

## Ideas para un diccionario automatizado del español actual

Manuel Alvar

### EXCURSO HISTÓRICO

Hablar de diccionarios automatizados es situarnos en el eslabón, por hoy, último de una larga cadena que empezó hace muchos siglos. Pero este mismo proceso nos ha venido a delimitar claramente los conceptos. Hace bien poco, y como cuestión general, aunque con mirada puesta en nuestra lengua, se formuló una pregunta nada retórica: "¿Qué es un diccionario?"<sup>1</sup> y a la respuesta tendré que volver para separar conceptos que alguna vez aparecen entreverados: un diccionario no es un glosario, ni un vocabulario, ni una enciclopedia. Un diccionario es una realización tardía, a la que se ha llegado desde lejanos caminos. *In principio fuit glossa*. Y *glosa* es, lo he dicho ya, el primitivo quehacer de los lexicógrafos, el comentario o el escolio a un texto difícil<sup>2</sup>, y el *glosario* el 'catálogo de palabras oscuras o desusadas, con definición o explicación de cada una de ellas'<sup>3</sup>. Ya no nos extraña que *glosario* sea palabra del bajo latín, mientras que *diccionario* haya nacido mucho después y sólo, ayer mismo, se definiera, sin mucho rigor, como 'libro en que por orden comúnmente alfabético se contienen y explican todas las dicciones de uno o más idiomas'<sup>4</sup>.

<sup>1</sup> Manuel ALVAR EZQUERRA, en *Lingüística Española Actual* II (1980): 103-108. Del mismo autor, *Proyecto de lexicografía española*, Barcelona, 1976: 14-15.

<sup>2</sup> "Del glosario al diccionario automatizado", *Boletín de la Academia Puertorriqueña de la Lengua Española* VII (1979): 71-91.

<sup>3</sup> REAL ACADEMIA ESPAÑOLA, *Diccionario de la lengua española*, 19ª edic., Madrid, 1970, s.v. *glosario*.

<sup>4</sup> *Ibíd.*, s.v. *diccionario*.

Pero hoy nos instauramos en la necesidad de crear diccionarios<sup>5</sup>. Necesitamos diccionarios de muchas clases: normativos unos<sup>6</sup>, de arcaísmos otros<sup>7</sup>, de frecuencia<sup>8</sup> o inversos<sup>9</sup>, de técnicas<sup>10</sup> u oficios<sup>11</sup>, de una o de más de una lengua<sup>12</sup>, de palabras cruzadas<sup>13</sup> o de vocablos actuales<sup>14</sup>... Por doquier nos asaltan las preguntas y las respuestas, y no es extraño que nos sorprenda la perplejidad o la aporía. Y todo podría justificarse desde viejas perspectivas o desde nuestras propias urgencias. Pero mi campo va a acotar mucho, y aún a olvidar no poco. Quien quiera conocer los viejos caminos desde Finlandia a Italia, desde Polonia a Hispania, puede consultar el conocido trabajo de Manuel Díaz y Díaz<sup>15</sup>, pero obras como las de Goetz<sup>16</sup>, Lindsay<sup>17</sup> o Löfstedt<sup>18</sup> no son propiamente glosarios, aunque en el título de algunas de ellas pueda aparecer la palabra y, si se me permite un breve escolio, que nos hará pensar en la difícil tarea de hacer una lexicografía de tal carácter, repetiría que el latín medieval no es una lengua muerta, ni una lengua tradicional, ni una

<sup>5</sup> Para los problemas que trato, vid. M. ALVAR EZQUERRA, *Proyecto*, cit. en la nota 1.

<sup>6</sup> Tal es el carácter que tiene el *Diccionario académico*.

<sup>7</sup> Por ejemplo, los de CEJADOR (*Vocabulario medieval castellano*, Madrid, 1927), OELSCHLÄGER (*A Medieval Spanish Word List*, Madison, 1940), BOGGS (*Tentative Dictionary of Medieval Spanish*, Chapel Hill, 1946), etc. Para las realizaciones actuales, vid. M. ALVAR EZQUERRA, "Le 'Dictionary of the Old Spanish Language'" (DOSL), *Cahiers de Lexicologie XXXV* (1979): 117-132.

<sup>8</sup> Con referencia al español, vid. Ch. MULLER, "Un dictionnaire de fréquence de l'espagnol moderne", *Zeitschrift für romanische Philologie LXXXI* (1965): 476-493.

<sup>9</sup> Para el español, F.A. Stahl y G.E.A. Scavnick, *A reverse Dictionary of the Spanish Language*. Urbana-Chicago, 1975.

<sup>10</sup> Recuérdese el memorable del P. TERREROS (*Diccionario castellano con las voces de ciencias y artes y sus correspondientes en las tres lenguas francesa, latina e italiana*. Madrid, 1786-1793). Vid. Homero Seris, *Bibliografía de la lingüística española*. Bogotá, 1964, números 12801-12820.

<sup>11</sup> Por ejemplo, el *Vocabulario de oficios y profesiones*, de la Escuela Social (Madrid, 1946), el *Vocabulario de ocupaciones*, del Ministerio del Trabajo (Madrid, 1963), el *Vocabulario español de la caza*, del Ministerio de Agricultura (Madrid, 1950) o el *Léxico de alarifes de los siglos de Oro*, de F. García Salinero (Madrid, 1968). Vid. Seris, op. cit., números 12821-12964.

<sup>12</sup> ALVAR EZQUERRA, *Proyecto*, página 250.

<sup>13</sup> Véase, por ejemplo, R. EDWARDS, *The Crossword Anagram Dictionary*. Londres, 1978; Litero, *Diccionario de crucigramas*. Barcelona, 1974.

<sup>14</sup> Cfr. Pierre GILBERT, *Dictionnaire des mots contemporaines*. París, 1980.

<sup>15</sup> "Ruta crítica por la lexicografía medieval" *Helmantica XI*, 1960: 497-518).

<sup>16</sup> *Corpus glossariorum latinorum*, Leipzig-Berlin, 1888-1923.

<sup>17</sup> SEXTUS POMPEIUS FESTUS, *De verborum significatu quae sunt cum Pauli epitome*, Leipzig, 1913.

<sup>18</sup> *Philologischer Kommentar zur Peregrinatio Aetherae*, Upsala, 1911.

lengua universal de cultura: es eso y otras muchas cosas. De tal modo, al reflejarse ese mundo complejo en el espejo de su léxico, tendremos que considerarlo en lo que tiene de actitud vital para entender su propia evolución que, unas veces, llevará a las lenguas románicas, pero que otras creará una específica *latinidad medieval*, bien distinta de la clásica. Algo que, desde nuestra propia perspectiva, nos hace pensar en el áspero trabajo de hacer un diccionario de la baja latinidad<sup>19</sup>.

Y he aquí que la lejanía nos ha permitido llegar a nuestro propio esperadero. Porque situémonos en el español: las vías de su propio existir no serán distintas de las que acabo de enunciar. El español es una lengua que, en versos inolvidables de Unamuno,

... flota como el arca  
de cien pueblos contrarios y distantes,  
que las flores en ella hallaron brote  
de Suárez y Rizal, pues ella abarca  
legión de razas, lengua en que a Cervantes  
Dios le dio el Evangelio del *Quijote*<sup>20</sup>.

