

# El catalán: ¿iberorromance o galorromance?

*Elber Aguilar Porras*

Escuela de Literatura y Ciencias del Lenguaje

Universidad Nacional, Costa Rica



## **Resumen**

El objetivo de este estudio léxico-estadístico es comparar la relación entre la lengua catalana y las lenguas portuguesa y española del grupo iberorromance, así como del francés y el occitano del grupo galorromance, con el fin de probar su origen genealógico.

**Palabras Claves:** catalán, genealogía, lista de palabras Swadesh, lingüística

## **Abstract**

The aim of this lexico-statistical study research is to compare the relationship between the Catalan language and the Portuguese and Spanish ones –belonging to the Ibero-Romance group– or the French and Occitane languages –both belonging to the Galo-Romance group– in order to prove its genealogical origin.

**Key words:** Catalan, genealogy, the Swadesh list of words, linguistics

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 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 secciones siguientes de este trabajo. Según los objetivos de la lexicostatística, cuya meta es relacionar las lenguas, y por los objetivos de este estudio, 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 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 el 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**

### **Categorización de las lenguas**

#### **El catalán**

El catalán, la cual es una lengua romance que procede del latín vulgar y que forma parte del dialecto de varias regiones de España como lo es Valencia y 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, ocupando 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 y quien, con base en cuestiones de índole histórica, relacionó la lengua con el grupo galorrománico. Muchos han sido los lingüistas que han compartido dicha postura desde entonces; así, 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. A un nivel fonético, podemos mencionar similitudes entre el catalán y sus correspondientes protofonemas con las demás lenguas pertenecientes al grupo iberromance. Así, tenemos correspondencias fonéticas para \*ts donde esta se realiza como /s/ en catalán, español, gallego, aragonés, entre otras lenguas. Léxicamente hablando, formas derivadas del latín como *clave* mantienen una forma exactamente igual en lenguas occitanas como el aragonés y 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 (Ibero o Galorromance), existen aún otras teorías, principalmente la de Bennedet Vidos, que aseguran que el catalán no es ni galorromance ni iberorromance, sino "pirenaico-romance." Dicha afirmación supondría que esta lengua sería más bien un puente entre estos dos grandes grupos y agruparía el catalán con el gascón o el aragonés, por ejemplo, como diferenciado de las demás ramas.

### **Grupo galorrománico**

El grupo galorrománico es el más septentrional de la rama italo-continental. 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 en, al menos, 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 Québécoise 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 (Toulouse, 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, que 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 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 (Boc, 1973): 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 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 de personas. Sin embargo, en estudios más recientes, se asegura que la cifra aumentó a 500 millones de hablantes, los cuales se encuentran 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 personas. A pesar de esto, datos recientes indican que es el idioma oficial de 240 millones de hablantes distribuidos 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 estremenho Extremeño.

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

## **Fundamentos teóricos y metodológicos de la léxicoestadística**

### **La léxicoestadística como teoría**

Para establecer correspondencias entre las lenguas por estudiar, como ya se anunció, se recurrirá en este trabajo a la léxicoestadística, 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, cuerpos celestes o 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 revisten 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 aun 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 considera que las conclusiones de este trabajo son solamente tentativas.

## 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.

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 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:

**Vino:** el catalán está emparentado genealógicamente con todas las lenguas.

**Celeste:** el catalán no está emparentado con ninguna lengua.

**Verde:** el catalán solamente está emparentado con una de las lenguas galorrománicas.

**Naranja:** el catalán solamente está emparentado con una de las lenguas iberorrománicas.

**Café:** el catalán está emparentado con una lengua de cada subgrupo.

**Azul:** el catalán está emparentado solamente con las dos lenguas galorrománicas.

**Rosado:** el catalán está emparentado solamente con las dos lenguas iberorrománicas.

**Gris:** el catalán está emparentado con las dos lenguas iberorrománicas y con una galorrománica.

**Morado:** 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 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	todo	tot	tout	tot
2. and	y	e	i	et	e
3. animal	animal	animal	animal	animal	animal
4. ashes	cenizas	cinzas	cendres	cendre	cendre
5. at	a	a	a	à	a
6. back	espalda	costas	esquena	dos	esquina
7. bad	malo	mau	dolent	mauvais	marrit
8. bark	corteza	casca	escorça	écorce	rusca
9. because	porque	porque	perquè	parce que	perque
10. belly	vientre	ventre	ventre	ventre	ventre
11. big	grande	grande	gran	grand	grand
12. bird	pájaro	pássaro	ocell	oiseau	aucèl

