre el 4 de gener del 1493 i el 5 de febrer de 1505 i que els entesos estan d'acord a considerar autògrafs. Tampoc no qüestiona ningú la qualitat de la transcripció feta per a l'edició citada.

Aquests escrits han estat introduïts a l'ordinador i tractats amb el paquet TACT (Text Analyzing Computer Tools) del *Centre de càlcul per a les humanitats* de la Universitat de Toronto<sup>2</sup> (Canadà) i amb el programa Simple Concordances Program (SCP).

Totalitzen 39895 ocurrencies de 5222 mots diferents.

Revisats els 5222 mots diferents (annex 1), se'n varen seleccionar els que presenten vacil·lacions en alguna de les grafies corresponents als fonemes triats o sense presentar vacil·lacions coincideixen amb el català i no amb el castellà. Les d'aquest segon grup, pendents de confrontar-les amb la grafia predominant en el castellà coetani, operació de què parlarem després.

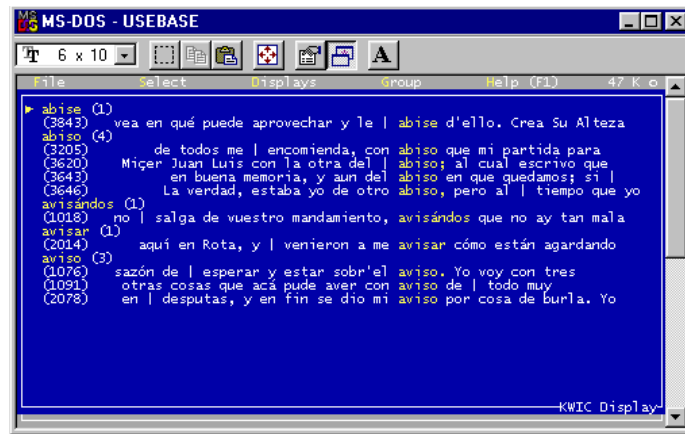
La recerca s'ha fet sense cap contaminació semàntica: no vàrem llegir les cartes, sinó que triàrem els mots a partir dels llistats generats pel SCP i el TACT, excepte en casos com 'avierta' en què es podia dubtar de si es referia a 'abierta', com és el cas, o a 'advierta'; en el primer cas pertoca a la variable b-v, mentre que el segon es referiria a reducció vs manteniment de grups consonàntics, que no va ser objecte d'estudi.

La imatge següent mostra unes concordances amb el TACT:

1 "Cristóbal Colón, Textos y documentos completos", Edición de Consuelo Varela; Nuevas Cartas: Edición de Juan Gil, Alianza Universidad, Alianza Editorial, edición 1992

2 Copyright (c) 1994 by John Bradley, Lidio Presutti, Michael Stairs and Centre for Computing in the Humanities University of Toronto





A l'annex núm. 2 es trobaran els llistats d'aquestes paraules.

## 7.- Primers resultats

Una primera aproximació numèrica a la incidència d'aquestes gràfies inestables o contradictòries que ens permeten entreveure els hàbits fonètics adquirits per Colom és la següent:

<b>mots</b>	<b>absolut</b>	<b>percentatge</b>
alternança a-e	13	0,03
alternança o-u	233	0,58
alternança b-v	711	1,78
alternança s-ç-ss-c	1063	2,66
conjunta	1940	4,86
ocurrències	39895	100

Doncs, quasi un cinc per cent de les ocurrències del corpus presenta alguna de les alternances estudiades (algun mot n'acumula més d'una), cosa que ja és per si mateixa significativa, i a més cadascuna de les alternances, amb excepció de la primera, presenta valors clarament significatius. Com demostrarem després, Colom controla força el seu castellà (els errors no es distribueixen aleatòriament) i això explicaria dos fets: a) que alguna d'aquestes alternances presenti els valors propis d'un aprenent avançat, força diferents dels d'un principiant i b) que l'alternança a-e escapi a les nostres prediccions per tal com Colom hi parés més atenció que a la resta. La poca presència de mots amb vacil·lacions a-e també es pot atribuir al fet que la neutralització d'aquests sons no és realitzada amb un dels concurrents, com passa amb b-v, realitzada amb [b], i amb o-u, realitzada amb [u], sinó que la neutralització de les grafies a-e en síl·laba àtona dóna vocal neutra [a e ε] → [ə]. En aquest fet, se n'hi afegeix un altre: un nombre important de neutralitzacions de a-e es dóna en alternança entre síl·laba tònica (mantinguda) i àtona (neutralitzada), com passa entre formes rizotòniques i arizotòniques dels paradigmes verbals: ['ba'lə bə'lɛm 'deʃu də'ʃem 'pɛnsu pən'sem]. En resum, la variable a-e no va ser una bona tria en la confrontació català-castellà, que ens hauríem pogut estalviar si haguéssim fet el treball més reposadament, però fer recerca amb el productor audiovisual que espera resultats enganxat a l'altre cap del telèfon pot tenir aquests desavantatges.