Pensamos que ésta pudiera ser una situación semejante a la latina, y así se ha creído y así ha dado pie a la famosa polémica entre Cuervo y Valera<sup>21</sup>, a las encontradas opiniones de Menéndez Pidal y Dámaso Alonso<sup>22</sup>. Pero no quiero entrar en la cuestión, aunque mis ideas no van hacia los pesimismos metafísicos: las cosas mucho han cambiado en casi un siglo y, a pesar de la "deslealtad lingüística"<sup>23</sup> de algunos, hay, también, la fidelidad observante de los más. Quiero seguir conduciendo el hilo de Ariadna de mi humanismo, aunque luego tenga que hablar de cuestiones hacia las que los humanistas solemos estar poco inclinados. Nos sigue valiendo la comparación latina aducida antes: he dicho que el latín había evolucionado obe-

<sup>19</sup> Ciertamente que existe, y es benemérito, pero ya es harina de otro costal el rigor técnico que presente: DU CANCE, *Glossarium ad scriptores mediae et infimae latinitalis*.

<sup>20</sup> OSCAR FERNÁNDEZ DE LA VEGA recogió una breve antología de elogios del español (Nueva York, 1977).

<sup>21</sup> Vid. PACIENCIA ONTANÓN, *La posible fragmentación del español en América. Historia de un problema*, México, 1967.

<sup>22</sup> R. MENÉNDEZ PIDAL, *La unidad del idioma*, Madrid, 1944; "Nuevo valor de la palabra hablada y la unidad del idioma", *Memoria del II Congreso de Academias de la Lengua Española*, Madrid, 1956: 487-495; DÁMASO ALONSO, "Unidad y defensa del idioma", *ibid.*: 33-48.

<sup>23</sup> Comunicación de Gregorio SALVADOR al Congreso de la Sociedad Española de Lingüística (Madrid, diciembre de 1980).

deciendo a principios vitales. Es cierto: el latín medieval no era un fósil, sino un ser vivo. Ese latín se apartaba del clásico porque la vida había cambiado, y la lengua que servía a la vida, también. Veámoslo en unos escolios léxicos: la evolución fonética había llevado la homofonía a *colus* 'rueca' y *collus* 'cuello', a *estrabo* 'moscarda' y *strabo* 'bizco', a *poto* 'beber' y *puto* 'podar'; la proximidad léxica confundió *colorium* con *color* y *colyrium*, *cucufa* con *cuculla* y *cufia*, *torribulum* con *torreo* y *turibulum*, *cetarium* con *setaceum* y *cestarium*, *endenia* con *indicta* y *nenia*; surgieron numerosos compuestos (*novinupta*, *sanguifixum*, *archiclavis*, *linipedium*), se inventaron infinitud de derivados (*fulgerium*, *gerularius*, *intitulo*) y se crearon voces extrañas, pero de vida cierta, porque se han recogido con reiteración (*cirlicitricopacinium* 'telaraña', *cirocitrociconium* 'rodezo de molino', *moglodulus* 'hojaldre', *sandapitarius* 'balletero'). Todo esto, es verdad, indica una vida fluyente, pero junto al manar irrestañable seguían vivos los arcaísmos (*aureax*, *imberbus*, *famina*, *lura* 'os utris', *subunculus*), los vulgarismos (*femarius* por *fimarius*, *gulfus* por *gonfus*, *agur* por *augurius*), los grecismos, habitualmente bíblicos o eclesiásticos (*aforus*, *apostrofo*, *anforio*, *drama* 'madeja', *inconomus*, *ornis*, *perifera*), los arabismos modernos (*almiratus*, *gabela*, *metalufum*, *napigiium*, *xarabula*), los germanismos (*baro*, *foforicium*, *cofia*, *marchio*), algunos hebraísmos... Todo este mundo nuevo, nos podemos preguntar, ¿existió fuera de las escuelas en las que los glosarios se usaban? Y la respuesta no podría ser sino ésta: la lengua recibía influjo de intereses muy variados y tuvo una vida de singular interés como instrumento universal de cultura: entonces debemos pensar que el vocabulario de una lengua se utiliza para fines enormemente complejos y distantes, no los de aclarar textos, que tal sería la pretensión de un glosario, sino para inventariar exhaustivamente el léxico de una lengua, que tal es el fin del diccionario. Y al llegar todos esos elementos sorprendentes hasta las columnas de una gran compilación lexicográfica, en ella habían alcanzado perpetuidad, por más que la vida brotara de manaderos que pudieran parecer encenegados. Y hoy, en este 1981 en que estoy escribiendo, ¿diferenciaríamos el borbotar de lo que se piensa enlojado, de lo que se sintió como limpio hontanar? Mejor, desde este latín que aquí estamos hablando, envejecido por más de mil quinientos años de historia<sup>24</sup>, ¿nos significa algo lo que no sea sino la ejecutoria que da el paso del tiempo?

<sup>24</sup> Iorgu IORDAN - Maria MANOLIU, *Manual de lingüística románica* (revisado por M. Alvar), I, Madrid, 1972: 71-74.

Pensemos en el español. Es una lengua viva, muy extendida, con mil enmarañamientos. De su origen se ha dicho que es un complejo dialectal<sup>25</sup>, de su andadura podríamos hablar de esas evoluciones fonéticas que en unos sitios llevó a confundir 'hinojo' y 'rodilla', porque tanto *feniculum* como *genuculum* hubieran tenido la misma forma<sup>26</sup>, como *ibam* y *habebam* convertidos en *ía*<sup>27</sup>, *rapida* y *rabia* en un mismo final<sup>28</sup>, y no hablemos de esas hablas vivas de hoy en que el singular de *coce(s)* es *patá*, porque el simple *có* (< c a l c e) se hubiera confundido con *codo* y con *col*<sup>29</sup>, o que se hacen las cosas *malamente*, porque de *hacerlas ma(l)* no se podría saber si se hacían *má(s)* o multitud de ellas (*la ma(r) de cosas*)<sup>30</sup>. También hoy la proximidad léxica lleva a confundir, como en los ejemplos de hace cientos y cientos de años, a cosas tan heterogéneas como *putativo* 'tenido por padre' y 'filius manceris'<sup>31</sup>, *tolete* 'estaquilla para el remo' y 'duro de mollera'<sup>32</sup>, los mil casos de homonimia<sup>33</sup>, y evito el fácil mundo de los chistes<sup>34</sup>. ¿Diríamos que es nueva la forma de hacer compuestos del tipo *putidoncella*<sup>35</sup>, *platinoche* o *baciyelmo*<sup>36</sup>? ¿La de los derivados como *madrigalizar*<sup>37</sup>, *reconstituyente* o *encarnadura*? ¿La de palabras tan sorprendentes como *guisquería*, *pincha-*

<sup>25</sup> Vicente GARCÍA DE DIEGO, "El castellano como complejo dialectal y sus dialectos internos", *Revista de Filología Española* XXXIV (1950): 107-124.

<sup>26</sup> Gerhard ROHLFS, *Estudios sobre el léxico románico* (trad. M. Alvar), Madrid, 1980: 108.

<sup>27</sup> "El imperfecto 'iba' en español", *Homenaje a F. Krüger*, I, Mendoza, 1952: 41-45.

<sup>28</sup> Ramón MENÉNDEZ PIDAL, "Modo de obrar el sustrato lingüístico", *Revista de Filología Española* XXXIV (1950): 5.

<sup>29</sup> Nota de M. ALVAR en Iorgu IORDAN, *Lingüística románica. Evolución, corrientes, métodos*. Madrid, 1967, p. 276, nota 47. *Atlas Lingüístico y etnográfico de Andalucía*, II, mapa 1612.

<sup>30</sup> *Ibid.*, VI, mapas 1620 y 1625.

<sup>31</sup> *Ibid.*, V, mapa 1347.

<sup>32</sup> Pancho GUERRA, *Léxico de Gran Canaria*, Las Palmas, 1977: 285 y 499 (por error pone *tilete*).

<sup>33</sup> IORDAN - MANOLIU, *op. cit.* en nota 17: 270 y ss.

<sup>34</sup> Henry N. BERSHAS, *Puns on Proper Names in Spanish*, Wayne University Press, 1961.

