

AYUDAS INFORMÁTICAS A LA DESCRIPCIÓN RELACIONAL INTERPRETATIVA DE LOS TEXTOS COMPÓSITOS: APLICACIÓN AL PRIMER VIAJE DE CRISTÓBAL COLÓN

VÉRONIQUE HUYNH-ARMANET
Catedrática de Lingüística-Informática
en colaboración con

FRANCISCO JAVIER SÁNCHEZ PÉREZ
Investigador-Responsable de la coordinación Informática
Laboratorio de Análisis Relacional Informático de Textos (A.R.I.T.)
Universidad de París VIII (Francia)

Esta aplicación últimamente desarrollada en nuestro Laboratorio será propuesta bajo dos enfoques constantes en nuestras investigaciones:

— uno fundamental, la descripción por vía ascendente de los modelos subyacentes a la plasticidad genuina de los idiomas, y más especialmente de la escritura, que desde hace varios años ¹, a través una dialéctica experimental ha progresado y progresa con el apoyo del ordenador, en ambientes tecnológicos diversos,

— otro instrumental, en micro-informática integrado a un proyecto de Enseñanza-Investigación ² que nuestro Laboratorio ARIT ha practicado desde 1984.

La Relación del Primer Viaje de Cristóbal Colón constituye actualmente el centro de esta investigación. Es uno de los textos privilegiados que plantean el problema de exploración de una obra parcialmente borrada pero no total-

¹ VERONIQUE HUYNH-ARMANET, *I. Des structures syntaxiques de l'espagnol à l'analyse relationnelle des textes*, París, Honoré Champion, 1977; *II. L'analyse textuelle du langage, L'Infra-littérature en Espagne au XIX et XXe siècle*, Presses Universitaires de Grenoble, págs. 321-332, 1977; *III. La ponderación contextual en la interpretación lingüística de los textos*, Actas del III Coloquio de Geografía cuantitativa, «Teoría, metodología y técnicas cuantitativas en las ciencias», Universidad de Extremadura, 26-29 de septiembre de 1988, págs. 174-194.

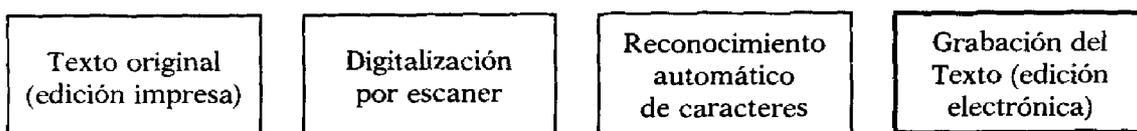
FRANCISCO JAVIER SÁNCHEZ PÉREZ, *Techniques informatiques de désambiguïsation et d'organisation par blocs des données textuelles et procédures initiales pour l'étude de l'interlocution*, Cahiers de Linguistique Relationnelle Informatique, Laboratoire ARIT, Université de Paris, VIII, 1988.

² Esta enseñanza desarrollada en nuestro Centro de *Informática Para Todos (IPT)*, dirigido por Daniel Lepage, profesor de Informática de Ciencias Humanas en nuestra Universidad, permite una iniciación precoz y una formación de investigadores de alto nivel.

mente, en una lectura-compendio, la de Bartolomé de las Casas que parece a menudo deseoso de presentar a través de varias citas autenticadas de Cristóbal Colón, «las propias palabras» del Almirante. Este texto polifónico plantea problemas difíciles incluso a metodologías bien adaptadas al estudio de la dinámica genuina de la escritura. Por eso, no pretendemos proponer, hoy, soluciones definitivas a este doble problema. Sólo intentaremos progresar hacia tales objetivos examinando este texto con nuestras metodologías informáticas más actuales.

Aplicando nuestro último protocolo experimental, hemos realizado sucesivamente:

- La lectura editorial del texto por escaner³ que representamos aquí gráficamente:



Esta grabación podrá muy pronto compararse con otras sacadas de ediciones asequibles más recientes.

- Una codificación editorial específica que además de la indicación de las páginas anota las informaciones cronológicas proporcionadas por Cristóbal Colón.

