

La lexicografía y la lista Swadesh: ¿Un método efectivo para establecer un vínculo genealógico entre lenguas?

Elber Aguilar Porras

Escuela de Literatura y Ciencias del Lenguaje
Universidad Nacional
Costa Rica

Resumen

La lexicoestadística, la cual permite analizar y comparar diferentes lenguas en estudio, reconociendo similitudes entre sí y estableciendo su vínculo genealógico, ha sido aplicada a diferentes análisis de lenguas con el propósito de discernir las cuestiones que atañen a la correcta calificación. Haciendo uso de la lista de vocabulario básico propuesta por Swadesh, este método ha ganado credibilidad pero, así mismo, críticas de varias teorías de estudio. Por tal razón, en este trabajo, se evalúa la efectividad de dicha metodología al comparar el catalán con otras lenguas tales como el portugués, el francés y el occitano, tratando de ubicar a esta lengua en uno de los siguientes grupos: iberorromance o galorromance. Entre otras conclusiones, los resultados muestran que el catalán se relaciona con el francés o el occitano a nivel de microfilo.

Palabras clave: Lexicoestadística, Swadesh, cognados, galorromance, iberorromance

Abstract

As a comparative approach, the lexicostatistic method has been employed to analyse and compare divergent under-study languages worldwide. This methodology has been able to recognize genealogical links among languages, which has allowed linguists to place them into a

diachronically given group. Getting the Swadesh list as a starting point, the lexicostatistic- as a method- has been both recognized but also criticized by several theories of study. Thus, the main purpose of this article is to evaluate the effectiveness of such approach by comparing the Catalan language with other languages such as Portuguese, French and Occitan as a means to establish a clear belonging to the Iberorromance or Galorromance groups. Among other conclusions, the results show microphyll level of relation between Catalan and French or Occitan.

Keywords: Lexicostatistic method, Swadesh, cognates, Galorromance, Iberorromance

Introducción

A mediados del siglo XX, el lingüista estadounidense-mexicano Morris Swadesh propuso una lista tanto de 100 como de 200 palabras que, en su opinión, eran representativas del vocabulario básico de todas las lenguas. A partir de este vocabulario se pueden comparar lenguas distintas relacionándolas entre sí genealógicamente.

Dichas listas de palabras son utilizadas en la lexicoestadística así como en la glotocronología. Para propósitos de este trabajo, la lexicoestadística será el procedimiento utilizado con el fin de determinar si dicha lista de palabras y dicho método utilizado por la lexicoestadística es del todo confiable al concluir si el catalán está más emparentado genealógicamente con el español y el portugués (grupo iberorromance) o con el francés y el occitano (grupo galorromance), ambos pertenecientes a la familia románica de las lenguas indoeuropeas.

Como es conocido, esta relación ha sido tratada ampliamente por varios investigadores, quienes parecen tener tanto criterios como resultados discordantes al respecto; dichas posturas y resultados serán tratados más a fondo en las sesiones siguientes de este trabajo. Según los objetivos de la lexicoestadística, cuya meta es relacionar las lenguas, y por los objetivos de este trabajo, se proporcionará un aporte importante a las divergentes discusiones que los investigadores han tenido por largo tiempo en cuanto a la relación genealógica y el parentesco de dichas lenguas. Así, el claro objetivo principal y general de este estudio es evaluar la utilidad de la lista de palabras expuesta por Swadesh y la metodología empleada por la lexicoestadística al discernir el parentesco que existe entre el catalán con uno de los grupos al cual pertenecen las demás lenguas que trataremos de relacionar desde una perspectiva genealógica; en síntesis, el agrupamiento del catalán con el subgrupo galorrománico o iberorrománico, a partir de

la aplicación de dicho método y del análisis etimológico de las “listas de Swadesh” correspondientes al francés, occitano, español, portugués y catalán, sería el objetivo que se persigue en este trabajo.

Estado de la cuestión

En este apartado, se presenta una categorización de las lenguas en estudio donde se describe, de manera muy general, la ubicación geográfica donde se habla cada una de ellas así como información etimológica de las mismas. Las lenguas son presentadas en orden de relevancia – según este estudio-, de relación histórica y en relación con su aparente parentesco – según los autores citados en cada una de ellas-.

Categorización de las lenguas

El catalán

El catalán, lengua romance que procede del latín vulgar y que forma parte del dialecto de varias regiones de España como lo es Valencia, de Francia como lo es Rosellón, e Italia en la ciudad de Alguer, ha sido motivo de una dilatada pugna en cuanto a su ubicación genealógica. Inclusive, dicha lengua fue recientemente catalogada, según la revista *Ethnologue* (2009), como una de las lenguas más habladas en el mundo al ocupar la posición setentaicinco, siendo esta hablada en alrededor de dieciocho países. Según Munteanu Colán (2004), el primero en

proponer una filiación para la lengua fue Morel-Fatio quien, con base en razones de índole histórica, relacionó la lengua con el grupo galorrománico. Muchos han sido los lingüistas que han compartido dicha postura desde entonces y han realizado comparaciones con el léxico galorromance, por una parte, y con el de las lenguas iberorromances, por otra.

Los resultados parecen ser concluyentes; así lo afirma Enwistle (1969) quien apunta que la semejanza entre el catalán y el provenzal es tal que es casi imposible discernir si las entradas léxicas pertenecen a una u otra lengua, si son préstamos muy tempranos o si son coevoluciones. Desde la perspectiva de este autor, las semejanzas con el iberorrománico se explicarían debido a la historia que comparten las lenguas de la península desde la invasión árabe y la posterior reconquista. Para muchos otros autores, la relación entre el catalán y el occitano es semejante a la existente entre el español y el portugués, es decir, de lenguas cuya semejanza léxica, y hasta estructural, son de una alta semejanza (Corominas, citado por Munteanu Colán).

Si bien hay similitudes entre el catalán y las lenguas iberorromances, las mayores semejanzas se dan con el grupo galorromance, y principalmente con el occitano. En un nivel fonético, podemos mencionar similitudes entre el catalán y sus correspondientes

protofonemas con las demás lenguas pertenecientes al grupo iberorromance. Así, tenemos correspondencias fonéticas para *ts donde esta se realiza como /s/ en catalán, español, gallego y aragonés, entre otras lenguas. Léxicamente hablando, formas derivadas del latín como *clave* mantienen una forma exactamente igual en lenguas occitanas como el aragonés y el aranés. Enwistle cita, entre muchas otras, las siguientes afinidades: pérdida de vocales átonas, neutralización de vocales finales, desarrollo de negación reforzada y construcción de participios de pasado sobre tema de pretérito. Sin embargo, como es común en las lenguas, no se puede marcar una línea clara para todas estas características, pues algunas se presentan también en dialectos septentrionales del español, como el aragonés y el leonés.