<sup>35</sup> Un soneto atribuido a QUEVEDO comienza así: "Melancólicos estáis, *putidoncella*" (R. Foulché-Delbosc, "136 Sonnets Anonymes" [Nº 59], *Revue Hispanique*, VI (1899): 328-407).

<sup>36</sup> Para *platinoche* y *baciyelmo*, véase el brillante artículo de Leo SPITZER, "Notas sintáctico-estilísticas a propósito del español 'que'", *Revista de Filología Hispánica* IV, 1942: 105-126 y 253-265. *Platinoche* fue usado por GARCÍA LORCA y *baciyelmo*, por CERVANTES.

<sup>37</sup> "*Madrigalizaré* junto a tus labios" es un verso de Rubén DARÍO en "Divagación" (*Prosas Profanas*, apud *Poesías Completas*, edic. Méndez Plancarte-Oliver Delmás, Madrid, 1968, p. 555).

*discos o paralelepípedo*?<sup>38</sup> Sí; como hace milenios, la lengua es un fluir sin descanso y en ella, también como hace milenios, duran lo que mal llamamos arcaísmos, vulgarismos y, por supuesto, las etiquetas trocadas, grecismos, galicismos o italianismos; los arabismos, anglicismos, y los hebraísmos, rusismos, *et sic et coeteris*. Nada nuevo, pero la vida no puede detenerse. Quienes hacían lexicografía en el siglo XIII o en el siglo XV debían ordenar los mismos materiales que nosotros (mudado sólo el aspecto), pero no tenían las técnicas que nosotros poseemos, y aquí van a empezar las discrepancias, aunque antes de señalarlas deba recoger un hilo que quedó suelto.

El latín medieval no era una lengua muerta, ni tradicional, ni universal, sino que, siendo todas esas cosas, era otras muchas y, siendo latín, era "otro latín". Pero esto no por ser precisamente "esa lengua", sino de manera mucho más simple y general: por ser "lengua". Todas las constituciones políticas ayudan al hombre a identificar su propio linaje, y, verdad es, todas las lenguas para su propia identidad han de reconocerse, también, en una estirpe. Y a este punto quería llegar: la lexicografía, mil veces se ha dicho, es una obra colectiva. Todos aprendemos de todos, pero al decir "todos" no me refiero al lexicógrafo como técnico o al lingüista como teorizador, sino también, al ignorado e ignorante usuario y a las desconocidas generaciones que aún han de venir. Necesitamos reconocer nuestro pasado, inventariar nuestro presente y colaborar con nuestro futuro. El latín nos ha servido mucho en unos casos, algo en otros y nada en cuanto voy a decir. Pero gracias a la historia se puede hoy hacer un diccionario; gracias a ella, aprendemos de la relatividad, pero gracias al presente damos el testimonio de nuestra propia vida. Y este testimonio quiero limitarlo a la realidad de nuestro quehacer lexicográfico.

Porque vivimos, hemos de decir que hemos vivido. Nos manifestamos a través de una lengua y ésta es nuestra posibilidad de expresión. Sin hacer historia, pero volviendo los ojos a ella, las generaciones que nos han precedido y nosotros mismos, no hemos hecho demasiado por los quehaceres lexicográficos, pero a pesar de ello, algo se puede saber del español medieval o del áureo (antiguos vocabularios de un escritor<sup>38</sup>, concordancias recientes de otros<sup>39</sup>, léxicos especiali-

<sup>38</sup> Resultan ya muy incompletos los datos bibliográficos de Gerhard ROHLFS, *Manual de filología hispánica. Guía bibliográfica, crítica y metódica*, Bogotá, 1957: 129-130. Cfr. H. SERÍS, *Bibliografía de la lingüística española*, Bogotá, ICC, 1964, números 12967-13029.

<sup>39</sup> Vid. bibliografía en ALVAR EZQUERRA, *Proyecto*, *op. cit.* en nota 1: 251. Son ya numerosas las concordancias de autores españoles. Como más recientes citaré las de este investigador sobre la *Vida de San Ildefonso* (Málaga, 1980) e

zados <sup>40</sup>); sin embargo, poco de ese mundo cambiante y vivo que es la lengua de cada día <sup>41</sup>. Padecemos necesidad de enfrentarnos con nuestra propia realidad y abrir las puertas al futuro, sin olvidar, por supuesto, que cada sincronía se genera en el pasado.

#### ACERCAMIENTO AL PROBLEMA

En estas páginas quisiera ofrecer un proyecto concreto de investigación. De su validez dependerá la aplicación general y, de cualquier modo, aspiramos a ofrecer el testimonio de una parcela del español de nuestros días y la metodología que nuestro tiempo exige. Pienso en la realidad del español peninsular, pero valdría cualquier otro ámbito de nuestra lengua en los años en que nos ha tocado vivir. Son razones clarísimas las que me obligan a la selección: geografía en la que me encuentro, testimonios que conozco, circunstancias que me cercan. Pero, y además, todos estos factores nos hacen ver el momento dinámico en que se encuentra ese fragmento lingüístico bajo el que posamos nuestra mirada. Vamos a acotar veinticinco años de historia: pienso que cinco lustros vividos hoy constituyen un período de tiempo suficientemente amplio como para llevar a cabo mi proyecto con posibilidades de terminarlo. Dispon-

“Infancia y muerte de Jesús”, *Archivo de Filología Aragonesa* XXVI-XXVII (1980): 421-460.

<sup>40</sup> Fácil sería colocar aquí un enorme fichero. Me voy a permitir unas breves referencias que completen la nota 11: W. T. BLESSUM - C. J. SIPPL, *Vocabulario de la caza*, Madrid, 1940; *Vocabulario de ocupaciones*, Madrid, 1963; W. T. BLESSUM - C. J. SIPPL, *Glossary for Medical and Health Sciences*, Nueva York, 1972; Gerhardt Christian AMSTUTZ, *Glossary of Mining Geology. English, Spanish, French and German*, Stuttgart, 1971; Karin R. BAND-KUZMANY, *Glossary of the Theatre*, Amsterdam - Nueva York, 1969; Françoise BIASS-DUCROUX, *Glossary of genetics in English, French, Spanish* [...], Amsterdam - Nueva York, 1970; W. E. CLASSON, *Elsevier's Dictionary of Library Science* [...] in six Languages, Amsterdam - Nueva York, 1973; W. E. CLASON, *Elsevier's Dictionary of Metallurgy in six Languages*, Amsterdam - Nueva York, 1967; *Elsevier's Nautical Dictionary*, Id., 1965 y otros varios de la misma editorial. Véanse, además, los que ha publicado el CONSEIL INTERNATIONAL DE LA LANGUE FRANÇAISE: *Langage medical moderne* (París, 1974), *Vocabulaire de la chasse et la venerie* (1974), *Vocab. de l'Environnement* (1976), *Vocab. de l'Océanologie* (1976), *Dictionnaire d'Agriculture* (1977), *Vocab. des sciences et techniques spatiales* (1978) y otros muchos (de hidrología y meteorología, de administración, de ecología, comercial, de radiografía, de geomorfología, de topografía, de micrografía, de astronomía, de equitación, de fauna y flora). No olvidemos la preciosa tradición multilingüe de los diccionarios náuticos (Röding, Jal, Segditsas, Ysita, etc.).

<sup>41</sup> Es útil para nuestro objeto el repertorio de *Diccionarios españoles* publicado por el INSTITUTO NACIONAL DEL LIBRO ESPAÑOL, Madrid, 1980.

mos —lo vamos a ver— de un inmenso caudal de información al que tenemos acceso y —también lo veremos— el caminar de la historia acompasa su paso a ritmos apresurados. En los años que van de 1955 a 1980 asistimos a un dinamismo trepidante: mutaciones culturales, socioeconómicas y políticas nos desazonan con su inquieto fluir. Pensemos, por ejemplo, lo que significa el acceso de las masas a cualquier nivel de instrucción, la multiplicación de centros docentes, la proliferación de universidades, la masificación estudiantil y la complejidad de los problemas que ha determinado: todo ello se cumplió en la década de los sesenta y culminó en mayo del 68. La conmoción sacudió a todos los países y, de una u otra forma, la impronta aún nos sigue marcando.