En canvi, podem aprendre algunes coses en relació a les alternances o-u i b-v. En tots dos casos, una de les dues possibilitats sembla més previsible, car coincidiria amb la de la llengua de referència. Per a la nostra hipòtesi, es podria pensar que això implica que esperaríem moltes "u" en lloc de "o",



moltes "b" en lloc de "v", molt poques "o" en lloc de "u" i molt poques "v" en lloc de "b". Però els experts en aprenentatge de llengües saben que això és així en les primeres etapes de l'aprenentatge, on predomina la interferència; en canvi, a mesura que l'aprenent va esdevenint familiar amb la llengua apresada, comença a prendre consciència de la seva propensió a cometre certs errors, que intenta corregir. En aquesta etapa, intermèdia entre les inicis i el domini satisfactori, el detall delatador és la hipercorrecció: l'aprenent aconsegueix corregir molts dels seus errors sobreactuant. Per posar-ho en clau moderna, l'aprenent d'anglès que no diferencia en la seva llengua materna 'b' de 'v', a l'inici pronuncia 'berigut' per 'very good'; passat un temps, potser dirà 'veriviutiful' per 'very beautiful'. En aquesta etapa intermèdia els errors de la primera donen pas a un còctel d'encerts i hipercorreccions (correccions errònies de falsos errors). Aparentment sembla que domini la hipercorrecció, cosa que ens indicaria que no és un aprenent recent, encara que després veurem que aquestes dades presenten una cronologia curiosa:

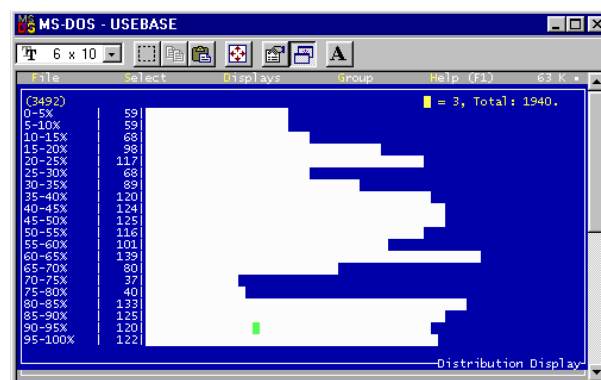
s'esperava "o" i fa "u" (interferència)	49
s'esperava "u" i fa "o" (hipercorrecció)	186
s'esperava "v" i fa "b" (interferència)	306
s'esperava "b" i fa "v" (hipercorrecció)	416

Descartem l'alternança a-e per poc significativa i aparquem de moment l'alternança s-ss-ç-ce/ci per la dispersió de la seva multilateralitat; les dades precedents relatives a a les alternances b-v i o-u permeten concloure:

- estem davant d'una persona que confon *b* amb *v* i, en síl·laba àtona, *o* amb *u*
- si els sons interferents són *v* i *o* àtona, es troba començant la fase intermèdia del seu aprenentatge i la seva llengua materna és força rara (no té ni 'u' àtona ni 'b'). En aquest cas, caldria reformular la taula precedent intercanviant *interferència* per *hipercorrecció* i viceversa.
- si els sons interferents són *b* i *u* àtona, es troba en la fase intermèdia-avançada del seu aprenentatge i la seva llengua materna està entre les que no tenen 'o' àtona ni 'v'

## 8.- Resultats detallats: b-v i o-u

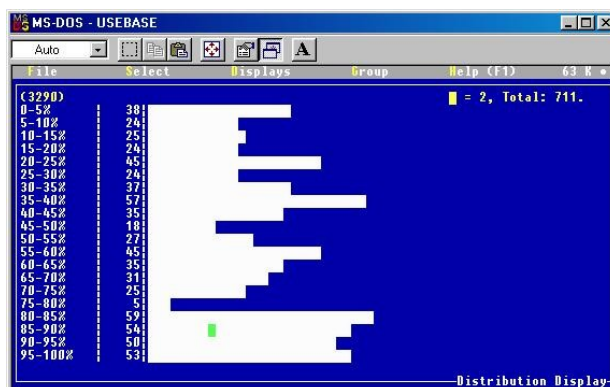
Ara és el moment de donar un cop d'ull a la distribució cronològica d'aquestes interferències i hipercorreccions. Atès que els documents del corpus estan ordenats cronològicament, la seva aleatorietat o determinades regularitats poden ser pertinents. Utilitzant la funcionalitat del TACT de mostrar la distribució de les paraules seleccionades, cercarem de veure què ens aporta això en relació a les nostres variables. Mostrem en primer lloc la distribució global de totes les 1940 ocurrències filtrades repartides en 20 blocs consecutius del 5% del corpus cadascun:



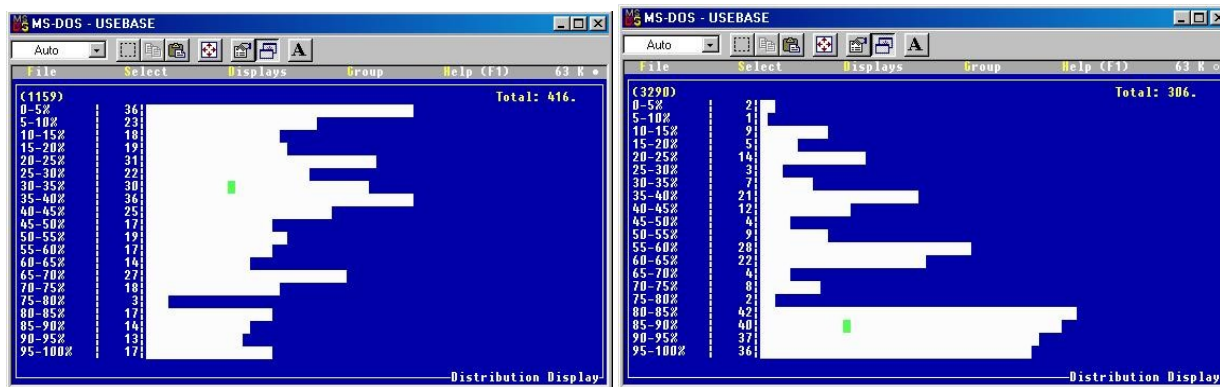
Aquest histograma s'interpreta llegint a cada fila les dades corresponents a un 5% del corpus, identificat a la primera columna. A la segona columna hi ha el valor absolut de la variable observada respecte al 5% corresponent i a la seva dreta la longitud de la barra horitzontal és proporcional al valor de la segona columna amb una escala expressada amb █ = 3. Així, la primera vintena part del corpus té 59 casos del conjunt de variables observades, valor que es repeteix en la segona vintena part, etc.

Aquesta gràfica ens deixa clar que, amb excepció del *forat* entre el 70% i el 80%, la distribució és moderadament ascendent dins d'una certa regularitat. Més endavant ens fixarem en aquest *forat*.

Si ens centrem en el cas b-v, la seva gràfica conjunta és la següent:



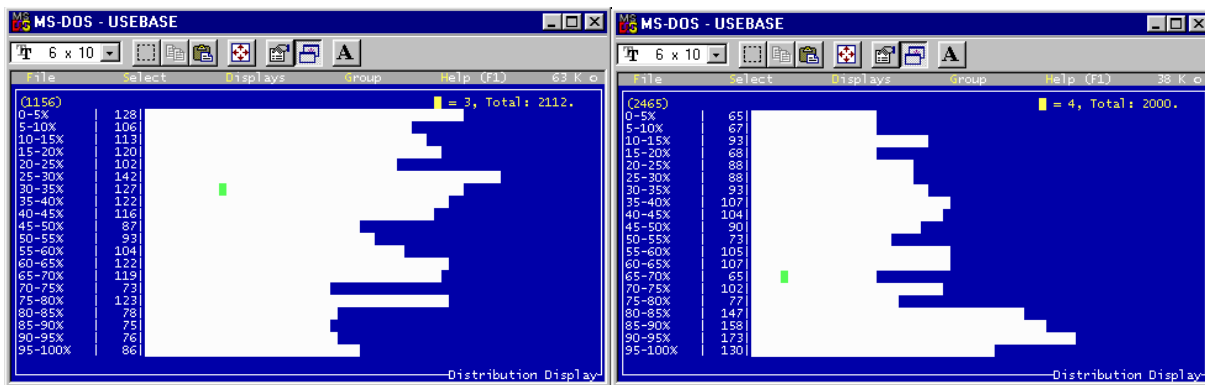
...amb la mateixa tendència i el mateix forat. Si separem les dades de la v dels de la b, obtenim



Esperàvem 'b' i trobem 'v'

Esperàvem 'v' i trobem 'b'

amb dues constatacions clares: a) les dues grafies es complementen amb una distribució reveladora d'alguna modificació de les pautes d'escriptura de Cristòfor Colom en el sentit que les quatre darreres parts (el darrer 20%, el més tardà) concentren quasi el 40% del total de 'v' produïdes en comptes de 'b'; i b) que el 'forat' no solament té importància per si mateix sinó que podria interpretar-se com la causa del canvi posterior. Així les coses, ens preguntem si aquest canvi només afecta les paraules que hem triat com a possiblement reveladores d'interferència o hipercorrecció o si això afectaria el conjunt de paraules que contenen 'b' o 'v'. Si estiguéssim confrontant 's' contra 'a', ateses les peculiaritats de la morfologia del castellà, una modificació de la proporcionalitat de 's' en favor de 'a' o viceversa podria deure's al fet que en aquell fragment predominessin els plurals o els singulars. Però, tractant-se de la parella b-v, no se'ns acut cap motiu casual que pugui justificar-ho. Vegem, doncs, la distribució global de tots els mots amb 'b' i amb 'v':