Swadesh	Español	Port	Catalán	Fr	Occitano.
13. to bite	morder	morder	mossegar	mordre	mossegar
14. black	negro	negro	negre	noir	negre
15. blood	sangre	sangue	sang	sang	sang
16. to blow	soplar	soprar	bufar	souffler	bufar
17. bone	hueso	osso	os	os	òs
18. to breathe	respirar	respirar	respirar	respirer	respirar
19. to burn	quemar	queimar	cremar	brûler	cremar
20. child	niño	menino	nen	enfant	enfan
21. cloud	nube	nuvem	núvol	nuage	nívol
22. cold	frío	frio	fred	froid	freg
23. to come	venir	vir	venir	venir	venir
24. to count	contar	contar	comptar	compter	comptar
25. to cut	cortar	cortar	tallar	couper	copar
26. day	día	dia	dia	jour	jorn
27. to die	morir	morror	morir	mourir	morir
28. to dig	cavar	cavar	cavar	creuser	cavar
29. dirty	sucio	sujo	brut	sale	brut
30. dog	perro	cão	gos	chien	can
31. to drink	beber	beber	beure	boire	beure
32. dry	seco	seco	sec	sec	sec
33. dull	romo	rombo	rom	émoussé	bercat
34. dust	polvo	poeira	pols	poussière	pols
35. ear	oreja	orelha	orella	oreille	aurelha
36. earth	tierra	terra	terra	terre	tèrra
37. to eat	comer	comer	menjar	manger	manjar
38. egg	huevo	ovo	ou	œuf	uòu
39. eye	ojo	olho	oill	œil	olh
40. to fall	caer	cair	caure	tomber	tombar
41. far	lejos	longe	lluny	loin	luenh
42. fat	grasa	gordura	greix	grasse	grais
43. father	padre	pai	pare	père	paire

Swadesh	Español	Port	Catalán	Fr	Occitano.
44. to fear	temer	temer	témer	craindre	crentar
45. feather	pluma	pluma	ploma	plume	pluma
46. few	poco	pouco	poc	peu	pauc
47. to fight	luchar	lutar	lluitar	se battre	se batre
48. fire	fuego	fogo	foc	feu	fuòc
49. fish	pez	peixe	peix	poisson	peis
50. five	cinco	cinco	cinc	cinq	cinc
51. to float	flotar	flutuar	flotar	flotter	flotar
52. to flow	fluir	fluir	fluir	couler	rajar
53. flower	flor	flor	flor	fleur	flor
54. to fly	volar	voar	volar	voler	volar
55. fog	niebla	névoa	boira	brouillard	nèbla
56. foot	pie	pé	peu	pied	pè
57. four	cuatro	quatro	quatre	quatre	quatre
58. to freeze	helar	gelar	gelar	geler	gelar
59. fruit	fruta	fruta	fruita	fruit	fruch
60. to give	dar	dar	donar	donner	donar
61. good	bueno	bom	bo	bon	bon
62. grass	hierba	erva	herba	herbe	èrba
63. green	verde	verde	verd	vert	verd
64. guts	entrañas	entranhas	budellada	boyaux	budelada
65. hair	pelo	pêlo	pèl	cheveux	pel
66. hand	mano	mão	mà	main	man
67. he	él	ele	ell	il	el
68. head	cabeza	cabeça	cap	tête	cap
69. to hear	oír	ouvir	oir	entendre	ausir
70. heart	corazón	coração	cor	cœur	còr
71. heavy	pesado	pesado	pesat	lourd	pesuc
72. here	aquí	aqui	ací	ici	aicí
73. to hit	golpear	bater	copejar	frapper	tustar
74. to hold	sostener	segurar	sostenir	tenir	tener