Esta falta de límites claros entre ambos idiomas, divulgada más profusamente gracias a trabajos dialectológicos, ha servido de base para discutir la filiación de la lengua en tiempos más modernos. Se parte de que el español actual, por ejemplo, no es necesariamente el representante más fiel de los dialectos iberorrománicos antiguos, por lo que no es el candidato idóneo para comparar con el catalán y determinar claramente una asociación genética (Menéndez, 1950). Estas evidencias lingüísticas se pueden asociar a interpretaciones históricas y sirven

para rebatir la supuesta pertenencia del catalán al grupo galorrománico.

Amado Alonso y Menéndez Pidal, en ese sentido, proponen que el catalán nació en España, desde donde pasó hacia los Pirineos por razones políticas.

A veces observamos varios límites confundidos en un haz o conjunto de líneas que marchan juntas en toda o gran parte de su extensión: límites coincidentes, que obedecen a la irradiación de grandes masas de fenómenos lingüísticos; es decir, una porción de fenómenos fonéticos, morfológicos o sintácticos alcanzan un mismo límite llevados por el mismo impulso propagador. La coincidencia de los límites es por lo común el resultado del choque de dos lenguas o dialectos que se sienten entre sí como diversos; hay conciencia de la distinción entre las lenguas colindantes y ambas se excluyen mutuamente en gran parte de sus fenómenos característicos [...] Los límites lingüísticos entre [...] el aragonés y el catalán desde los Pirineos hasta hacia Benabarre, son principalmente sueltos o entrecruzados, mientras que más al sur son límites por lo general coincidentes... (Menéndez Pidal, 1916, 77-78)

Según esta teoría, el catalán era una modalidad más de los dialectos hablados en el siglo X en la península, donde formaban un *continuum* desde Galicia hasta Cataluña. Posteriormente, con

la imposición de la Marca Hispánica por Carlomagno, el idioma influyó en otras zonas a las que no había tenido acceso con anterioridad y estrechó relaciones con las lenguas d'Oil a las que accedió. Esta teoría, conocida como de origen mixto galo e ibérico, ha sido defendida a partir del estudio de textos antiguos y su confrontación con textos modernos.

Muchas son las teorías y diversos son los puntos de vista y las conclusiones a las que han llegado los diferentes estudiosos de la rama. Si bien la mayoría está de acuerdo ya sea con una u otra teoría (iberro- o galorromance), existen aún otras teorías, principalmente la de Bennedet Vidos, que aseguran que el catalán no es ni galorromance ni iberorromance, sino “pirenaicoromance.” Dicha afirmación supondría que esta lengua sería más bien un puente entre estos dos grandes grupos y agruparía al catalán con el gascón o el aragonés, por ejemplo, como diferenciada de las demás ramas.

Grupo galorrománico

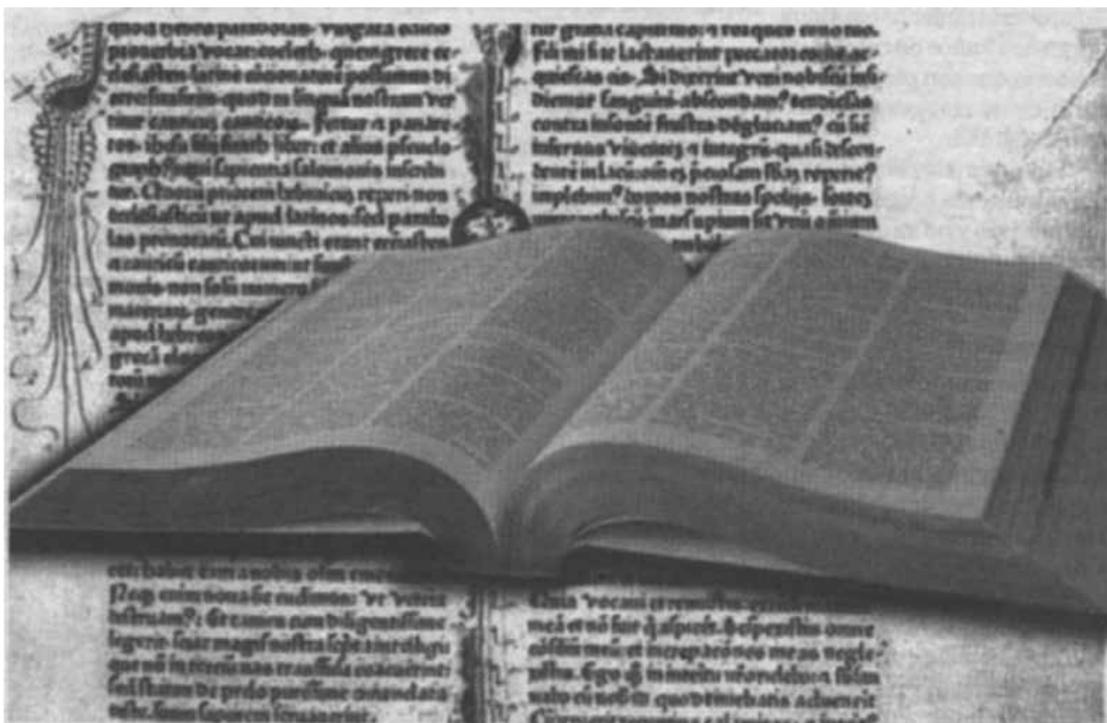
El grupo galorrománico es el más septentrional de la rama italocontinental. Se ubica físicamente en Francia, el norte de Italia, algunas partes de Bélgica y Suiza y los territorios que han tenido relación con estos países alrededor del mundo.

Las lenguas más conocidas son probablemente el francés, idioma de gran presencia internacional, y el occitano, el cual posee una considerable tradición literaria y política. Para efectos de esta investigación, ambos se considerarán lenguas distintas y representativas de los dos subgrupos en que, según datos de *Ethnologue*, se divide el tronco galorromance: oil y oc, respectivamente.

El francés

El francés posee una dilatada historia como lengua desde sus orígenes como un romance hablado en las provincias galas del imperio romano (con sustrato céltico) hasta la posterior influencia recibida por los pueblos francos que invadieron la región. El francés moderno surgió durante la Edad Media en París, gracias al impetuoso desarrollo de esta ciudad, lo cual la convirtió en la urbe más poblada de Europa alrededor del siglo XV. La teoría más aceptada es que la lengua es una koiné, producto de la unión de los múltiples dialectos romances vecinos (Lodge, 2010).

Con el tiempo, el francés fue desplazando a otras lenguas en la región, hasta que se consagró como lengua nacional mediante la constitución dictada al finalizar la Revolución Francesa. Posteriormente, debido al expansionismo francés (y en alguna medida belga) se propagó por prácticamente todos los continentes, donde hay presencia de la lengua hasta la actualidad.