Pero pensemos, también, en todo lo que la nación se modificó al producirse cambios sociales como el de la industrialización, la mecanización agraria, el absentismo rural y la macrocefalia urbana. En 1953 iniciamos el Atlas de Andalucía; en 1964, el de Aragón: los cambios estructurales afectaron profundamente a nuestro trabajo, y muchas veces recogíamos ya historia. Pueblos en los que habían desaparecido el arado y el trillo, sustituidos por el tractor o la cosechadora, y las desapariciones arrastraron a las yuntas, a los yugos, a la siega, a la trilla... En Azanuy los hombres no me sabían armar un arado de madera, en Caparroso no quedaba ni un par de mulas. Con razón pudo escribir Caro Baroja al reseñar los primeros tomos del Atlas de Andalucía “que nadie será capaz en lo futuro de reunir unos materiales tan impresionantes [...] sobre la vida y la cultura de Andalucía”<sup>42</sup>.

Y pensemos en un último botón de muestra, lo que es pasar de la autarquía a las multinacionales, del direccionismo político a la democracia, del corporativismo vertical a los sindicatos de partido.

Es posible que tanta mutación se haya producido alguna vez al mismo ritmo que el que ahora se nos impone, pero para ningún período de tiempo poseemos tantos materiales al alcance del investigador ni tantos recursos para comprobar el pulso de la lengua<sup>43</sup>.

<sup>42</sup> Sobre Manuel ALVAR, “El Atlas lingüístico de Andalucía”, *Revista de Dialectología y Tradiciones Populares* XXI (1965): 438.

<sup>43</sup> Vid. Geoffrey KNIGHT (edit.), *Linguistics and Textual Data Processing: Computer Applications in the Liberal Arts*, Washington, 1969; David G. HAYS *Introduction to Computational Linguistics*, Nueva York, 1967; J. W. DE BAKKER (ed.), *Foundations of Computer Science*, Amsterdam, 1976-1979. En 1971 tuvo lugar en Debrecen (Hungria) un Congreso Internacional de Lingüística Computativa; sus *Papers* se editaron en La Haya, 1976.

Intentar hacer un diccionario de ese cuarto de siglo significa asentar los fundamentos del español futuro desde una etapa en la que convergen mil intereses distintos, y, por supuesto, desde la historia que hemos recibido. Sólo así —aceptando la herencia que se nos legamos— podremos entender nuestro propio tiempo. Por eso no resulta impertinente cualquier consideración del pasado, si de ella sabemos obtener enseñanzas para el presente. Y la vida del hombre es distinta en cada criatura, pero todos los seres tienen de común su propia humanidad. Y la lengua, criatura viva, será distinta hoy que hace mil quinientos años, pero los problemas universales se repiten una y otra vez en el ser histórico al que llamamos lengua, según hemos tenido ocasión de comentar.

#### LA AUTOMATIZACIÓN LINGÜÍSTICA

Mil veces se ha dicho, y de mil formas diferentes, que la lengua es el más social de los bienes humanos<sup>44</sup>, que, sin ella, no cabría ni la división del trabajo<sup>45</sup>, que en su interior se encierra —como el tesoro en el cofre— la experiencia de las generaciones que nos precedieron<sup>46</sup>. Bástenos esto. Pero si la lengua cumple todos esos fines es porque sirve como instrumento de comunicación; de otro modo no existiría. Comunicar exige un transmisor y un receptor, esos seres que hablan de una u otra forma, con uno u otro procedimiento, pero que usan un lenguaje de algún tipo<sup>47</sup>. Hablamos de lenguajes naturales y de lenguajes artificiales; el lenguaje natural es el que constituye cualquier lengua humana, principio de sociabilidad<sup>48</sup>. Este carácter social de la lengua nos obliga a inventariar todas sus posibles realizaciones; pues de otro modo difícilmente poseeríamos la perfección de tan sutil instrumento. Es necesario elaborar —desde nuestra atalaya

<sup>44</sup> Eugenio COSERIU lo ha expuesto con reiteración; vid., por ejemplo, *Sincronía, diacronía e historia*, Madrid, 1973: 70.

<sup>45</sup> Leonard BLOOMFIELD, *Lenguaje* (trad. Alma F. de Zubizarreta), Lima, 1964: 25-26.

<sup>46</sup> Véanse las ideas de Wilhelm von HUMBOLDT expuestas por IORDAN - MARNOLI, *op. cit.* en nota 17: 182-191.

<sup>47</sup> "Informática y Lingüística", *Revista Española de Lingüística* VII (1977): 192-212.

<sup>48</sup> Cfr.: Paul L. GARVIN, *Natural Languages and the Computer*, Nueva York, 1963; *Approaches to Natural Language*, edit. K. J. J. HINTIKKA *et al.*, Dordrecht-Boston, 1973; Peter KÜMMEL, *Formalization of Natural Languages*, Berlín-Nueva York, 1979; Terry WINOGRAD, *Understanding Natural Language*, Nueva York, 1972; Leonard BOLC (edit.), *Natural Language Communication with Computers*, Berlín-Nueva York, 1978.

lexicográfica— un corpus tan amplio como, idealmente, nos sea posible. Hoy no podemos pensar en cumplir tales fines por procesos manuales<sup>49</sup>: nuestro tiempo ha dado la respuesta llamada automatización<sup>50</sup>.

Enunciar las ventajas de semejante método puede resultar pueril o, cuando menos, redundante<sup>51</sup>. Si queremos salir del atolladero en que nuestra lexicografía se encuentra, hemos de recurrir, como en todas partes, al despojo automatizado de nuestro léxico<sup>52</sup>. Si así no se hace, cuando lleguemos al final seremos autoridades históricas para nuestros herederos. Si se me permite el enunciado de las ventajas<sup>53</sup> que ofrece este tipo de investigación, las resumiría en unos pocos enunciados:

1º Objetividad, por cuanto los materiales son allegados de manera exhaustiva.

2º Rapidez, porque supera a cualquier procedimiento manual.

3º Seguridad, ya que la máquina no comete más errores que los imputables a “fallos humanos”.

4º Utilidad, pues se pueden recuperar los informes para cuantos programas<sup>54</sup> necesitemos<sup>55</sup>.

Un diccionario del español realizado de este modo sería un “diccionario del español actual”, si lo lleváramos a cabo dentro de unos

<sup>49</sup> Cfr. International Federation of Automatic Control, *Multilingual Dictionary of Automatic Control Terminology*, edit. D. T. BROADBENT, Pittsburgh, 1967; M. COYAUD - N. SIOT - DECAUVILLE, *L'analyse automatique des documents*, París, 1967; Harold BORKO (edit.), *Automated Language Processing*, Nueva York, 1967; Erich BÜRGER, *Technical Dictionary of Data Processing [...]*. Oxford-Nueva York - Pergamon, 1970; Alfred WITTMANN, *Dictionary of Data Processing [...]*, Amsterdam - Nueva York, 1973. Nos interesa citar aquí la revista *Beiträge zur Linguistik und Informationsverarbeitung* (Munich, desde 1963).

<sup>50</sup> Es útil, por referirse a la lingüística y dar información sobre programación de lenguajes para los ordenadores, el libro de Raoul CHAPKIS, *Exercises in Computational Linguistics*, Amsterdam, 1970.