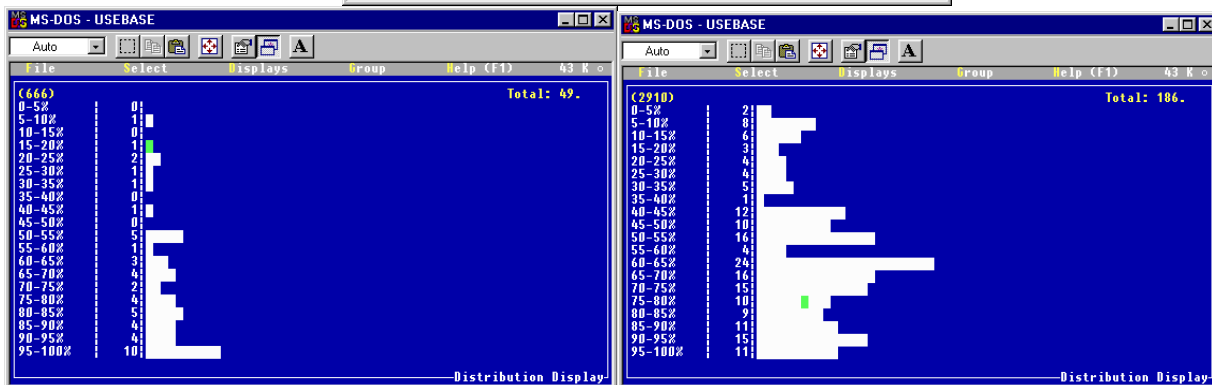
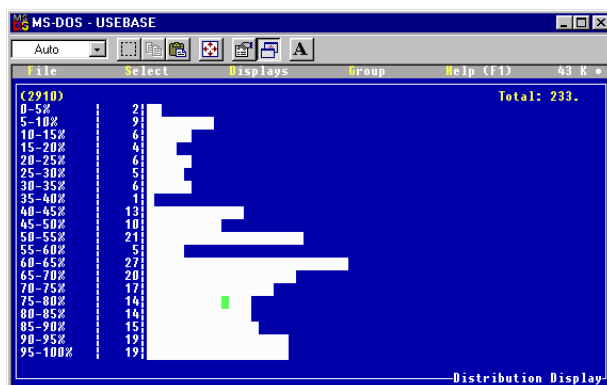


Mots que contenen "v"

Mots que contenen "b"

És evident que passa alguna cosa que no té relació amb el contingut (doncs, amb el lèxic) o amb pautes discursives: simplement canvien els criteris ortogràfics; segurament, a més, podríem haver estat molt més inclusius a l'hora de confegir la llista de mots reveladors d'interferència.

Trobem una coincidència no negligible en parell o-u



Esperàvem 'o' i trobem 'u'

Esperàvem 'u' i trobem 'o'

Cal descartar que aquestes confusions fossin influència dels castellanoparlants de l'entorn de Colom o que puguin venir d'alguna altra llengua. Arribats en aquest punt, doncs, per poder donar tota la seva força a les dades que hem obtingut, ens sembla pertinent separar les llengües eventualment candidates a llengua materna de Colom que podrien exercir aquestes interferències de les que no i per això en tindrem prou amb l'alternança b-v.

André Martinet, en el seu monumental *Economia de los cambios fonéticos*, pàgines 441-448, considera que el castellà perd la distinció b-v a començament del S. XVII, cent anys després de la mort de Colom i que aquesta pèrdua només es troba abans a les zones del nord d'influència èuscarà

(substrat i repoblació) o pròximes al lleonès, a l'aragonès o al bearnès (gascó).

Heinrich Lausberg, a les pàgines 310-311 i 356 de la seva *Lingüística romànica*, classifica les llengües romàniques en relació a l'alternança b-v:

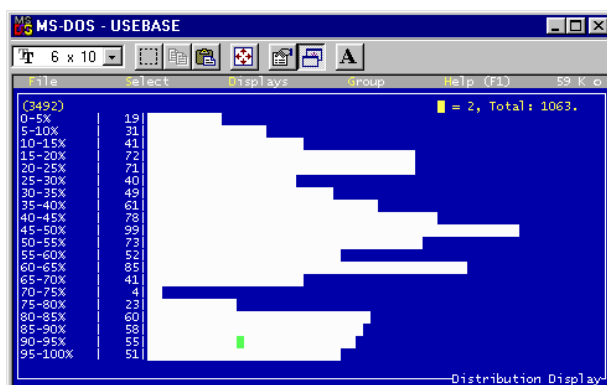
La coincidència fonològica de llatí b- i v- (§ 300) presuposa para llatí v- (§ 297) el grau fonètic [β]. Las regiones que dan al llatí v- una articulació labiodental [v] (cf. § 297), evitaren la confusió de ambos sonidos en principi de dicción. Son éstas; centro y norte de Italia, retorromano, francoprovenzal, francés, provenzal (pero cf. § 300), portugués y rumano.