- Una codificación estructural determinada por la presencia alternada en este texto, junto a fragmentos presentados con más o menos énfasis como las propias palabras del autor (dice el Almirante), de bloques-compendios propuestos, a veces claramente, como referencias más anónimas (diz que...), pero otras veces difíciles de situar, por lo menos en el estado actual de nuestra investigación. No nos detendremos aquí en ellos.

- Una codificación suplementaria edito-estructural de bloques determinados por las tres etapas esenciales, según el mismo Colón, de este primer viaje, la **Ida** (del 3 de Agosto al 11 de Octubre de 1492), el **Descubrimiento** (del 11 de Octubre al 31 de Diciembre de 1492), la **Vuelta** (del 31 de Diciembre de 1492 al 15 de Marzo de 1493). Notaremos que los tres bloques tienen aproximadamente el mismo espacio temporal (2 meses y 8 días, 2 meses y 20 días, 2 meses y 15 días, respectivamente). El código utilizado integra naturalmente las distinciones precedentes entre citas autenticadas del Almirante y los bloques llamados aquí provisionalmente: Las Casas.

Hemos resumido en un cuadro sinóptico la codificación del texto:

	CODIFICACION
Cristóbal Colón	C C
Las Casas	L C

³ La grabación del texto por escaner se realizó a partir de una edición de las más corrientes: Colección Austral, Espasa-Calpe, 1971. Como sabemos bien se impone el deshojar el volumen de manera a poder someterlo convenientemente a la digitalización.

Viaje de Ida	IDA
Descubrimiento del Nuevo Mundo	DESC
Viaje de Vuelta	VUEL
Codificación de los meses	Aoû (Août = Agosto)
	Sep (Septembre = Setiembre)
	Oct (Octobre = Octubre)
	Nov. (Novembre = Noviembre)
	Déc (Décembre = Diciembre)
	Jan (Janvier = Enero)
	Fev (Fevrier = Febrero)
Codificación de los días	1, 2, 3, ... 29, 30, 31
	15, 16 ... 77, 78
Páginas de la edición	15, 16 ... 77, 78

Para ilustrar nuestra metodología de lectura informática propondremos seguidamente con un fragmento de la Relación del Primer Viaje la aplicación al texto del dicho código:

Día (11)	Mes (Oct)	Bartolomé de Las Casas (LC) -carácter normal-
----------	-----------	---

Jueves 11 de octubre.- Navegó al Ouesudueste. Tuvieron mucha mar y más que en todo el viaje habían tenido. Vieron pardelas y un junco verde junto a la nao. Vieron los de la carabela Pinta una caña y un palo, y tomaron otro palillo labrado alo que parecía con hierro, y un pedazo de caña y otra hierba que nace en tierra, y una tablilla. Los de la carabela Nifia también vieron otras señales de tierra y un palillo cargado de escaramojos. Con estas señales respiraron y algráronse todos. Anduvieron en este día, hasta puesto el sol, veintisiete leguas. Después del sol puesto, navegó a su primer camino al Oeste: (...)

Esto que se sigue son palabras formales del Almirante, en su libro de su primera navegación y descubrimiento de estas Indias. "Yo (dice él), porque nos tuviesen mucha amistad, porque conocí que era gente que mejor se libraría y convertiría a nuestra Santa Fe con amor que no por fuerza, les di a algunos de ellos unos bonetes colorados y unas cuentas de vidrio que se ponían al pescuezo, y otras cosas muchas de poco valor, con que hobieron mucho placer y quedaron tanto nuestros que era maravilla. (...)

Page 30

Página (Pg30)	Cristóbal Colón (CC) -citación en itálica-
---------------	--

Esta codificación nos ha permitido generar varios tipos de listas destinadas a manifestar las características generales y/o específicas del texto. Así pues, varios tipos de índices han sido realizados:

- *Índice global alfabético y jerárquico descendente* de formas fuera de contexto.

- *Índices selectivos alfabéticos y jerárquicos descendentes* de Colón y de Las Casas fuera de contexto.

- *Índices de los bloques edito-estructurales* predefinidos: Ida, Descubrimiento, Vuelta, alfabéticos y jerárquicos fuera de contexto.