La publicación *Ethnologue*, en su 16ta edición, afirma que el idioma es hablado, al menos, en 30 países y cuenta con 67,8 millones de hablantes. La Organización Internacional de la Francofonía (OIF), por su parte, estimaba en 2010 cerca de 200 millones de francófonos, de los cuales 72 millones eran catalogados como hablantes parciales.

De hecho, la organización *Internet World Stats* reporta que para 2010 hay 59,8 millones de internautas de habla francesa, con lo que este idioma es el octavo más representado en la Web. La Unesco, adicionalmente, afirma que para 2007, 4,4% de las páginas web estaban escritas en francés, lo que

significaba un tercer lugar, por detrás del inglés y del alemán.

En lo concerniente a los dialectos, hay una “ortodoxia” comúnmente tomada como la más representativa de qué es el francés, merced a la Academia Francesa. Esto hace que el parisino sea el dialecto más reconocido, aunque hay varios más, como el quebequense (que cuenta con su propia institución reguladora del idioma, la *Office Québécois de la langue française*), los dialectos del Magreb y subsaharianos (producto de la colonización francesa y belga en el continente africano), y las variedades belga, suiza y monegasca, por ejemplo, en el continente Europeo. La misma Francia presenta dialectos regionales,

que han resurgido en los últimos años con la concienciación de los hablantes de su identidad histórica (como el valón, el picardo o el normando).

En todo caso, para efectos de este estudio, tomaremos como insumo el dialecto francés parisino, pues es el más sencillo de rastrear y facilita la obtención y revisión de la lista de palabras necesarias, además de que es el más reconocido internacionalmente, como ya se dijo.

El occitano

Con respecto a la segunda de este grupo, el occitano, cabe señalar que su origen se puede explicar, según algunos autores (Bec, 1973), en las características geográficas del sur de Francia, donde se presentan tres grandes divisiones naturales que limitan la Occitania histórica: los Pirineos, los Alpes y el Macizo Central. Tal división generó desde épocas tempranas un romance muy característico de la región, distinto al del norte (dialectos de *langue d'oïl*) y, más aún, al del este y al del sur (dialectos retorrománicos y dialectos ibéricos, respectivamente).

Las primeras referencias del occitano se tienen gracias a las *trovas* medievales, que extendieron el idioma por toda la Europa cortesana e influyeron notablemente en la cultura y visión de mundo del continente. Sin embargo, los centros de poder de habla occitana (Tolosa, Burdeos y Marsella, principalmente) no incursionaron en

campañas militares permanentes y fueron absorbidos posteriormente por Francia; por ello, el idioma se mantuvo confinado a su región histórica, que comprende el actual Mediodía francés (a excepción del País Vasco y los Pirineos Orientales), el Valle de Arán, en España y el Piamonte, en Italia.

En territorio francés, el occitano se vio muy perjudicado por las políticas monolingües del Estado, las cuales han desestimado el uso de las lenguas vernáculas a favor del idioma nacional; por ello, actualmente el occitano está relegado al ámbito familiar y casi confinado a las zonas rurales en Francia (Manzano, 2004), mientras que sí goza de carácter cooficial (junto al español y al catalán) en el Valle de Arán, donde se le conoce también como aranés.

En años recientes (y se podría decir que desde la fundación del *Félibrige* por Frédéric Mistral), ha habido un esfuerzo más o menos sostenido, de parte de varios grupos independientes, por revigorar la lengua occitana y fundar organismos encargados de su difusión. Hasta el momento, dos de los más reconocidos son el Institut d'Estudis Occitans y el Conselh de la Lengua Occitana.

Comúnmente, se proponen tres grupos dialectales para la lengua (Béc, supra): el septentrional, que comprende el limosino, el auvernés y el alpino;

el meridional, con el languedociano y el provenzal, y el aquitano, cuyo único dialecto es el gascón. En total, *Ethnologue* calculaba para 2009 un aproximado de 2 050 000 hablantes, en su mayoría bilingües con la lengua del territorio donde habitan. Para este estudio tomaremos como base el provenzal, pues es tal vez el dialecto con mayor proyección.

Grupo iberorrománico

El grupo iberorrománico, originalmente, comprende todas las lenguas habladas en la península ibérica e islas Baleares, si se excluye el controvertido catalán. En este estudio se tomarán como representativos el español y el portugués, lenguas que se caracterizan a continuación.

El español

El español, también conocido como castellano, es una lengua de la rama iberorrománica de la familia indoeuropea. Los primeros textos encontrados, y que se conservan aún, son las *Glosas Emilianenses* y *Silenses*, las cuales datan del siglo XI. Las glosas, al provenir de una región que fue punto de encuentro de diversas culturas, permiten distinguir los rasgos castellanos, navarro-aragoneses e, incluso, germánicos que han caracterizado a la lengua desde sus inicios.

Con respecto al carácter formal de la lengua, es importante señalar dos

aspectos fundamentales. En primer lugar, el intento de regularización de la lengua a partir del siglo XIII gracias a la labor de Alfonso X el Sabio, quien intentó renovarla y enriquecerla en los aspectos ortográficos, gramaticales y léxicos. En segundo lugar, cabe destacar la importancia de la creación de la Real Academia Española de la Lengua en 1714, pues desde ese momento la institución tiene autoridad normativa sobre el idioma.

Ahora bien, según datos tomados de *Ethnologue*, para el año 1996 la cantidad de hablantes se estimaba en alrededor de 329 millones. Sin embargo, en estudios más recientes, se asegura que la cifra aumentó a 500 millones de hablantes distribuidos en los cinco continentes: América (desde Estados Unidos hasta Argentina, se incluyen territorios insulares), Europa (en España como lengua oficial, aunque se pueden encontrar grupos aislados en otros países), Asia (especialmente en Filipinas donde fue lengua oficial hasta 1973), África (en Guinea Ecuatorial como lengua oficial y grupos pequeños en el norte, Marruecos) y en Oceanía (en la isla de Pascua que pertenece a Chile).

El español, como otras lenguas ampliamente difundidas debido a los procesos de colonización, está sujeto a variaciones regionales y sociolingüísticas. A pesar de esta heterogeneidad, el grado de variación que presenta no

es tan grande como para producir la mutua ininteligibilidad.

Finalmente, entre las variedades dialectales españolas se señala el andaluz, el aragonés, el murciano, el navarro, el castellano, el español de las islas canarias - más conocido como isleño - las diferentes variedades del español de América y el silbo gomero.

El portugués

El portugués, al igual que el español, es una lengua de la rama iberorrománica de la familia indoeuropea. Este idioma comenzó a desarrollarse en la parte occidental de la Península Ibérica a partir del latín implantado por los romanos en el siglo III a.C. Sin embargo, su diferenciación con las otras lenguas romances tuvo inicio luego de la caída del Imperio Romano y las invasiones bárbaras que se dieron alrededor del siglo V.