Llatí -v- (...) el resultado normal es [v] en norteitaliano, retorromano, francoprovenzal, francés, provenzal y portugués.

Cristòfor Colom escrivia sota la influència d'uns hàbits fonatoris en què no existia l'oposició b-v en general ni l'oposició o-u en general o en algunes posicions. Això descarta una part dels orígens italians (incloent el genovès) i una part dels orígens catalans (però no els majoritaris).

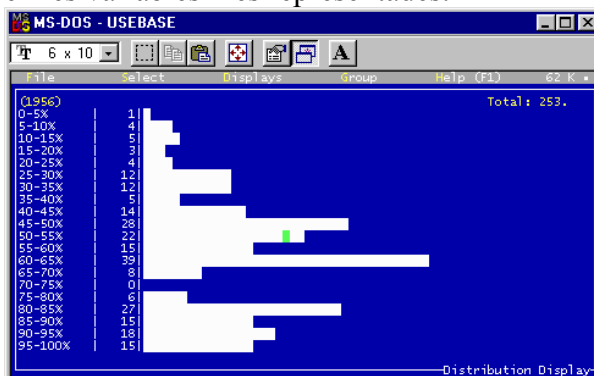
### 9.- Resultats detallats: s-ss-ç

Passant al cas de la confusió s-ss-ç, el primer que podem observar és que el conjunt dels mots observats també presenta un forat entre el 70% i el 75%:

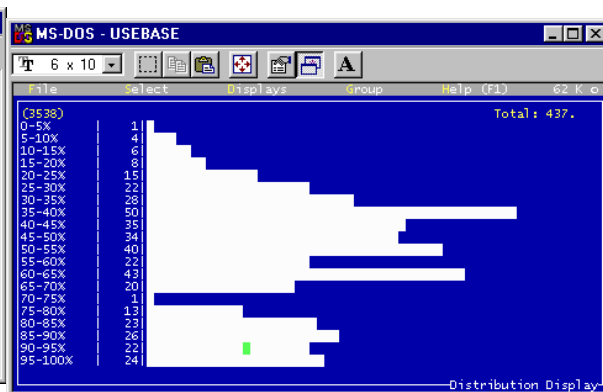


Mots amb s-ss-ç

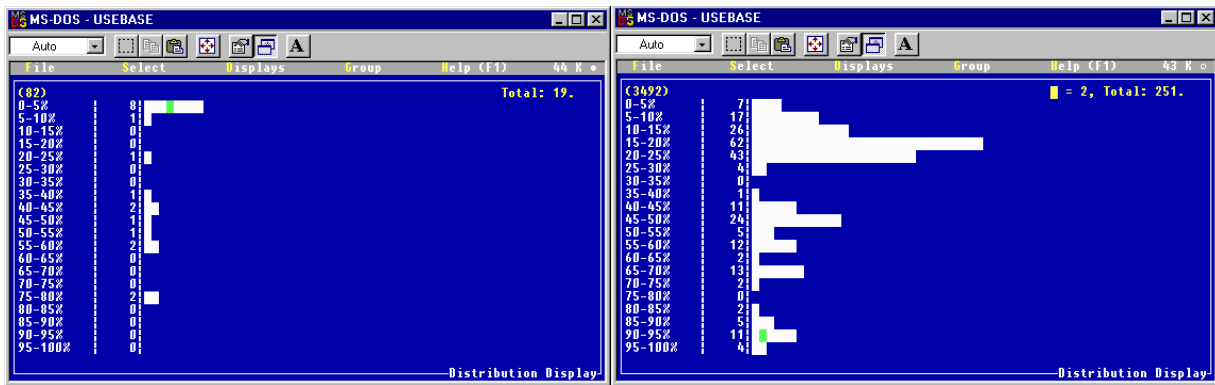
Si ho desglossem en les diverses components, observem unes tendències clares: 'ss' és testimonial, però quasi el 50% es concentra a l'inici; els altres tres també contenen el famós 'forat', més visible en les variables més representades.



Mots amb "çe"



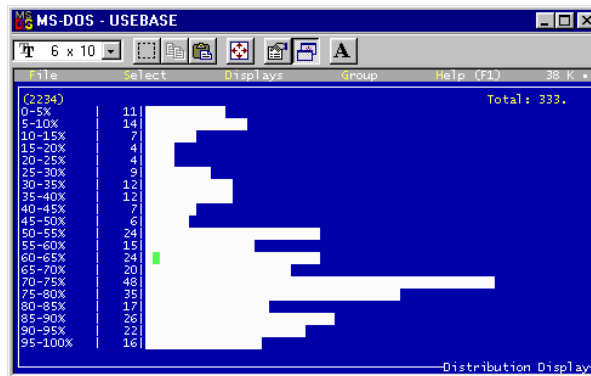
Mots amb "çi"