<sup>51</sup> “Informática y Lingüística”, *op. cit.* en nota 40; John E. HOPCROFT, *Introduction to Automata Theorie, Languages and Computation*, Addison - Wesley, 1979.

<sup>52</sup> Cfr. el libro coordinado por J. ŠTINDLOVÁ, *Les machines dans la linguistique*, Praga, 1968. Vid., también, ALVAR EZQUERRA, *Proyecto*, *op. cit.* en nota 1: 203-229.

<sup>53</sup> Ténganse en cuenta las muy importantes *Atti del Convegno internazionale d'elaborazione elettronica in lessicologia e in lessicografia*, edit. por A. ZAMPOLLI, Florencia, 1973.

<sup>54</sup> Para la voz *programme*, vid. J. GUILHAUMOU, *Lexique de l'informatique*, 2ª ed., París, 1970.

<sup>55</sup> Vid. M. F. BOTT, “Lingüística computacional”, J. LYONS, *Nuevos horizontes de la lingüística*, Madrid, 1975; “Informática y Lingüística”, *op. cit.* en nota 40: 197.

límites como los que he precisado, pero sería, además, un “diccionario vivo”, ya que estaría abierto a futuras incorporaciones. Por ejemplo, cada diez años se podrían añadir todos los cambios léxicos que hubieran afectado a la tecnología, pero, y no hay que olvidarlo, un diccionario de estas características se podría proyectar también hacia el pasado para constituir un nuevo *Diccionario histórico de la lengua española*, distinto del admirable que vamos —tan lentamente— poseyendo. Y, por último, podría extenderse sobre todas las tierras que “hablan en español”, según se fueran realizando otros proyectos parciales. Sólo el día en que, a partir del momento presente, poseyéramos el inventario de nuestra lengua según realidades diacrónicas, diatópicas, diacrónicas y sincrónicas, habríamos cumplido ese ideal, tal vez inaccesible, del diccionario panhispánico. Bien sé que a partir de la realidad contingente he saltado a la utopía, pero diré que las utopías alguna vez son realidades. Por ejemplo, el *Tresor* de la lengua francesa, con sus 250 millones de papeletas, no se imprimirá nunca<sup>56</sup>. Se va a terminar por editar lo que ya está en trance; es decir, los años de los siglos XIX y XX que constituyen el francés contemporáneo, y aún esa parcela lleva ya publicados ocho grandes volúmenes. El resto, hacia el pasado, va a ser un inmenso banco de datos<sup>57</sup>, utilizable sobre una pantalla de televisión activada desde nuestra mesa de trabajo<sup>58</sup>. Sé que muchos nostálgicos, y acaso también yo, volverán los ojos a unos años en los que todavía vivieron. Es comprensible, pero, ¿no hubo gentes que creyeron que el cine mudo sobreviviría o que el de color no lograría imponerse? Algún egregio romanista se me lamentaba del microfilm y de la xerocopia: la investigación había perdido el encanto de los viajes y la emoción de tener un libro viejo entre los dedos, pero ¿ese mismo romanista no me pidió una xerocopia de la *Toponimia árabe* de Asín? Todos hemos leído el réquiem por el libro: la era de Gutenberg agoniza y las microfichas y las pantallas acabarán con el libro. No merece la pena hacer de zahoríes: el futuro está ahí y nosotros somos quienes he-

<sup>56</sup> Tomo información de Josse DE KOCK, *Introducción a la lingüística automática en las lenguas románicas*, Madrid, 1974: 57.

<sup>57</sup> Cfr., para el francés, *La banque des mots*, París, desde 1971.

<sup>58</sup> Por otra parte, y bajo la dirección de Bernard QUEMADA, se publican los *Matériaux pour l'histoire du vocabulaire français*, cuyo volumen más reciente, el 18, data de 1980; el primero vio la luz en 1949. Se trata de dos series, la primera de tres volúmenes (*abacot - cytotropismo* (1959-1965)) y la segunda, por hoy, de 18. En cada tomo hay unas mil documentaciones que, luego, se reúnen en unas “tables cumulatives” (alfabéticas, cronológicas, metódicas, inversas) que agruparán el contenido de los tomos aislados; se dispone de las primeras, que comprenden los volúmenes 1-15 de la segunda serie.

mos de intentar crearlo. Pretender hacer proyectos para dentro de siglos no deja de ser también una utopía: los materiales los necesitamos *nunc et hic*. Aquí y ahora son términos relativos, pero su relatividad es, justamente, lo más opuesto a un ensueño.

#### LÍMITES DEL PROYECTO

El proyecto, pues, tiene que tener unos límites. O precisando mejor, el diccionario debe plantearse dentro de unos límites. Por eso he hablado de hacer el inventario de veinticinco años decisivos, sin que las puertas queden cerradas a cualquier oreo. Veinticinco años en los que la lengua no sólo ha ido a remolque de las realidades socio-culturales, sino que ha servido de instrumento para crear nuevas formas de humor o de invención de léxicos que van fluyendo hacia niveles superiores y que cuentan con anuencia y proyección en los medios de comunicación de masas. Pero es que hippies, pasotas<sup>59</sup> o lo que sean ¿no representan hablas jergales como las de cualquier otra época? Y favorecido el intercambio hoy, por las posibilidades de la técnica y la disposición de la sociedad a explicar cualquier fenómeno de marginación. A caballo entre lo que antes dije (democratización) y lo que ahora apunto (pérdida de valores tradicionales) hay un fenómeno harto importante y que no sólo afecta a nuestra lengua: me refiero al desprestigio que padecen las normas de cortesía y, con ellas, ciertas formas de tratamiento (hecho morfosintáctico y léxico), el arraigo de un vocabulario habitualmente proscrito y el desdén por otro prestigiado. Y todo esto, empezado antes, ha precipitado en menos de un cuarto de siglo<sup>60</sup>.

Al hablar de la creación de un diccionario del español actual no debemos pensar sólo en realizar un tesoro lexicográfico, y ya sería un servicio impagable el que con ello rendiríamos a nuestra lengua, sino que pretendemos algo mucho más importante: crear un método científico para poder cumplir tareas más ambiciosas. Pensemos qué podrá ser un índice de frecuencias para poder seguir los derroteros del español; o denunciar la presencia de los anglicismos, fantasma

<sup>59</sup> Como muestras recientes del quehacer lexicográfico que estas actitudes han condicionado, vid. Jaime MARÍN, *Diccionario de expresiones malsonantes del español*, 2ª edic., Madrid, 1979; Juan VILLARÍN, *Diccionario de argot*, Madrid, 1979; VÍCTOR LEÓN, *Diccionario de argot español*, Madrid, 1980; Felipe NAVARRO y Julen SORDO, *Diccionario del pasota*, Barcelona, 1979.

<sup>60</sup> Pensemos lo que significa, a pesar de su limitación, el utilísimo *Dictionnaire des mots contemporaines*, de Pierre GILBERT (París, 1980).

que tanto nos aterra; o el aplebeyamiento del habla como reflejo sutil de lo que en la sociedad va cambiando; o la aparición de infinitos metalenguajes con los que cada día se enfrenta la perplejidad del pobre hablante que no entiende la jerga de los sociólogos, de los políticos, de los economistas, de los banqueros y, desgraciadamente, tampoco la de los lingüistas y de los críticos literarios <sup>61</sup>.