- *Índices en contexto o Concordancias* centradas sobre algunas formas claves significativas.

Antes de estudiar ciertos aspectos estadísticos más generales de este texto, presentaremos aquí dos índices jerárquicos paralelos y dos selecciones también paralelas de formas pertenecientes esencialmente a inventarios abiertos. En los dos últimos casos hemos integrado, sin embargo, las formas personales y posesivas dominantes (yo me se su).

INDICES SELECTIVOS JERARQUICOS DE LOS LEXICOS COLON Y LAS CASAS

CRISTOBAL COLON			BARTOLOME DE LAS CASAS		
Formas	Occurrencias	Porcentage	Formas	Occurrencias	Porcentage
y	895	7,39	y	2.539	5,86
de	691	5,71	de	2.093	4,83
que	573	4,73	que	2.081	4,79
la	349	2,88	la	1.369	3,13
a	265	2,19	el	977	2,23
en	242	2,00	a	912	2,09
el	161	1,33	en	743	1,75
no	160	1,32	los	593	1,38
yo	137	1,13	por	509	1,15
las	127	1,05	al	498	1,15
se	124	1,02	no	453	1,04
los	114	0,94	del	437	1,01
es	113	0,93	las	393	0,92
muy	111	0,92	Almirante	380	0,86
por	104	0,86	con	354	0,80
porque	83	0,69	se	340	0,78
lo	82	0,68	muy	283	0,63
del	81	0,67	lo	266	0,63
como	80	0,66	tierra	266	0,61
isla	79	0,65	un	265	0,60
con	78	0,64	había	264	0,60
así	76	0,63	una	253	0,59
al	75	0,62	leguas	243	0,56
esta	75	0,62	porque	238	0,54
para	74	0,61	le	231	0,53
me	67	0,55	su	212	0,50
una	66	0,54	como	190	0,43
todos	63	0,52	para	181	0,42
todo	62	0,51	era	180	0,40

**INDICES JERARQUICOS DE INVENTARIOS ABIERTOS
Y FORMAS PERSONALES DOMINANTES**

CRISTOBAL COLON			BARTOLOME DE LAS CASAS		
Formas	Occurrencias	Porcentage	Formas	Occurrencias	Porcentage
yo	137	1,13	Almirante	354	0,86
<i>se</i>	124	0,93	<i>se</i>	340	0,78
es	113	0,65	tierra	249	0,61
isla	79	0,55	había	246	0,60
<i>me</i>	67	0,47	leguas	228	0,56
tierra	57	0,39	<i>su</i>	212	0,50
gente	47	0,39	era	178	0,43
son	47	0,37	noche	165	0,40
Altezas	45	0,36	dice	152	0,37
Vuestras	44	0,30	isla	151	0,37
había	36	0,28	día	138	0,34
era	34	0,27	millas	138	0,34
oro	33	0,24	gente	135	0,33
vide	29	0,22	es	132	0,32
hacer	27	0,22	puerto	120	0,29
hombres	27	0,22	diz	106	0,26
nao	27	0,21	Cabo	101	0,25
islas	26	0,21	río	101	0,25
noche	26	0,21	Leste	100	0,24
ver	26	0,21	sol	91	0,22
cosas	25	0,21	viento	91	0,22
hay	25	0,19	oro	88	0,21
cosa	23	0,18	parte	81	0,20
cabo	22	0,18	vido	80	0,20
creo	22	0,18	carabela	78	0,19
ha	22	0,18	tenía	77	0,19
haber	22	0,17	camino	75	0,18
tienen	21	0,17	indios	75	0,18
agua	20	0,16	anduvo	73	0,18
parte	19	0,16	islas	72	0,18
tengo	19	0,16	Sueste	71	0,17
tiempo	19	0,15	estaba	70	0,17
dar	18	0,15	hay	68	0,17