*Mots amb "ss"*

*Mots amb "seseo"*

Sembla totalment evident que aquest forat té un pes específic en la caracterització del corpus, de manera que ens crida l'atenció la variable complementària za-zo-zu, que destaca justament allà on les altres variables tenen el forat



*Mots amb "za-zo-zu"*

## 10.- Què ens diu el forat?

Hem observat en la majoria dels histogrames que en l'interval entre aproximadament el 70% i el 80% hi ha una modificació de la pauta ortogràfica de Colom. Aquest fet és molt significatiu, especialment quan és tan evident que no pot ser fortuït. Ens indica que en els textos en qüestió hi va haver algun canvi respecte a les variables controlades. També ens confirma que les variables triades són significatives. Hem de recordar que no solament hi ha el forat, sinó que el canvi es detecta en el conjunt del vocabulari amb 'b' (que decreix des d'aquell punt) i amb 'v' (que creix). Doncs:

- algú va influir sobre Colom quan redactava o copiava aquests escrits
- aquest influx va persistir (segurament perquè Colom va interioritzar part de les correccions)
- Colom està més pendent dels seus escrits des d'aquell moment

Ens vàrem plantejar fins i tot la possibilitat que les cartes LXIII i LXIV no fossin autògrafes de Colom i en vàrem demanar confirmació a la Dra. Consuelo Varela, la qual va confirmar que són autògrafes, sense descartar que haguessin estat fetes a partir d'esborranys revisats.

En paraules del Sr. Albardaner, "es produeix en un moment històric decisiu: la seva detenció per Bobadilla i la seva caiguda en desgràcia. Colom fa el tercer viatge de 1498 al 1500 i fa el quart viatge del 1502 al 1504. Els textos 63 i 64 són per reclamar els seus drets. A més a més que la raó de la Dra. Varela sigui encertada, que Colom va copiar autogràficament un informe fet per una tercera persona, resulta que Colom perd la confiança, la poca confiança que tenia, en els reis. Per

tant està escarmentat i es torna molt cuidados, recelós i caut en tot el que fa. Doncs, no fóra d'estranyar que a partir del desembre del 1500 els textos colombins tinguin moltes més *hipercorreccions* que no pas abans de 1500 !"

## 11.- Falsabilitat

Com es poden falsar les nostres aportacions?

- demostrant que Badia i Margarit i el comú de la romanística s'equivoquen respecte a la cronologia de les neutralitzacions b-v, o-u o c-s-ç-ss en els dialectes del català que les tenen o respecte al fet que els dialectes protoitalians del nord d'Itàlia no les tenen.
- demostrant que la transcripció que hem utilitzat no és correcta (perquè la versió publicada no s'ajusta al facsímil o perquè el corpus no s'ajusta a la versió publicada)
- demostrant que els escrits analitzats no són de Colom
- aportant-ne d'altres (de Colom) que ofereixin uns resultats diferents en cercar-hi les nostres variables
- demostrant que el corpus analitzat evidencia altres interferències que són incompatibles amb la possibilitat que la L1 de Colom sigui el català
- demostrant que altres llengües amb les mateixes neutralitzacions que el català ofereixen versemblança històrica de ser la llengua materna de Colom (descartariem, v.g., el quítxua o el suahili per motius diferents)

Algunes d'aquestes proves invalidarien totalment les nostres conclusions; d'altres simplement les relativitzarien.

## 12.- Conclusions fins al punt

Cal dir que el nostre objectiu inicial (determinar si la llengua materna de Colom podria haver estat el català) està àmpliament corroborat perquè

- la casuística de les variables triades apriorísticament (tres localitzades satisfactòriament vs una no significativa, justament la que tenia menys possibilitats), mesurada a través de paraules escollides sense cap mena de prejudici (ni per lectura prèvia dels documents ni pel seu nombre d'ocurrències) presenta una fenomenologia significativa;
- la seva distribució no és aleatòria, com es podria pensar d'una variabilitat no induïda o independent, sinó que presenta uns perfils cronològics congruents amb l'evolució natural en un procés d'adquisició de L2
- el nombre total d'ocurrències de les paraules escollides està molt per sobre dels mínims exigibles

però això no exclou que altres llengües que presentin les mateixes oposicions fonològiques que el català en relació als sons estudiats també podrien considerar-se candidates a llengua materna de Colom; atès que això exclou el protoitalià parlat a Gènova, caldria trobar explicació als motius d'amagar aquest altre origen. Tenint en compte que en el moment pressuposat de la naixença de Colom la corona d'Aragó senyorejaven tot el Mediterrani occidental i tenia sota la seva sobirania els regnes de Sicília i Nàpols, aquest catalanoparlant podria ser un subdit català de naixement sicilià o napolità o d'antecessors genovesos (Pere III, en esdevenir rei de Sicília, va mantenir-la sota l'administració dels genovesos que hi dugué Carles d'Anjou).