Quiero insistir: el proyecto debe tener unos planteamientos rigurosos y concretos <sup>62</sup>. La especulación es necesaria para que la práctica no sea fruto del diletantismo, pero no me interesa especular en abstracto, sino disponer de unos medios que, teóricamente, faciliten apoyos al proyecto, para que desde la realidad puedan levantarse nuevas teorías y sepamos cuáles deben ser las medidas que ayuden a resolver los problemas todavía inéditos. Porque esos materiales precisos y concretos no sólo nos deben dar "un" diccionario, sino muchos diccionarios de los que apenas si poseemos algunos: diccionarios inversos <sup>63</sup>, de palabras largas y cortas <sup>64</sup>, de compuestos y derivados <sup>65</sup>, de estructuras <sup>66</sup>, de . . . Es fácil seguir acumulando posibilidades, y aún así no se habrán cegado todas las vías para que la vida vaya entrando sin descanso en las tareas del proyecto, ni se habrá remansado la historia en su quietud. El diccionario abierto y la programación informática nos brindan posibilidades que hubiéramos creído de fecundidad inalcanzable.

<sup>61</sup> La insolidaridad de muchas de estas tendencias se intenta salvar en obras como la de Rose NASH, *Multilingual Lexicon of Linguistics and Philologie*, Coral Gables, Fla., 1968; Maxim NEWMARK, *Dictionary of Science and Technology in English-French-German-Spanish, containing 10.000 Currents Terms* [. . .], Nueva York, 1963; Zoltán PIPICS, *Dictionarium bibliothecarii practicum (Ad usum internationalem in XX linguis)*, Munich, 1969.

<sup>62</sup> Cfr. Bernard QUEMADA, "L'automatisation de la recherche lexicologique: état actuel des tendances nouvelles", *Meta* I (1973): 10-138. Vid. también, Jitka ŠTINDLOVÁ (redactor científico), *Colloque International sur la mécanisation et l'automation des recherches linguistiques*. Praga, 1966 (las actas se publicaron en La Haya, 1968).

<sup>63</sup> Vid. antes, nota. Añádase: Jitka ŠTINDLOVÁ, "Les dictionnaires inverses", *Cahiers de lexicologie* II (1960): 79-86.

<sup>64</sup> Pienso lo que podrían significar, por ejemplo, para la estructura de la palabra.

<sup>65</sup> Desde un punto de vista metodológico, interesa el *Russian Derivational Dictionary*, de D. S. WORT, A. S. KOZAK y D. B. JOHNSON, Nueva York, 1970.

<sup>66</sup> Cfr. H. COTTEZ, *Dictionnaire des structures du vocabulaire savant*, París, 1980.

## ANTECEDENTES, ESTUDIOS Y EXPERIENCIAS

He llegado a esta situación desde unas tareas lexicográficas, pero no como diccionarista, sino como dialectólogo o geógrafo lingüista. He trabajado y trabajo en el *Atlas Linguarum Europae*<sup>67</sup>, automatizado según los principios del profesor Wolfgang Putschke, de Marburgo<sup>68</sup>, pero en él, yo no he tenido acceso sino a una etapa puramente especulativa y me vi obligado a tentar mi propio método. Durante muchos meses colaboré con técnicos del Centro de Cálculo de la Universidad Complutense de Madrid. Evidentemente, lo que tratábamos de encontrar era un procedimiento de automatización de la geografía lingüística, pero, desde los tiempos de Gilliéron, la geografía lingüística trata de devolver su dignidad a la palabra<sup>69</sup> y así, buscando la utilización del *plotter*<sup>70</sup>, había llegado a ordenar un diccionario. El camino fue más complejo que las pocas palabras que acabo de escribir, pero a ellas se pueden reducir. Con Manuel Verdejo redacté un trabajo en 1977, que sólo vio la luz en 1980<sup>71</sup>. La ventaja de este estudio se basaba en una realidad previa: pretendíamos automatizar el Atlas de Andalucía, que ya estaba confeccionado por procedimientos manuales<sup>72</sup>. Ibamos, es cierto, a descubrir el Mediterráneo, pero sin el Mediterráneo anterior nuestras singladuras hubieran sido de incierto navegar. La conclusión a la que debíamos llegar era ésta: si el nuevo método era válido, deberíamos obtener la misma información que por medios tradicionales. Y éste era el servicio impagable de la vieja cartografía: nos servía de comproba-

<sup>67</sup> Cito bibliografía y trabajos realizados en "Ordenadores y geografía lingüística. El proyecto del Atlas plurilingüe de Europa", *Revista de la Universidad Complutense* XXV (1976): 79.

<sup>68</sup> "Ueber ein Computerprogramm zur Herstellung von Sprachkarten", *Germanistische Linguistik* I (1969): 45-114, y "Planung einer Projektdurchführungs Automatische Kartierung des Atlas Linguarum Europae", *ib.*, VII (1972): 545-577.

<sup>69</sup> Vid., por ejemplo, Albert DAUZAT, *La géographie linguistique*, París, 1948: 32, *passim*, y, sobre todo, Karl JABERG, *Aspects géographiques du langage*, París, 1936: 14.

<sup>70</sup> Cfr. "Ordenadores y geografía lingüística": 84. Véase también Julio FERNÁNDEZ-SEVILLA, "Ordenadores electrónicos y atlas lingüísticos", *Revista de la Universidad Complutense* XXV (1976): 87-100; Suzanne H. PETERSEN, "Representación cartográfica de datos complejos mediante ordenador" (en la misma publicación, 207-216).

<sup>71</sup> "Automatización de Atlas Lingüísticos", *Revista de Dialectología y Tradiciones Populares* XXXIV (1978 [1980]): 23-48. Vid. también la bibliografía y antecedentes que aduzco en "Ordenadores y geografía lingüística": 80-81.

<sup>72</sup> Cfr. FERNÁNDEZ-SEVILLA, "Ordenadores electrónicos y atlas lingüísticos", ya citados: 95.

ción en cada uno de los pasos que estábamos dando. Llegamos al final del trabajo y los resultados podían utilizarse con seguridad: fue entonces cuando llevé a cabo la automatización del Atlas de Santander<sup>73</sup>, según he descrito en un trabajo reciente<sup>74</sup> y, al poseer la codificación de todos esos materiales, se ha procedido a la redacción de unos cuantos índices<sup>75</sup>, de entre los muchos que pude haber seleccionado. Pero aquí quiero traer a consideración los de palabras y terminaciones.

Un atlas lingüístico es un mundo inmenso que suele arredrar a los diccionaristas, según he demostrado no hace mucho<sup>76</sup>. Tenemos que poner al alcance de todos unos materiales tan cuantiosos como jamás se hubiera podido imaginar. Intentar un procedimiento manual es factible, claro, pero de incierto acabamiento. Pondré un ejemplo: en el Atlas de Andalucía transcribimos más de medio millón de formas. El índice manual ha costado unos ocho años de trabajo de los colaboradores de mi cátedra (podemos calcular un término medio de dos ayudantes por año). Ocho años, y prescindo de costos, previsibles errores, etc., para disponer sólo de un índice de palabras; es decir, de una especie de diccionario parcial del dialecto andaluz. Pero quinientas mil formas constituyen un conjunto sobrecogedor. ¿Y los cientos de miles de Canarias, de Aragón, de Navarra, de Rioja, de Santander, de los marineros peninsulares? ¿Y el Atlas en marcha de España y Portugal? ¿Y si hacemos el de América?<sup>77</sup>. Pienso que nuestros archivos guardan, por lo menos, cuatro millones de palabras, que duermen esperando la mano que sepa hacerlas vibrar. Esto es lo que me llevó a pensar en un *Diccionario del español actual*, porque lo que hacemos en dialectología es transcribir lengua viva, en un ámbito geográfico limitado, con un cuestionario restringido y con unos límites cronológicos muy precisos (los materiales de todas estas obras se han recogido en períodos de ocho o diez años)<sup>78</sup>. Ya

<sup>73</sup> Manuel ALVAR y María Pilar NUÑO, "Un ejemplo de atlas automatizado: el ALES", *Lingüística Española Actual* III (1981).

<sup>74</sup> "El atlas lingüístico y etnográfico de la provincia de Santander (España)", *Revista de Filología Española* LIX (1977 [1980]): 81-118.