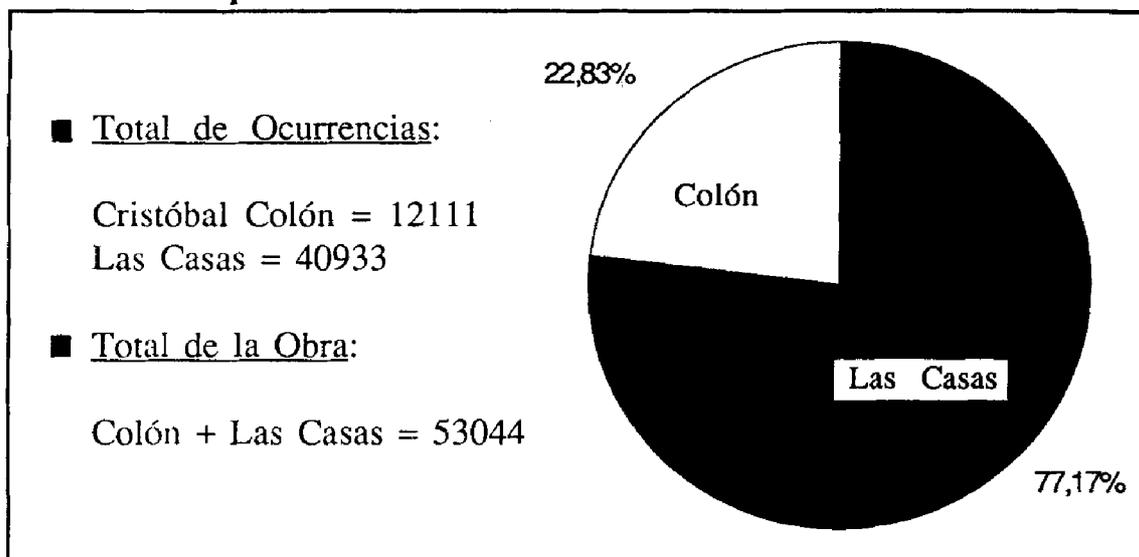
Notaremos que las formas contabilizadas aquí pueden ser formas homógrafas y por consiguiente comportan cierta ambigüedad. Si advertimos además que las citas autenticadas de Colón representan globalmente en este texto 12.111 ocurrencias, mientras que el compendio de Las Casas comporta 38.516 ocurrencias, parece imposible hacer comparaciones significativas, mientras no hayamos acabado la constitución de las Bases de Datos adecuadas. Éstas nos permitirán la desambigüisación de las formas homógrafas y el estudio de las relaciones intratextuales.

Sin embargo, si consideramos provisionalmente las formas como elementos matemáticos, podemos examinar su distribución en los diferentes fragmentos y sacar así otros tipos de informaciones.

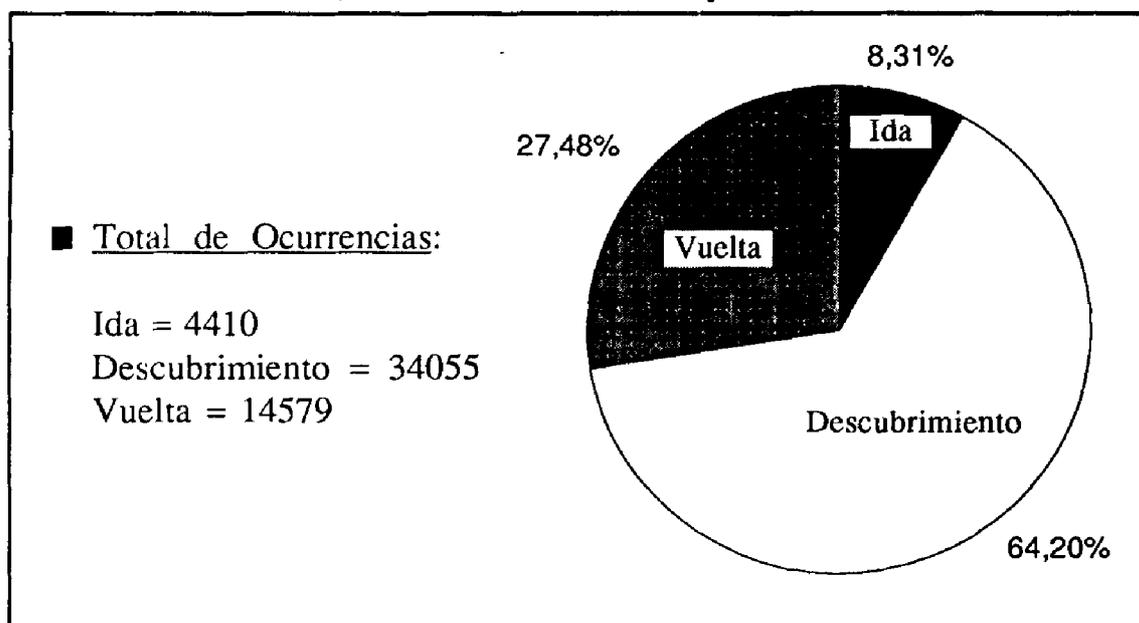
Por eso hemos hecho varias representaciones gráficas:

- De la oposición global de los dos vocabularios (Las Casas/Colón).
- De la comparación global de los tres fragmentos Ida, Descubrimiento y Vuelta.
- De la distribución de los léxicos Las Casas y Colón en los tres fragmentos Ida, Descubrimiento y Vuelta.

Comparación estadística de los dos léxicos



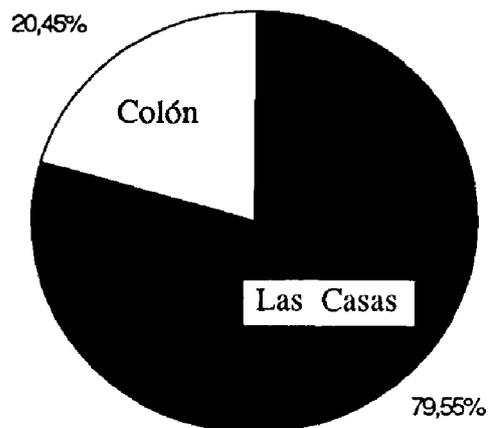
Comparación estadística del léxico global en los bloques: Ida, Descubrimiento y Vuelta



Distribución de los léxicos Colón y Las Casas en los bloques: Ida, Descubrimiento y Vuelta

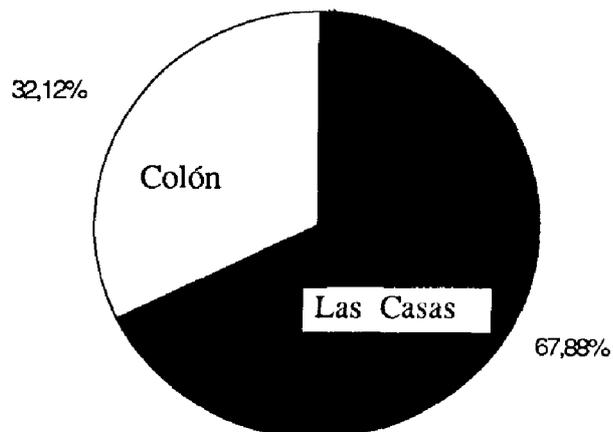
■ Ocurrencias Ida:

Colón = 902
Las Casas = 3508



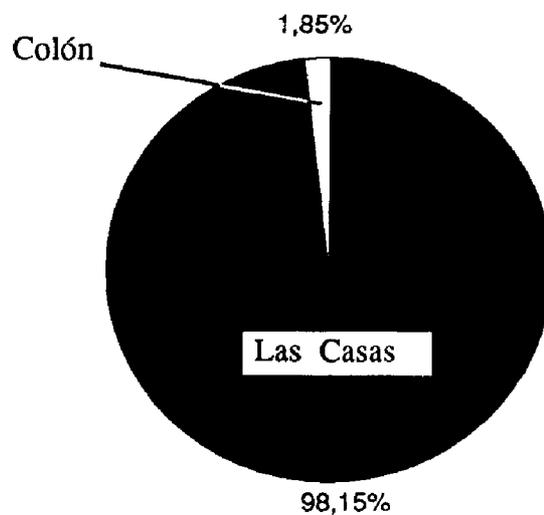
■ Ocurrencias Descubrimiento:

Colón = 10940
Las Casas = 23115



■ Ocurrencias Vuelta:

Colón = 269
Las Casas = 14310

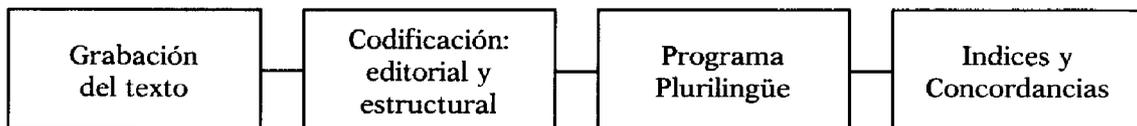


Se nota que las reparticiones de los fragmentos autenticados de Colón y de los del Compendio varían mucho a lo largo del texto. En la Vuelta parece que las citas autenticadas son muy poco numerosas. Es evidente también que la extensión textual del Descubrimiento es mucho más importante que la de la Vuelta y sobre todo que la de la Ida, aunque como ya hemos notado los tres períodos son más o menos, cronológicamente, equivalentes.

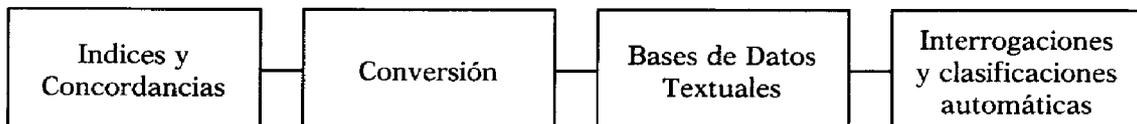
Para progresar hacia verdaderas interpretaciones, hace falta como ya lo hemos dicho, introducir en nuestros métodos, procesos de dinamización e interrogación posteriores a la generación de los listados, siempre demasiado estáticos.

Por eso, utilizando los índices y las codificaciones morfo-sintácticas, ya largamente experimentadas en nuestro Laboratorio⁴ hemos desarrollado procesos de conversión para recuperar dichos índices hacia Bases de Datos Textuales.

Anteriormente, podíamos resumir nuestro protocolo experimental de la manera siguiente:



Actualmente, insertamos una nueva etapa después de la generación de índices y concordancias que resumimos gráficamente:



Esta metodología se inscribe en un proceso de compatibilidad ascendente que permite la recuperación de los trabajos realizados anteriormente con otros tipos de adaptaciones e instrumentos. Se encamina también hacia la utilización de ordenadores de gran capacidad.

Para ilustrar este proceso daremos en la página siguiente un fragmento de concordancia y un ejemplo de inserción en una Base de Datos Textuales

Y aunque nuestra presentación metodológica queda aquí incompleta trataremos sin embargo de examinar la productividad de nuestras ayudas informáticas para la interpretación de textos de alto nivel de complejidad como éste.

Uno de los resultados más notables de nuestras primeras investigaciones textuales, en el primer viaje de Cristóbal Colón, fue sin duda la posibilidad de alcanzar lo que parece ser la escritura del Almirante a partir de un conjunto de fragmentos autenticados por Bartolomé de Las Casas.

Es evidente, por ejemplo, que la presencia exclusiva de la forma «yo» en los fragmentos autenticados de Colón, y la presencia exclusiva de la forma «Almirante» en la parte que hemos llamado Las Casas, nos permite comprobar

⁴ VERONIQUE HUYNH-ARMANET, cap. 4 «les étapes de l'Analyse», párrafo 4.3.0. «Le Code», *Des structure syntaxiques de l'espagnol à l'analyse relationnelle des textes*, París, Honoré Champion, 1977.