A petició nostra, aquestes conclusions varen ser sotmeses a contraanàlisi. La productora de Discovery Channel va escollir com a auditor extern el Pfr. Dr. Dr. Max Pfister, eminent romanista de la Universitat de Saarland, i una versió electrònica del corpus va ser facilitada al seu col·lega Hans Uszkoreit<sup>1</sup>, del Centre d'Intel·ligència Artificial de la mateixa universitat, per tal que presentessin objeccions o retrets al mètode aplicat. Aquests investigadors varen disposar per a la seva contraanàlisi de més temps del que nosaltres vàrem tenir per a l'anàlisi original i tot i això no en varen presentar cap, d'objecció o retret al mètode aplicat .

### 13.- Ampliació: contrast amb CORDE

Amb posterioritat al treball desenvolupat a petició de Discovery Channel i que ens va permetre arribar a les conclusions anteriors, ateses les reserves amb què vàrem haver de fer algunes de les nostres estimacions, ens va semblar oportú analitzar quina part de la variabilitat ortogràfica de Colom podria estar contaminada pels usos habitualment vacil·lants dels seus coetanis, per tal com la rigidesa de la norma ortogràfica és un concepte modern que no es pot aplicar anacrònicament sense flexibilitat.

Per mesurar aquesta qüestió, vàrem cercar al Corpus Diacrònic del Español (CORDE) de la Real Academia Española de la Lengua (castellana, és clar), que conté una bona col·lecció de textos medievals en castellà, totes les paraules en què es basen les nostres mesures al corpus colomí. Cada paraula la vàrem cercar dues vegades (amb *v* o amb *b*, etc), en algun cas per més d'una variable, com *cudicia-codiçia* i *cudiçia-cudicia*. 761 parells (1522 cerques) varen donar positiu en el sentit que almenys una de les dues estava documentada.

Els resultats de l'annex 3 s'interpreten segons el fragment següent:

		1	2	3	4	5
començar	comenzar	367	92	6	4	4 freqüentíssim
comienço	comienzo	1802	148	54	16	16 freqüentíssim
comparaçión	comparación	46	16	66	20	rar
composiçión	composición	8	5	44	10	raríssim
conbalesçerán	conbalescerán	3	2	0	0	irrellevant
conçebtos	concebtos	3	2	0	0	irrellevant
conçedido	concedido	23	17	52	39	raríssim
conçedieron	concedieron	4	3	9	6	raríssim
conçedió	concedió	5	4	7	5	rar
conçedo	concedo	5	4	16	10	raríssim
conçertada	concertada	28	15	27	20	freqüent

Al costat de les dues formes cercades (essent la primera la candidata a interferència o hipercorrecció) hi ha les columnes següents:

- 1 nombre de casos de la primera forma
- 2 nombre de documents en què apareix la primera forma
- 3 nombre de casos de la segona forma
- 4 nombre de documents en què apareix la segona forma
- 5 classificació

<sup>1</sup> **Hans Uszkoreit** Professor of Computational Linguistics at the Dept. of Computational Linguistics and Phonetics of Saarland University at Saarbrücken. / Scientific Director at the German Research Center for Artificial Intelligence (DFKI) and Head of DFKI Language Technology Lab. segons la seva pàgina web



Per fer aquesta classificació, hem considerat irrelevant els mots que no totalitzessin 10 ocurrències entre les columnes 1 i 3 (el CORDE coincideix amb el nostre corpus en algunes cartes colombines i distorsionaria els nostres resultats un cas com *conbalesçerán*, per tal com totes les ocurrències trobades podrien pertànyer justament als escrits colombins; doncs, aquest coincidència en comptes de mostrar que serien freqüentíssimes indicaria que són exclusives de Colom)<sup>1</sup>. La fórmula per calcular els altres casos ha estat

=SI(G503+I503<10;"irrelevant";SI(G503>I503;SI(G503>I503\*2;"freqüentíssim";"freqüent");SI(G503=I503;"igual";SI(G503\*2<I503;"raríssim";"rar"))))

que vol dir que hem considerat que la grafia observada a Colom era freqüent si el valor de la columna 1 és més gran que el de la columna 3, era igual si coincideixen i era rar en cas contrari. Hem dividit els freqüents i els rars afegint els superlatius quan la columna 1 dobla la 3 (freqüentíssim) o la 3 dobla a 1 (raríssim). Doncs, *raríssim* vol dir que la grafia colombina no arriba a una tercera part en el CORDE i *freqüentíssim* vol dir que hi supera els dos terços.