<sup>75</sup> Manuel ALVAR y Carola MARTÍNEZ, *Automatización de índices en los atlas lingüísticos* (en prensa, en el Homenaje a Luis Flórez, Bogotá).

<sup>76</sup> "Atlas lingüístico y diccionario". (Leí unas pocas páginas en el Congreso Internacional de Hispanistas; el trabajo completo aparecerá en *Lingüística Española Actual* IV, 1982).

<sup>77</sup> Cfr. Manuel ALVAR, "Hacia una geografía lingüística de América", *Perspectivas de la investigación lingüística en Hispanoamérica*, México, 1980: 79-92.

<sup>78</sup> Para cuestiones referentes al mundo románico, vid. Sever POPE, *La dialectologie I*, Gembloux, 1950.

no quedaba sino dar el salto: ampliar la geografía, no limitarnos a la vida rural, pero mantener un tiempo abarcable.

#### PREPARACIÓN DE MATERIALES

He dicho que no hablo de utopías, aunque las utopías se alcancen desde la contingencia. Por eso creo en la realización del trabajo y creo que no se podrá conseguir en fases aisladas, sino en procesos de elaboración, que deberán coexistir con frecuencia. He aquí cómo pienso que debiéramos proceder:

1º Selección de los materiales para ordenar el *corpus*<sup>79</sup>.

2º Preparación de los primeros *documentos*<sup>80</sup> por un equipo. Si se decide la incorporación del material de manera exhaustiva, podría disponerse en el *soporte*<sup>81</sup> elegido para que un equipo de lingüistas procediera a una primera selección.

3º Cada tres meses se podría entregar el material para su selección.

4º Al terminar el segundo año de estas tareas podría procederse a la redacción.

Pienso que  
de palabras  
de esto es

#### COMENTARIOS SOBRE LOS TRABAJOS QUE DEBEN REALIZARSE

Los extremos anteriores necesitan algún comentario. En primer lugar la *s e l e c c i ó n* de los materiales. De acuerdo con lo que en otros sitios se ha hecho<sup>82</sup> y con lo que parece lógico en razón de lo que es nuestra propia realidad, propondría las siguientes proposiciones:

<sup>79</sup> 'Conjunto de enunciados escritos o grabados de los que se sirve la descripción lingüística' (apud Georges MOUNIN, *Dictionnaire de la linguistique*, París, 1974, s.v.).

<sup>80</sup> Vid. lo que digo en "Informática y lingüística", *op. cit.* en nota 40: 199. Allí puede encontrarse bibliografía.

<sup>81</sup> 'Soporte material sobre el que están registrados los datos (tarjetas perforadas, cintas magnetofónicas, bandas perforadas, etc.)', GUILHAUMOU, *op. cit.* en nota 54, s.v. *support*.

<sup>82</sup> Mis ideas y mis informaciones tienden a la realización de un diccionario, digamos, autónomo; mientras que las expuestas por ALVAR EZQUERRA insertaban el diccionario del español actual dentro de un gran *Tesoro* de nuestra lengua (*Proyecto*, *op. cit.* en nota 1: 150-169). Las pretensiones son distintas.

Habla viva	10%
Radio y Televisión	10%
Textos literarios	10%
Prensa .	40%
Textos didácticos	5%
Oratoria . .	5%

Queda, como es lógico, una cantidad flotante (el 20%) para atender a las necesidades que vayan surgiendo a lo largo de la obra y que, antes de empezar, son imprevisibles.

Con esta selección se podría elaborar un *corpus* de 40 ó 50 millones de palabras, que podría descomponerse así:

*H a b l a v i v a*. Pensando en la realidad de España, seleccionarían 42 de las 50 capitales de provincia; en cada una de ellas podrían grabarse 10 horas; el total sería unos tres millones de palabras, según los cálculos hechos sobre los materiales del habla de Madrid<sup>83</sup>. Estos materiales (unas 280 horas) podrían ser utilizados sin esfuerzos excesivos y, habida cuenta del conglomerado social que es la capital, añadiríamos otros dos millones de palabras transcritas. Así, pues, el habla viva nos suministraría cinco millones de palabras.

Evidentemente, no figuran todas las capitales españolas porque evito los problemas de bilingüismo, aunque utilizaría alguna de ellas para saber cuál es la modalidad castellana que allí se ha fraguado. (Por ejemplo, de las cuatro provincias catalanas, elegiría sólo una).

Además, si la recogida de materiales se hace poniendo un límite, las grabaciones no podrán cubrir los 25 años del proyecto, sino sólo los que sirvan de referencia al comienzo y al final de las encuestas.

*R a d i o y t e l e v i s i ó n*. Se puede grabar a partir de la constitución de los equipos de trabajo<sup>84</sup>, aunque podemos disponer de muchísimos materiales archivados en las emisoras. El conjunto de estos materiales nos daría otros cinco millones de palabras.

<sup>83</sup> Los informes léxicos se encuentran en estos momentos en la etapa de lematización. Su publicación está prevista para 1983.

<sup>84</sup> Cfr. R. SCHREYER *et al.*, *Dictionary of Photography and Cinematography*, Londres - Nueva York, 1961; Cécile GITEAU, *Dictionnaire des arts du spectacle* [...], *théâtre, cinéma, cirque, danse, radio, marionettes, télévision, documentologie*, París, 1970; Emilio NOVOA, *Terminologie usuelle dans les sciences et les techniques des télécommunications*, 3ª edic., Madrid, 1967; *Vocabulaire de la radiodiffusion*, del CONSEIL INTERNATIONAL DE LA LANGUE FRANÇAISE, París, 1972; *Vocabulaire de la publicité*, id., 1976; *Lexique photo-cinéma*, id., 1976.

*Textos literarios.* Pienso que podríamos seleccionar cuatro obras de las que se editaron en cada uno de esos veinticinco años. La selección podría hacerse de acuerdo con unos criterios fijos: de una parte, traducciones; de otra, obras de creación. El primer grupo no debería ser sino la mitad del segundo. La proporción puede parecer grande, pero de otro modo apenas si iban a significar gran cosa, teniendo en cuenta el amplio espectro (novela, teatro, ensayo, etc.) que debe cubrirse. Además, si el traductor cumple dignamente con su oficio, sus versiones sirven también como representación del estilo literario. Pienso, también, que las traducciones, de merecer tal nombre, son un camino importante por el que accede a nuestra lengua un caudal léxico nada desdeñable o ayudan a renovaciones parciales del vocabulario: pensemos, por ejemplo, lo que ha significado la revolución que llamamos *ensayo*<sup>85</sup>. Es lógico que buscáramos traducciones de lenguas distintas para comprobar, si las hubiera, diferencias en los criterios de esos traslados y, es lógico también, que drama, comedia, humor, costumbrismo, etc., tuvieran su propia representación. Mis cálculos hacen que estos despojos asciendan a un millón y medio de palabras.

En cuanto a las obras de creación, las proporciones de los inventarios serían: 45% de *n o v e l a* (porque es la literatura más leída y, por tanto, con mayor trascendencia); 15% de *c u e n t o s*; 15% del ancho espectro al que llamamos *e n s a y o* y en el que se cobija un mundo muy complejo de posibilidades; 10% de *r e l a t o s d e v i a j e s*, forma muy específica de hacer literatura viva, directa, pero con menor incidencia social que las anteriores; 10% de *t e a t r o* y 5% de *p o e s í a*. Ni qué decir tiene cuán necesaria considero la variedad en el carácter de los textos que puedan escogerse. El conjunto de esta información rebasaría los tres millones de palabras.