Las Casas	Page 17	Agosto	Día 6	Contexto Izquierdo	Forma Polo	Contexto Derecho
LC	Pg17	Aout	6	pesaba ir aquel viaje; y	dice	el Almirante que antes que p
LC	Pg17	Aout	6	rabela sin su peligro, y	dice	que alguna pena perdía con s
LC	Pg18	Aout	9	re con la Pinta adobada.	Dice	el Almirante que juraban mu
LC	Pg18	Aout	9	tro tanto con juramento.	Dice	aquí el Almirante que se acu
LC	Pg18	Aout	9	de una manera. Y también	dice	que se acuerda que lo mismo
LC	Pg20	Sep	16	aba sino oír ruiñeños.	Dice	él
LC	Pg21	Sep	17	l guardó el Almirante. Y	dice	que aquellas fueron señales
LC	Pg21	Sep	17	de la Niña mataron una.	Dice	aquí el Almirante que quell
LC	Pg21	Sep	17	En aquella mañana	dice	que vido un ave blanca que s
LC	Pg22	Sep	22	nas pardelas y otra ave.	Dice	aquí el Almirante
LC	Pg23	Sep	23	s asombraba, por lo cual	dice	aquí el Almirante
LC	Pg23	Sep	25	afirmación, el Almirante	dice	que se echó a dar gracias a
LC	Pg25	Sep	30	esto	dice	aquí el Almirante. También e
LC	Pg26	Oct	3	ía pintadas en su carta.	Dice	aquí el Almirante que no se

Inserción en la Base de Datos

Base de Datos: El Compendio del Primer Viaje de C. Colón

Referencias textuales => Pluma: LC
Página: Pg23
Mes: Sep
Día: 25

Representación de la Concordancia:

afirmación,el Almirante dice que se echó a dar gracias a

1 Désignation COLON.WDB

la calidad de la codificación que ha sido aplicada. Si advertimos además que en la parte reconocida como de Colón aparecen no sólo el «yo», sino también formas verbales tales como «digo», «dije», etc., y que en la parte opuesta dominan al contrario el «su» y las formas de tipo «dice», «dijo», «diz», parece que cada uno de los dos bloques corresponden a una lengua distinta, en el sentido informático de «lengua»⁵. Para tratar de alcanzar a Critóbal Colón a partir de

⁵ VERONIQUE HUYNH-ARMANET, Actas del VIII Congreso de la Asociación Internacional de Hispanistas, Brown University, agosto 1983, Ed. Istmo, Madrid, 1986. Cit. pág. 23: «... El adjetivo plurilingüe estriba en la definición computacional de lenguas: una **lengua**, en tal definición, no es sólo un idioma en el sentido habitual de la palabra, sino un fragmento declarado mediante una codificación adecuada como lengua distinta; podemos, por consiguiente, declarar que una novela está escrita en dos lenguas, la de la narración y la del diálogo...»

su escritura haría falta no sólo seguir con la descripción de los fragmentos que acabamos de aislar, sino también aplicar a las Cartas Columbianas, las metodologías que acabamos de esbozar y someter los datos así recogidos a estadísticas multivariadas.

Es evidente que para conocer también mejor al lector-editor Las Casas, tendríamos que examinar con muchísima atención, los temas que el autor del Compendio abandona al Almirante, los que comparte con él y los que se reserva, eventualmente.

La recolección de tales informaciones por vía automática, prepara evidentemente la constitución de diccionarios contextuales ya clásicos en las «industrias de las lenguas» y la definición de gramáticas textuales indispensables a los traductores como a los que tratan de resucitar las épocas pasadas. Parece que podría llevarnos más allá de estos objetivos tangibles y que hará falta alcanzar. Es por lo menos lo que nos han sugerido los primeros estudios experimentales que hemos emprendido en torno a la pluricontextualidad cultural de las terminología amerindias⁶ que llamamos, con cierto anacronismo, de Primera Generación. Terminologías que la mayoría de los pueblos europeos ya no reconoce como exóticas...

La capacidad de describir correctamente las alternancias de bloques en textos compositos, valiéndose de metodologías que admiten comprobaciones y correcciones, podría llevarnos a dejar de proyectarnos sobre las obras más complejas del pasado, creyendo analizarlas. Lo que nos permitiría sin duda volver a descubrirlas...

⁶ VERONIQUE HUYNH-ARMANET, *Les conquérants espagnols et les terminologies indo-américaines: Etude sur le lexique dit des Eléments Naturels dans «La Historia general y natural de las Indias» de Gonzalo Fernandez de Oviedo*. Actes du Colloque de «Terminologie es industries de la langue», Les Presses de l'Université de Montréal, Meta vol. 34, núm. 3, septembre 1989.