Ahora bien, como lengua escrita, el portugués fue utilizado hasta el siglo IX. Además, los autores distinguen dos periodos de la lengua: el arcaico (desde el siglo XII hasta el XVI, aunque también se ha señalado el IX como el inicio de este periodo) y el moderno (desde mediados del siglo XVI hasta la actualidad).

Según *Ethnologue*, para el año 1996 el portugués era la lengua de 177 981 570 de hablantes. A pesar de esto, datos recientes indican que es el idioma

oficial de 240 millones de personas, las cuales están distribuidas en África, América, Europa y Asia.

Dialectalmente, el portugués es una lengua muy rica; sin embargo, la diferenciación más grande se da entre el portugués de Portugal y el de Brasil (brasileiro). En el portugués de Portugal se establecen dos categorías principales. En primer lugar, la Septentrional, la cual conserva muchas afinidades con el gallego; en segundo lugar, la Meridional y central, que corresponde al territorio reconquistado y por lo tanto está más afectada por el superestrato árabe.

Otros dialectos, según datos de *Ethnologue*, son el beiran (Beirão), el alentejan (Alentejano), el algarvian (Algarvio), el minhota (Minhoto), el transmontan (Transmontano), el madeirese (Madeirense), el azorean (Açoriano) y el (Estremeño).

Finalmente, cabe señalar que la variedad aceptada como estándar desde el siglo XVIII corresponde al portugués meridional (dialecto estremeño), el cual está representado por el lenguaje de las ciudades universitarias de Coimbra y Lisboa.

Como se observa en este apartado, la mayoría de los estudios han relacionado, generalmente, al catalán con lenguas como el occitano mayormente y han descartado, en su mayoría, que

esta lengua tenga una relación parentética con las demás lenguas citadas anteriormente. Sin embargo, ciertas teorías ponen en tela de juicio que esta relación parentética entre catalán y occitano – o bien entre el grupo galorrománico- sea tan tajante y clara. Es por esta razón que se ha tomado a estas lenguas como una forma de constatar la efectiva utilidad y precisión de la lexicoestadística y la lista Swadesh en cuanto a los resultados obtenidos al comparar al catalán con las demás lenguas de los grupos iberorromance y galorrománico, lo cual podrá esclarecer (o dar una idea más clara) de la pertenencia del catalán a uno de estos dos grandes grupos de la familia indoeuropea.

Fundamentos teóricos y metodológicos de la lexicoestadística

La lexicoestadística como teoría

Para establecer correspondencias entre las lenguas por estudiar, como ya se anunció, se recurrirá en este trabajo a la lexicoestadística, técnica utilizada para proponer relaciones genealógicas entre idiomas. Por ello, a continuación se presenta una pequeña descripción del método.

Con la lexicoestadística se someten a comparación dos o más lenguas, a partir de una lista de vocabulario denominado como básico. Las listas contienen 200 o 100 palabras, y

fueron propuestas por el lingüista Morris Swadesh a partir de sus observaciones de que algunas palabras cambiaban más o menos a través del tiempo en las lenguas que había estudiado. Calculando cuánto tiempo pasa para que una de estas palabras “básicas” cambie, Swadesh propuso que se puede argüir si una lengua está relacionada con otra genealógicamente, y en qué nivel, a partir del porcentaje de semejanza entre los vocabularios de las dos lenguas.

Así Pues, entre 100% y 81%, las variedades son dialectos de una misma lengua; entre 80% y 36%, las variedades son lenguas de una misma familia en sentido estricto; entre 35% y 12%, las variedades son lenguas de un mismo tronco; entre 11% y 4%, las variedades pertenecen a un mismo microfilo, y entre 3% y 1%, las variedades pertenecen a un mismo mesofilo.

Los supuestos de la lexicoestadística, sin embargo, han sido sometidos a muchas críticas. En primer lugar, como apunta Sihler (2000), el vocabulario básico es una noción más fácil de nombrar que de definir.

Tradicionalmente, se ha considerado vocabulario básico el relacionado con las partes del cuerpo, los primeros numerales, los cuerpos celestes o las relaciones familiares. Sin embargo, múltiples trabajos en antropología lingüística demuestran lo falaz de este

supuesto ya que algunas partes del cuerpo son tabú para ciertas culturas, o las relaciones familiares no revisiten igual importancia para todos. Por ejemplo, los numerales se han mantenido relativamente estables en las lenguas indoeuropeas, pero han sufrido fuertes cambios en lenguas del pacífico asiático.

Además, tal como demuestra Rosenfelder (2002), palabras semejantes fonéticamente pero no relacionadas genealógicamente son muy posibles, dado el relativamente limitado haz de contrastes fonológicos que los seres humanos usamos en las lenguas.

Aunado a esto, se debate mucho la influencia de los préstamos y si estos son tan raros en el vocabulario básico como se pensó en un momento. Autores como Starostin (2010) o Sihler mencionan la dificultad que existe para diferenciar cognados de préstamos cuando solamente se tiene la lista de palabras y, sobre todo, cuando se sabe que las lenguas por comparar comparten un origen relativamente común, pero que también han estado en contacto durante largo tiempo. En caso de que no se tengan conocimientos de la lengua madre, la elección suele parecer poco menos que arbitraria, pues no hay consenso para delimitar “hasta dónde” deben parecerse dos lenguas y por qué.

A pesar de estas dificultades, innegables en el método, éste es todavía ampliamente utilizado pues, como reconoce de nuevo Starostin, otros métodos podrían traer otras limitantes aún mayores. Se siguen utilizando las listas de vocabulario básico propuestas por Morris Swadesh, primero una de 200 palabras y luego otra de 100, aunque parece predominar la visión con respecto a que la lexicoestadística y las listas de cognados solamente sirven para abrir el debate en cuanto a la situación de una lengua en relación con otras, y no para terminarlo (Strazny, 2000). Por ello, se espera hacer un aporte, primero, al evaluar la utilidad y exactitud de los resultados obtenidos al aplicar la metodología de la lexicoestadística y, segundo, al discernir, si la metodología es exitosa, al cual de los grupos pertenece esta lengua.

Metodología

Como se señaló, para realizar esta investigación se tomaron en cuenta las herramientas que se utilizan en el ámbito denominado lexicoestadística. Esto permitió analizar y comparar las diferentes lenguas en estudio para, de esta manera, resolver las cuestiones que atañen a la correcta calificación del catalán dentro de la familia románica así como para evaluar la efectividad de la lista de Swadesh y la metodología de la lexicoestadística.

En primer lugar, hay que señalar que las listas de vocabulario para las cinco lenguas (español, portugués, catalán, francés y occitano) fueron obtenidas de Wikipedia. Las listas se confrontaron con la que presenta, para el inglés, Lehmann (1962). Así, se corroboró que todas contenían las mismas palabras en los respectivos idiomas.