*La prensa.* Evidentemente constituye un mundo de heterogénea variedad. No podemos silenciar su complejidad: cabeceras<sup>86</sup>, editoriales, artículos de fondo, información política, deportiva, taurina, secciones financieras, religiosas, laborables<sup>87</sup>... Además debemos fijar el tiempo de nuestra realización y los límites que hemos

<sup>85</sup> "Historia de la palabra 'ensayo' en español", *Ensayo. Reunión de Málaga de 1977*, Málaga, 1980: 11-43.

<sup>86</sup> Cfr. la tesis doctoral (1977) de José Antonio MAYORAL, *El énfasis en la prensa diaria: Estudio lingüístico de los titulares de prensa*, Universidad Complutense de Madrid.

<sup>87</sup> Cfr. Graciela REYES, *La lengua de la prensa*, tesis doctoral (1979), dos tomos, Universidad Complutense de Madrid.

impuesto a nuestro *corpus*. Pensando en esos 40-50 millones de palabras, seleccionaría para cada año un periódico-diario entre los diez de más tirada; la elección puede hacerse por sorteo o, tal vez fuera mejor, estableciendo una rotación (con ella la proporcionalidad de cada rotativo sería la de una vez cada dos años y medio). Si pensamos en que cada periódico (tipo *Le Monde*)<sup>88</sup> tiene medio millón de palabras, conseguiríamos una selección de doce millones y medio de formas, o podríamos aumentar la frecuencia de cada periódico hasta conseguir el total.

Lugar aparte merecen las *revistas*. No podemos atender sino a las de tipo general, pues representan una lengua más uniformada. Y no cabría otra cosa que elegir dos cada cinco años, con lo que los diez ejemplares facilitarían otros diez millones de palabras. Total: veintidós millones y medio.

En este apartado hay que hacer referencia a las *agencias de noticias*. Su transcendencia, para bien y para mal, es inmensa. Una de las más importantes del mundo es española; transcribe trescientas mil palabras cada día y su responsabilidad social es incalculable: de una u otra forma se vinculan a ella todas las emisoras de España y América y todos, sin excepción, los grandes rotativos del Nuevo Mundo. Si sus despachos se tradujeran mal, el daño a nuestra lengua sería irreparable; si se traducen bien, los beneficios serán impagables. En estos momentos sólo podría hablar de voluntad de acierto: ahí está la reciente comisión del "español de urgencia". No hago un apartado con estos despachos porque quedan incorporados al cuerpo de cada periódico, pero no se puede silenciar esta importantísima vía de modificación del léxico<sup>89</sup>.

*Textos didácticos*. Su importancia está en el modo de conformar el mundo cultural de millones de hablantes. No podemos ignorar que todos hemos estudiado segunda enseñanza y aunque nuestra disciplina esté muy lejos de la física, de la fisiología, de la química, de las matemáticas, en el fondo de nuestra conciencia ha quedado un poso de lo que aprendimos, lo mismo que sedimentó

<sup>88</sup> Cálculos personales, hechos en 1976.

<sup>89</sup> La comisión, formada por un grupo de lingüistas, tiene cinco académicos supervisores (tres técnicos, F. Lázaro, A. Tovar, M. Alvar; un creador, L. Rosales, y el secretario de la Comisión Permanente de Academias Hispanoamericanas en la Española, J. A. León Rey, de Colombia). Aunque se venía trabajando desde tiempo atrás, la presentación pública de las tareas se hizo el 16 de febrero de 1981, en una solemne sesión (Instituto Iberoamericano de Cooperación).

un trasfondo humanístico en quienes se dedicaron a ciencias experimentales. Con pocas palabras: el español de cada época se forma también con los textos de segunda enseñanza, que conforman nuestra propia peculiaridad cultural. Por ello elegiría diez obras científicas (matemáticas, física y química, ciencias naturales) por la heterogeneidad de campos en que se cultivan; cinco de geografía e historia; diez, de lengua y literatura (me refiero a historia y crítica de la literatura en nivel secundario), que deben ser distintas de las que incluyo en el apartado de ensayo. He hecho cálculo sobre textos de unas 250 páginas (40 líneas por página y diez palabras por línea) y los datos que alcanzo son 100.000 palabras por texto. Total: dos millones y medio.

#### ENTRADAS DEL DICCIONARIO

La cuestión ha suscitado numerosos problemas críticos<sup>90</sup>, pero descendiendo a un nivel práctico, de modo que mi exposición pueda quedar completa, he pensado que las entradas de la obra se deben formular de acuerdo con los siguientes principios:

1º Por palabras aisladas.

2º Por lexías<sup>91</sup> o sintagmas<sup>92</sup>; es decir, tantas referencias como elementos significativos los integren.

3º Por aparición cronológica, para conocer cuál es la marcha de la lengua y poder atisbar cuál será su caminar en el futuro. No nos dejemos vencer por el entusiasmo, sino por la morigeración: la lingüística del siglo XIX anheló saber el destino de las lenguas; no pretendamos tanto: que nos baste con intuir lo que pueda ser el porvenir con las limitaciones que, por ejemplo, ha establecido la fonología, por más que estemos en un campo diferente<sup>93</sup>.

4º Por niveles de lengua (prensa, radio, novela, conversación, etc.), ya que cada uno de ellos tiene distinta manera de manifestarse.

5º Por índices de terminaciones (sufijos, desinencias, etc.).

<sup>90</sup> Bibliografía en ALVAR EZQUERRA, *Proyecto, op. cit.* en nota 1: 254 y 172-177.

<sup>91</sup> Término utilizado por Bernard POTTIER para designar unidades superficiales de léxico. Se trata de las entradas del diccionario que comprenden los lemas y sus derivados y compuestos.

<sup>92</sup> 'Asociación de varias unidades susceptibles de aparecer en otros enunciados'. El término procede de SAUSSURE (cfr. Oswald DUCROT - Tzvetan TODOROV, *Dictionnaire encyclopédique des sciences du langage*, París, 1972: 139).

<sup>93</sup> Vid. los datos que aporto en "Informática y Lingüística", *op. cit.* en nota 40: 205.

6º Por doble ordenación en las formas verbales (por sí y en función del verbo al que pertenecen).

7º Por frecuencia numérica, que nos dará el registro de nuestras palabras y la proporción de los extranjerismos, menor en la lengua coloquial de lo que habitualmente nos hace creer la ocasional aparición en determinados metalenguajes.

#### CONCLUSIONES

La lexicografía hispánica ha estado condicionada por la actividad académica, y a ello tenemos que referirnos siempre. Cualquier obra de hombres es susceptible de mejorar, y no voy a ser yo quien diga qué podría hacerse para perfeccionar nuestro mejor diccionario. Sí quiero declarar mis dudas de que se pueda hacer mejor que lo que desde el siglo XVIII viene haciendo la Academia Española. Pero ésta es una cuestión y otra la que acabo de suscitar: un diccionario normativo tiene determinadas exigencias, distintas de las que plantea un diccionario de uso. Más aún, de uso en un momento determinado, con los valores que la frecuencia debe reconocer, con las exigencias de una vida trepidante, con los ojos puestos no en el pasado, sino en el porvenir de la lengua. El inmenso mundo al que pertenecemos no necesita de anarquías; anda falto de obras solidarias. Y eso es lo que he querido plantear. Si nuestra lexicografía ha de progresar, lo hará con las técnicas que la ciencia de hoy pone a nuestra disposición. Si queremos contar, y que con nosotros científicamente cuenten, no basta con decir que nuestros mayores trabajaron bien, sino que nosotros hemos de poner las manos en la manera y hacer que el arado hienda la tierra y la besana impida que los surcos se tuerzan. Somos nosotros quienes escribimos el futuro desde nuestra realidad de hoy. Como un día otros antes que nosotros supieron respetar el pasado, pero trabajaron para el presente en que ahora vivimos.

UNIVERSIDAD COMPLUTENSE  
Madrid