		<b>bv</b>	<b>vb</b>	<b>ou</b>	<b>uo</b>
freqüentíssim	290	89	22	3	17
freqüent	107	8	2	0	7
igual	5	0	0	0	1
rar	47	2	6	0	3
raríssim	185	18	69	13	22
irrelevant	126	19	26	1	10

Aquestes dades ens aporten tres conclusions:

- La primera, és que serà interessant ampliar aquest estudi, que és evidentment revelador, però que ho seria més si poguéssim fer una recerca molt més detallada (centrada en el lema i no en la forma), sobre tots els textos del CORDE exceptuant-ne els colombins i altres textos d'autors que no hi hagi certesa que siguin nadius i comparant els seus resultats amb corpus semblants de català medieval i d'italià coetani. Ja vàrem plantejar aquesta possibilitat a Discovery Channel i ells varen fer gestions, infructuoses, per tenir accés a corpus d'aquestes característiques.
- La segona, una observació contundent en aquests totals és el fet que les dues variables corresponents a interferència (vb - ou) destaquen pels casos raríssim, ergo raríssimament atribuïbles a l'influx dels castellans amb qui Colom tenia tractes escrits. Les altres dues variables van de l'equilibri al predomini del valor freqüent, cosa que és clau per tal com la tendència natural de l'aprenent a hipercorregir-se es veu reforçada pel fet que els propis nadius comenten les mateixes vacil·lacions.
- La tercera és que les grafies que hem considerat probablement interferides per la llengua materna de Colom sumen 82 dels casos raríssims, contra 40 de les altres dues, cosa que reforça l'observació anterior.

Sense perdre perdre de vista que quasi han passat cinc-cents anys de la mort de Colom, distància

<sup>1</sup> si mai disposem de personal per fer una recerca més minuciosa, caldria revisar aquests casos, perquè verificant-los individualment podrien resultar molt rellevants; a més, seria interessant cercar per lema i no per forma, car sovint hauríem trobat una forma pròxima a la cercada

cronològica que, afegida a l'escassetat dels materials, fa molt difícil arribar a conclusions irrefutables, el que no podem menystenir és que, de totes maneres, les xifres que hem pogut extreure van en el sentit de la hipòtesi: les vacil·lacions ortogràfiques de Colom coincideixen perfectament amb les que tindria qualsevol aprenent de castellà amb uns hàbits fonatoris que incloguessin la neutralització b-v en general i la de o-u en síl·laba àtona, i altres. Es podria objectar que a partir d'aquestes xifres no sabem gaire, però sabem molt més que abans. Sabem que la seva manera d'escriure canvia, que en determinats moments ho fa molt marcadament i que majoritàriament aquests canvis no es poden explicar per acomodació als usos nadius.

#### 14.- Accessibilitat dels materials

El corpus que ha servit per fer aquestes exploracions es pot baixar, per gentilesa del Centre d'Estudis Colombins, de les adreces següents (primer en format ANSI preparat per a SCP; després en format cp 850 = MS-DOS per a TACT):

[http://terminotica.upf.es/membres/de\\_yza/colom/colom.txt](http://terminotica.upf.es/membres/de_yza/colom/colom.txt)

[http://terminotica.upf.es/membres/de\\_yza/colom/colom850.txt](http://terminotica.upf.es/membres/de_yza/colom/colom850.txt)

#### 15.- Síntesi final

De la mateixa manera que no es pot esperar tenir evidència directa del Big Bang, però s'accepta com a teoria mentre sigui la que més satisfactòriament explica els interrogants sobre origen de l'univers, vam considerar provat que el català era la llengua materna de Colón mentre no es demostrés que una altra llengua podria explicar millor les interferències de base fonològica que es manifesten en els escrits autògrafs colombins d'autoria no contestada.

#### 16.- Agraïments

Aquesta comunicació no hauria estat possible sense la insistència del Sr. Nito Verdera, a qui vull agrair tots els esforços que ha fet. Hi han aportat algun tipus d'observació o revisió els Srs. Albardaner i Marí, del CEC, i la Dra. Varela, del CSIC. Finalment Imma Fradera i Auke Remmert Oosterhoff Sanders són responsables de la digitalització del Corpus, feta el 1992 al Centre de Lexicometria de la Universitat de Barcelona sota la direcció de l'autor d'aquesta comunicació.

## Referències bibliogràfiques

Alcina Franch, Juan i José Manuel Blecua, *Gramática española*, Ariel, Barcelona, 1979

Badia i Margarit, Antoni M., *Gramática histórica catalana*, Valencia, Tres i Quatre, 1981

Crystal, David, *The Cambridge encyclopedia of language*, Cambridge University Press, Cambridge, 1991 reimpress.

Lausberg, Heinrich, *Lingüística románica*, Madrid, Editorial Gredos, S.A., 1976, 2a reimpr.

Martinet, André, *Economía de los cambios fonéticos : tratado de fonología diacrónica*, Madrid, Editorial Gredos, 1974

Parellada i Cardellach, Caius, *Colom venç Colombo : la vera naturalesa catalana del descobridor del nou món*, Barcelona, 1987

Parellada Cardellach, Caius, *Cristòfor Colom i Catalunya : una relació indefugible*, Barcelona, Hogar del Libro, 1992

Verdera, Nito, *De Ibiza y Formentera al Caribe: Cristobal Colón y la Toponimia*, Barcelona, 2000