Ahora bien, cuando una entrada correspondía a dos traducciones diferentes en la lengua se tomó como criterio eliminar una de las palabras y conservar aquella que aparece como primera acepción en el diccionario utilizado. Además, cabe señalar que, a pesar de haberse tomado en cuenta dentro de la búsqueda y de que no fueron

eliminadas de la lista, para el análisis de los resultados no fueron tomadas en cuenta las palabras correspondientes a “there”, “this” y “that” debido a la dificultad para encontrar la correspondencia exacta entre todas las lenguas.

En segundo lugar, se procedió a marcar con colores (y números para efectos de publicación) los cognados que comparte el catalán con cada una de las lenguas. Para realizar este procedimiento fue necesario recurrir a diccionarios etimológicos que permitieran determinar si realmente había relación genealógica entre los lexemas o no. A continuación, se especifica el significado de cada color:

Vino¹	el catalán está emparentado genealógicamente con todas las lenguas.
Celeste²	el catalán no está emparentado con ninguna lengua.
Verde³	el catalán solamente está emparentado con una de las lenguas galorrománicas.
Naranja⁴	el catalán solamente está emparentado con una de las lenguas iberorrománicas.
Café⁵	el catalán está emparentado con una lengua de cada subgrupo.
Azul⁶	el catalán está emparentado solamente con las dos lenguas galorrománicas.
Rosado⁷	el catalán está emparentado solamente con las dos lenguas iberorrománicas.
Gris⁸	el catalán está emparentado con las dos lenguas iberorrománicas y con una galorrománica.
Morado⁹	el catalán está emparentado con las dos lenguas galorrománicas y con una iberorrománica.

Finalmente, se debe aclarar que la primera columna a la izquierda, correspondiente a Swadesh, no fue contemplada dentro del estudio y que los cognados en color negro (sin número) corresponden a aquellas que no están genealógicamente emparentadas con el lexema correspondiente al catalán.

Listas y señalamiento de cognados

A continuación se presenta la lista que se llenó a partir de la correspondiente al inglés de Swadesh.

Swadesh	Español	Port.	Catalán	Fr.	Occitano.
1. all	todo ¹	todo ¹	tot ¹	tout ¹	tot ¹
2. and	y ¹	e ¹	i ¹	et ¹	e ¹
3. animal	animal ¹	animal ¹	animal ¹	animal ¹	animal ¹
4. ashes	cenizas ¹	cinzas ¹	cendres ¹	endre ¹	endre ¹
5. at	a ¹	a ¹	a ¹	à ¹	a ¹
6. back	espalda	costas	esquena ³	dos	esquina ³
7. bad	malo	mau	dolent ²	mauvais	marrít
8. bark	corteza	casca	escorça ³	écorce ³	rusca
9. because	porque ¹	porque ¹	perquè ¹	parce que ¹	perque ¹
10. belly	vientre ¹	ventre ¹	ventre ¹	ventre ¹	ventre ¹
11. big	grande ¹	grande ¹	gran ¹	grand ¹	grand ¹
12. bird	pájaro	pássaro	ocell ⁶	oiseau ⁶	aucèl ⁶
13. to bite	morder	morder	mossegar ³	mordre	mossegar ³
14. black	negro ¹	negro ¹	negre ¹	noir ¹	negre ¹
15. blood	sangre ¹	sangue ¹	sang ¹	sang ¹	sang ¹
16. to blow	soplar	soprar	bufar ³	souffler	bufar ³
17. bone	hueso ¹	osso ¹	os ¹	os ¹	òs ¹
18. to breathe	respirar ¹	respirar ¹	respirar ¹	respirer ¹	respirar ¹
19. to burn	quemar ⁸	queimar ⁸	cremar ⁸	brûler	cremar ⁸
20. child	niño ⁷	menino ⁷	nen ⁷	enfant	enfan
21. cloud	nube ¹	nuvem ¹	núvol ¹	nuage ¹	nívol ¹
22. cold	frío ¹	frio ¹	fred ¹	froid ¹	freg ¹
23. to come	venir ¹	vir ¹	venir ¹	venir ¹	venir ¹
24. to count	contar ¹	contar ¹	comptar ¹	compter ¹	comptar ¹
25. to cut	cortar	cortar	tallar ²	couper	copar
26. day	día ⁷	dia ⁷	dia ⁷	jour	jorn
27. to die	morir ¹	morrer ¹	morir ¹	mourir ¹	morir ¹
28. to dig	cavar ⁸	cavar ⁸	cavar ⁸	creuser	cavar ⁸
29. dirty	sucio	sujo	brut ³	sale	brut ³
30. dog	perro	cão	gos ²	chien	can
31. to drink	beber ¹	beber ¹	beure ¹	boire ¹	beure ¹
32. dry	seco ¹	seco ¹	sec ¹	sec ¹	sec ¹

33. dull	romo ⁷	rombo ⁷	rom ⁷	émoussé	bercat
34. dust	polvo ¹	poeira ¹	pols ¹	poussière ¹	pols ¹
35. ear	oreja ¹	orelha ¹	orella ¹	oreille ¹	aurelha ¹
36. earth	tierra ¹	terra ¹	terra ¹	terre ¹	tèrra ¹
37. to eat	comer	comer	menjar ⁶	manger ⁶	manjar ⁶
38. egg	huevo ¹	ovo ¹	ou ¹	œuf ¹	uòu ¹
39. eye	ojo ¹	olho ¹	oill ¹	œil ¹	olh ¹
40. to fall	caer ⁷	cair ⁷	caure ⁷	tomber	tombar
41. far	lejos	longe ⁹	lluny ⁹	loin ⁹	luenh ⁹
42. fat	grasa ⁹	gordura	greix ⁹	grasse ⁹	grais ⁹
43. father	padre ¹	pai ¹	pare ¹	père ¹	paire ¹
44. to fear	temer ⁷	temer ⁷	témer ⁷	craindre	crentar
45. feather	pluma ¹	pluma ¹	ploma ¹	plume ¹	pluma ¹
46. few	poco ¹	pouco ¹	poc ¹	peu ¹	pauc ¹
47. to fight	luchar ⁷	lutar ⁷	lluitar ⁷	se battre	se batre
48. fire	fuego ¹	fogo ¹	foc ¹	feu ¹	fuòc ¹
49. fish	pez ¹	peixe ¹	peix ¹	poisson ¹	peis ¹
50. five	cinco ¹	cinco ¹	cinc ¹	cinq ¹	cinc ¹
51. to float	flotar ¹	flutuar ¹	flotar ¹	flotter ¹	flotar ¹
52. to flow	fluir ⁷	fluir ⁷	fluir ⁷	couler	rajar
53. flower	flor ¹	flor ¹	flor ¹	fleur ¹	flor ¹
54. to fly	volar ¹	voar ¹	volar ¹	voler ¹	volar ¹
55. fog	niebla	névoa	boira ²	brouillard	nèbla
56. foot	pie ¹	pé	peu	ped	pè
57. four	cuatro ¹	quatro	quatre	quatre	quatre
58. to freeze	helar ¹	gelar	gelar	geler	gelar
59. fruit	fruta ¹	fruta	fruita	fruit	fruch
60. to give	dar	dar	donar ⁶	donner ⁶	donar ⁶
61. good	bueno ¹	bom ¹	bo ¹	bon ¹	bon ¹
62. grass	hierba ¹	erva ¹	herba ¹	herbe ¹	èrba ¹
63. green	verde ¹	verde ¹	verd ¹	vert ¹	verd ¹
64. guts	entrañas	entranhas	budellada ⁶	boyaux ⁶	budelada ⁶
65. hair	pelo ⁸	pêlo ⁸	pèl ⁸	cheveux	pel ⁸
66. hand	mano ¹	mão ¹	mà ¹	main ¹	man ¹
67. he	él ¹	ele ¹	ell ¹	il ¹	el ¹
68. head	cabeza ⁸	cabeça ⁸	cap ⁸	tête	cap ⁸
69. to hear	oír ⁸	ouvir ⁸	oir ⁸	entendre	ausir ⁸
70. heart	corazón ¹	coração ¹	cor ¹	cœur ¹	còr ¹
71. heavy	pesado ⁸	pesado ⁸	pesat ⁸	lourd	pesuc ⁸
72. here	aquí ¹	aquí ¹	ací ¹	ici ¹	aic ¹
73. to hit	golpear ⁴	bater	copejar ⁴	frapper	tustar
74. to hold	sostener ⁴	segurar	sostenir ⁴	tenir	tener
75. how	cómo ¹	como ¹	com ¹	comment ¹	quinament ¹

76. to hunt	cazar ¹	caçar ¹	caçar ¹	chasser ¹	cassar ¹
77. husband	marido ¹	marido ¹	marit ¹	mari ¹	marit ¹
78. i	yo ¹	eu ¹	jo ¹	je ¹	eu ¹
79. ice	hielo ⁷	gelo ⁷	gel ⁷	glace	glaça
80. if	si ¹	se ¹	si ¹	si ¹	se ¹
81. in	en	em	dins ⁶	dans ⁶	dins ⁶
82. to kill	matar ⁷	matar ⁷	matar ⁷	tuer	tuar
83. know	saber ¹	saber ¹	saber ¹	savoir ¹	saber ¹
84. lake	lago ¹	lago ¹	llac ¹	lac ¹	lac ¹
85. to laugh	reír ¹	rir ¹	riurer ¹	rire ¹	rire ¹
86. leaf	hoja ¹	folha ¹	fulla ¹	feuille ¹	fuelha ¹
87. left	izquierda ⁸	esquerda ⁸	esquerra ⁸	gauche	esquerra ⁸
88. leg	pierna	perna	gamba ⁶	jambe ⁶	camba ⁶
89. lie	yacer ¹	jazer ¹	jeure ¹	gésir ¹	jàser ¹
90. to live	vivir ¹	viver ¹	viurer ¹	vivre ¹	viure
91. liver	hígado ¹	figado ¹	fetge ¹	foie ¹	fetge ¹
92. long	largo ¹	longo ¹	Llarg ¹	long ¹	larg ¹
93. louse	piojo ¹	piolho ¹	poll ¹	pou ¹	pesolh ¹
94. man	hombre ¹	homem ¹	home ¹	homme ¹	òme ¹
95. many	muchos ⁷	muitos ⁷	molt ⁷	beaucoup	fôrça
96. meat	carne ⁸	carne ⁸	carn ⁸	viande	carn ⁸
97. mother	madre ¹	mãe ¹	mare ¹	mère ¹	maire ¹
98. mountain	montaña ¹	montanha ¹	muntanya ¹	montagne ¹	montanha ¹
99. mouth	boca ¹	boca ¹	boca ¹	bouche ¹	boca ¹
100. name	nombre ¹	nome ¹	nom ¹	nom ¹	nom ¹
101. narrow	estrecho ¹	estreito ¹	estret ¹	étroit ¹	estrech ¹
102. near	cerca	perto	prop ³	près	probet ³
103. neck	cuello ⁹	pescoço	coll ⁹	cou ⁹	còl ⁹
104. new	nuevo ¹	novο ¹	nou ¹	nouveau ¹	nòu ¹
105. night	noche ¹	noite ¹	nit ¹	nuit ¹	nuech ¹
106. nose	nariz ¹	nariz ¹	nas ¹	nez ¹	nas ¹
107. not	no ⁸	não ⁸	no ⁸	ne pas ⁸	pas
108. old	viejo ¹	velho ¹	vell ¹	vieux ¹	vièlh ¹
109. one	uno ¹	um ¹	un ¹	un ¹	un ¹
110. other	otro ¹	outro ¹	altre ¹	autre ¹	autre ¹
111. person	persona ¹	pessoa ¹	persona ¹	personne ¹	persona ¹
112. to play	jugar ¹	jogar ¹	jugar ¹	jouer ¹	jogar ¹
113. to pull	tírar ⁹	puxar	tírar ⁹	tírer ⁹	tírar ⁹
114. to push	empujar ⁸	empurrar ⁸	empènyer ⁸	pousser ⁸	butar
115. to rain	llover ¹	chover ¹	ploure ¹	pleuvoir ¹	plover ¹
116. red	rojo	vermelho ⁴	vermell ⁴	rouge	roge
117. right	correcto ¹	correcto ¹	correcte ¹	correct ¹	corrècte ¹
118. right	derecha ¹	direita ¹	dreta ¹	droite ¹	drecha ¹

119. river	río ⁸	rio ⁸	riu ⁸	fleuve	riu ⁸
120. road	camino ⁵	estrada	cami ⁵	route	camin ⁵
121. root	raíz ¹	raiz ¹	arrel ¹	racine ¹	rasiga ¹
122. rope	cuerdas ¹	corda ¹	corda ¹	corde ¹	còrda ¹
123. rotten	podrido ¹	podrido ¹	podrit ¹	pourri ¹	poirit ¹
124. rub	frotar	esfregar ⁴	fregar ⁴	frotter	fretar
125. salt	sal ¹	sal ¹	sal ¹	sel ¹	sal ¹
126. sand	arena	areia	sorra ⁶	sable ⁶	sabla ⁶
127. to say	decir ¹	dizer ¹	dir ¹	dire ¹	dire ¹
128. scratch	rascar	coçar	gratar ⁶	gratter ⁶	gratar ⁶
129. sea	mar ¹	mar ¹	mar ¹	mer ¹	mar ¹
130. to see	ver ¹	ver ¹	veure ¹	voir ¹	vèser ¹
131. seed	semilla	semente	llavor ²	graine	grana
132. to sew	coser ¹	cozer ¹	cosir ¹	coudre ¹	cóser ¹
133. sharp	afilado ⁷	afiado ⁷	afilat ⁷	tranchant	trencant
134. short	corto ¹	curto ¹	curt ¹	court ¹	cort ¹
135. to sing	cantar ¹	cantar ¹	cantar ¹	chanter ¹	cantar ¹
136. to sit	sentarse ¹	sentar-se ¹	asseure ¹	s'asseoir ¹	s'assetar ¹
137. skin	piel ¹	pele ¹	pell ¹	peau ¹	pèl ¹
138. sky	Cielo ¹	céu ¹	cel ¹	ciel ¹	cèl ¹
139. to sleep	dormir ¹	dormir ¹	dormir ¹	dormir ¹	dormir ¹
140. small	pequeño ¹	pequeno ¹	petit ¹	petit ¹	pichon ¹
141. to smell	oler ⁴	cheirar	olorar ⁴	sentir	sentir
142. smoke	humo ¹	fumaça ¹	fum ¹	fumée ¹	fum ¹
143. smooth	liso ¹	liso ¹	llis ¹	lisse ¹	lis ¹
144. snake	serpiente ¹	serpente ¹	serp ¹	serpent ¹	sèrp ¹
145. snow	nieve ¹	neve ¹	neu ¹	neige ¹	nèu ¹
146. some	algunos ⁷	alguns ⁷	alguns ⁷	quelques	qualques
147. to spit	escupir ⁸	cuspir ⁸	escopir ⁸	cracher	escupir ⁸
148. to split	partir ⁷	partir ⁷	partir ⁷	fendre	fendre
149. to squeeze	apretar ⁷	apertar ⁷	apretar ⁷	presser	estrenher
150. to stab	apuñalar ¹	apunhalar ¹	apunyal ¹	poignarder ¹	emponhalar ¹
151. to stand	estar de pie ⁷	estar em pé ⁷	estar de peu ⁷	se lever	se levar
152. star	estrella ¹	estrela ¹	estrella ¹	étoile ¹	estela ¹
153. stick	palo ⁷	pau ⁷	pal ⁷	bâton	bròc
154. stone	piedra ¹	pedra ¹	pedra ¹	pierre ¹	pèira ¹
155. straight	recto ⁷	reto ⁷	recte ⁷	droit	drech
156. to suck	chupar ¹	chupar ¹	xuclar ¹	sucer ¹	chucar ¹
157. sun	sol ¹	sol ¹	sol ¹	soleil ¹	solelh ¹
158. to swell	hincharse ⁷	inchar-se ⁷	inflar ⁷	se gonfler	conflar
159. to swim	nadar ⁸	nadar ⁸	nedar ⁸	nager	nadar ⁸
160. tail	cola ¹	cauda ¹	cua ¹	queue ¹	coa ¹
161. that	ese	esse	aquest	cet	aqueste

162. there	ahí	aí	allà	là	aquí
163. they	ellos ¹	eles ¹	ells ¹	ils ¹	eles ¹
164. thick	espeso ¹	espesso ¹	espès ¹	épais ¹	espés ¹
165. thin	delgado	delgado	prim ³	mince	prim ³
166. to think	pensar ¹	pensar ¹	pensar ¹	penser ¹	pensar ¹
167. this	este	este	aquest	ce	aqueste
168. thou	tú ¹	tu ¹	tu ¹	tu ¹	tu ¹
169. three	tres ¹	três ¹	tres ¹	trois ¹	tres ¹
170. to throw	lanzar ¹	lançar ¹	llançar ¹	lancer ¹	lançar ¹
171. to tie	atar	atar	lligar ²	attacher	estacar
172. tongue	lengua ¹	língua ¹	llengua ¹	langue ¹	lenga ¹
173. tooth	diente ¹	dente ¹	dent ¹	dent ¹	dent ¹
174. tree	árbol ¹	árvore ¹	arbre ¹	arbre ¹	arbre ¹
175. to turn	girar ⁷	girar ⁷	girar ⁷	tourner	virar
176. two	dos ¹	dois ¹	dos ¹	deux ¹	dos ¹
177. to vomit	vomitarse ¹	vomitarse ¹	vomitarse ¹	vomir ¹	vomir ¹
178. to walk	caminar ⁸	caminhar ⁸	caminar ⁸	marcher	caminar ⁸
179. warm	cálido ¹	cálido ¹	càlid ¹	chaud ¹	caud ¹
180. to wash	lavar	lavar	rentar ²	laver	lavar
181. water	agua ¹	água ¹	aigua ¹	eau ¹	aiga ¹
182. we	nosotros ¹	nós ¹	nosaltres ¹	nous ¹	nosautres ¹
183. wet	mojado ¹	molhado ¹	mullat ¹	mouillé ¹	moth ¹
184. what?	qué ¹	que ¹	què ¹	quel ¹	qué ¹
185. when?	cuándo ¹	quando ¹	quan ¹	quand ¹	quand ¹
186. where?	dónde ¹	onde ¹	on ¹	où ¹	ont ¹
187. white	blanco ¹	branco ¹	blanc ¹	blanc ¹	blanc ¹
188. who?	quién ¹	quem ¹	qui ¹	qui ¹	qui ¹
189. wide	ancho ⁴	largo	ample ⁴	large	larg
190. wife	esposa	esposa	dona ²	épouse	esposada
191. wind	viento ¹	vento ¹	vent ¹	vent ¹	vent ¹
192. wing	ala ¹	asa ¹	ala ¹	aile ¹	ala ¹
193. wipe	limpiar	enxugar ⁹	eixugar ⁹	essuyer ⁹	eissugar ⁹
194. with	con	com	amb ³	avec	amb ³
195. woman	mujer	mulher	dona ²	femme	femna
196. woods	bosque ¹	bosque ¹	bosc ¹	bois ¹	bòsc ¹
197. worm	gusano	verme	cuc ²	ver	vèrme
198. ye	vosotros ¹	vocês ¹	vosaltres ¹	vous ¹	vosautres ¹
199. year	año ¹	ano ¹	any ¹	année ¹	annada ¹
200. yellow	amarillo	amarelo	groc ²	jaune	jaune

Discusión de los resultados y conclusiones

Según la clave de color empleada, y tras corroborar mediante diccionarios

etimológicos cuáles palabras están relacionadas genealógicamente, se creó la siguiente tabla:

Color/ Relación	Número de cognados	Porcentaje
Todas están emparentadas con el catalán ¹	125	63,45
Catalán no está emparentado con ninguna ²	11	5,59
Catalán-1 lengua galorrománica ³	8	4,06
Catalán- 1 lengua iberorrománica ⁴	6	3,04
Catalán-1 galorrománica y 1 iberorrománica ⁵	1	0,5
Catalán- 2 galorrománicas ⁶	8	4,06
Catalán – 2 iberorrománicas ⁷	19	9,64
Catalán- 2iberorrománicas -1 galorrománica ⁸	14	7,1
Catalán- 2 galorrománicas- 1 iberorrománica ⁹	5	2,53
TOTAL	197	99,97

Según los porcentajes propuestos por Swadesh, se puede observar, en primer lugar, que para la interpretación de datos lexicoestadísticos, las cinco lenguas pertenecen a una misma familia, lo cual estaba ya plenamente demostrado en la documentación existente.

En cuanto a la filiación del catalán con alguno de los dos grupos, los resultados obtenidos no parecen concluyentes. Las claves azul⁶ y verde³, que excluyen el español y el portugués, pero relacionan el catalán con el francés o el occitano, dan como resultado 8,12%, por lo que, según Swadesh, estaríamos hablando de un microfílo. La contraparte iberorrománica, es decir, la

suma de las claves naranja⁴ y rosado⁷, da 12,7%, resultado poco diferente al del galorromance. No obstante, si estaríamos al nivel de decir que el catalán forma parte de un tronco común con el portugués y el español (iberorromance) ya que estaríamos entre el rango del 35% y el 12%, lo que aclararía, de cierta forma, el debate sobre la pertenencia de estas lenguas a una u otra familia.

Sin embargo, a pesar de esta estadística y, aunque no está incluido en la clave, si se hace una comparación directa con una única lengua, se puede observar que el catalán comparte 152 cognados con el occitano (galorromance), de los cuales 8 son

exclusivos, es decir, son compartidos únicamente entre ambos idiomas. El número es interesante, sobre todo si se compara con cuántos cognados comparte exclusivamente con el español (iberorromance), 4 de 165. Es decir, de las palabras compartidas con el galorromance, 5% están relacionadas únicamente con el occitano, mientras que de las compartidas con el ibero-romance, 2,4% están relacionadas únicamente con el español. Por tal, no se ve una mayor cercanía con una de las dos lenguas con las que se suele agrupar el catalán. Esta estadística contrasta claramente con la anterior, lo que nos hace dudar, una vez más, del método bajo análisis.

Tampoco podemos concluir que el catalán está emparentado con las dos lenguas iberorrománicas y con una galorrománica, o bien con las dos lenguas galorrománicas y con una iberorrománica. Como se representa con la clave gris⁸, del total de los cognados, solo un 7,1% estaría emparentado con las dos lenguas iberorrománicas y con una galorrománica, lo cual, una vez más, sería un porcentaje demasiado pobre como para aventurarse a decir que forma parte de ambas lenguas más una Gala. Lo mismo sucede con la clave morada⁹, la cual supone que el catalán está emparentado con las dos lenguas galorrománicas y con una iberorrománica. En este caso, el porcentaje es aún más bajo, con tan solo un 2,53%, por lo que las variedades pertenecerían a un mismo mesofilo.

Se concluye, luego de evaluar -mediante la aplicación de este enfoque a las lenguas ya citadas- el método lexicostatístico y la lista propuesta por Swadesh, que éste no permite llegar a un criterio concreto sobre el cual basarse para ubicar el catalán sea con las lenguas galorromances, sea con las iberorromances. Según este método, es claro que está emparentado con ambas lenguas, pero los porcentajes no son concluyentes. Por ello, parece que es necesario utilizar otra metodología para lograr la ubicación. Debemos, entonces, coincidir con lo expuesto anteriormente por Strazny (2000) quien apunta que parece predominar la visión con respecto a que la lexicostatística y las listas de cognados son útiles para darnos una idea de la situación entre lenguas posiblemente emparentadas o cercanas pero no se debe basar solamente en este método para determinarlo por completo.

Tal como se vio en el estado de la cuestión, los criterios de ubicación del idioma son variados: por geografía, por historia o por razones fonéticas. Este último método es el que parece más indicado en este caso: por ejemplo, a pesar de que todas las glosas de “all” están relacionadas etimológicamente, hay una mayor similitud del catalán con las del grupo galorromance, que sonoriza la /*t/. Aspectos como estos, sistematizados con metodologías como el método comparativo, arrojarían mucha más luz sobre la cuestión.

Bibliografía

- Académie Française. *Dictionnaire de l'Académie Française* (9na ed.) Disponible en: <http://www.academie-francaise.fr/dictionnaire/>
- Bac, P. (1973). *La Langue Occitan* (3a ed.). París: Presses Universitaires de France.
- Ortega Cavero, D. (1966). *Diccionario español-portugués*. Barcelona: Editorial Ramón Sopena.
- Diccionario Collins Pocket francés*. (1997). México: Editorial Grijalbo.
- Fundació Enciclopèdia Catalana. (s.f.) *Diccionari.cat*. Disponible en: <http://diccionari.cat>.
- Institut d'Estudis Catalans. (2007). *Diccionari de la llengua catalana* (2da. ed.). Barcelona.
- Lehmann, W. (1969). *Introducción a la Lingüística Histórica*. Madrid: Gredos.
- Levy, E. (1966). *Petit dictionnaire provençal-français*. (4ta. ed.).
- Lodge, A. (2010). "The Sources of Standardisation in French – Written or Spoken?" en R. Ingham (ed.). *The Anglo-Norman Language and its Contexts*. York: York Medieval Press, pp 26-43.
- Luraghi, S. (2010). *Continuum Companion to Historical Linguistics*. Londres: Continuum International Publishing.
- Manzano, F. (2004). "Situation and use of Occitan in Languedoc". *International Journal of the Sociology of Language*. 2004 (169), pp. 63–89. En: <http://www.reference-global.com/doi/abs/10.1515/ijsl.2004.043>. Consultado el 29 de octubre de 2011.
- Menéndez Pidal, M. (1950). *Orígenes del español*. Madrid.
- Munteanu Colán, D. (2002). Sobre la posición del catalán en el conjunto de la Rumania. *Archivo de Filología Aragonesa*, (1) 59-60.
- Pimienta, D., D. Prado y A. Blanc (2009). Twelve years of measuring linguistic diversity in the Internet: balance and perspective. París: UNESCO. En: <http://unesdoc.unesco.org/images/0018/001870/187016e.pdf>. Consultado el 27 de octubre de 2011.
- Real Academia Española (2001). *Diccionario de la lengua española* (22nda ed.). Disponible en: <http://rae.es/rae.html>.
- Sihler, A. (2000). *Language History: An Introduction*. Amsterdam: John Benjamins Publishing Company.
- Starostin, G. (2010). Preliminary lexicostatistics as a basis for language classification: A new approach. Disponible en: <http://starling.rinet.ru/new100/Lexicostatistics.htm>. Consultado el 20 noviembre de 2011.
- Starostin, G., A. Kassian, A. Dybo y V. Chernov (2010). The Swadesh wordlist. An attempt at semantic specification. Disponible en: <http://starling.rinet.ru/new100/Swadesh.htm>. Consultado el 20 noviembre de 2011.