Swadesh	Español	Port	Catalán	Fr	Occitano.
75. how	cómo	como	com	comment	quinament
76. to hunt	cazar	caçar	caçar	chasser	cassar
77. husband	marido	marido	marit	mari	marit
78. i	yo	eu	jo	je	eu
79. ice	hielo	gelo	gel	glace	glaça
80. if	si	se	si	si	se
81. in	en	em	dins	dans	dins
82. to kill	matar	matar	matar	tuer	tuar
83. know	saber	saber	saber	savoir	saber
84. lake	lago	lago	llac	lac	lac
85. to laugh	reír	rir	riurer	rire	rire
86. leaf	hoja	folha	fulla	feuille	fuelha
87. left	izquierda	esquerda	esquerra	gauche	esquerra
88. leg	pierna	perna	gamba	jambe	camba
89. lie	yacer	jazer	jeure	gésir	jàser
90. to live	vivir	viver	viurer	vivre	viure
91. liver	hígado	figado	fetge	foie	fetge
92. long	largo	longo	llarg	long	larg
93. louse	piojo	piolho	poll	pou	pesolh
94. man	hombre	homem	home	homme	òme
95. many	muchos	muitos	molt	beaucoup	fòrça
96. meat	carne	carne	carn	viande	carn
97. mother	madre	mãe	mare	mère	maire
98. mountain	montaña	montanha	muntanya	montagne	montanha
99. mouth	boca	boca	boca	bouche	boca
100. name	nombre	nome	nom	nom	nom
101. narrow	estrecho	estreito	estret	étroit	estrech
102. near	cerca	perto	prop	près	probet
103. neck	cuello	pescoço	coll	cou	còl
104. new	nuevo	novo	nou	nouveau	nòu
105. night	noche	noite	nit	nuit	nuech

Swadesh	Español	Port	Catalán	Fr	Occitano.
106. nose	nariz	nariz	nas	nez	nas
107. not	no	não	no	ne pas	pas
108. old	viejo	velho	vell	vieux	vièlh
109. one	uno	um	un	un	un
110. other	otro	outro	altre	autre	autre
111. person	persona	persona	persona	personne	persona
112. to play	jugar	jogar	jugar	jouer	jogar
113. to pull	tirar	puxar	tirar	tirer	tirar
114. to push	empujar	empurrar	empènyer	pousser	butar
115. to rain	llover	chover	ploure	pleuvoir	plover
116. red	rojo	vermelho	vermell	rouge	roge
117. right	correcto	correcto	correcte	correct	corrècte
118. right	derecha	direita	dreta	droite	drecha
119. river	río	rio	riu	fleuve	riu
120. road	camino	estrada	camí	route	camin
121. root	raíz	raiz	arrel	racine	rasiga
122. rope	cuerda	corda	corda	corde	còrda
123. rotten	podrido	podrido	podrit	pourri	poirit
124. rub	frotar	esfregar	fregar	frotter	fretar
125. salt	sal	sal	sal	sel	sal
126. sand	arena	areia	sorra	sable	sabla
127. to say	decir	dizer	dir	dire	dire
128. scratch	rascar	coçar	gratar	gratter	gratar
129. sea	mar	mar	mar	mer	mar
130. to see	ver	ver	veure	voir	vèser
131. seed	semilla	semente	llavor	graine	grana
132. to sew	coser	cozer	cosir	coudre	cóser
133. sharp	afilado	afiado	afilat	tranchant	trencant
134. short	corto	curto	curt	court	cort
135. to sing	cantar	cantar	cantar	chanter	cantar
136. to sit	sentarse	sentar-se	asseure	s'asseoir	s'assetar

Swadesh	Español	Port	Catalán	Fr	Occitano.
137. skin	piel	pele	pell	peau	pèl
138. sky	cielo	céu	cel	ciel	cèl
139. to sleep	dormir	dormir	dormir	dormir	dormir
140. small	pequeño	pequeno	petit	petit	pichon
141. to smell	oler	cheirar	olorar	sentir	sentir
142. smoke	humo	fumaça	fum	fumée	fum
143. smooth	liso	liso	llis	lisse	lis
144. snake	serpiente	serpente	serp	serpent	sèrp
145. snow	nieve	neve	neu	neige	nèu
146. some	algunos	alguns	alguns	quelques	qualques
147. to spit	escupir	cuspir	escopir	cracher	escupir
148. to split	partir	partir	partir	fendre	fendre
149. to squeeze	apretar	apertar	apretar	presser	estrenher
150. to stab	apuñalar	apunhalar	apunyalar	poignarder	emponhalar
151. to stand	estar de pie	estar em pé	estar de peu	se lever	se levar
152. star	estrella	estrela	estrella	étoile	estela
153. stick	palo	pau	pal	bâton	bròc
154. stone	piedra	pedra	pedra	Pierre	pèira
155. straight	recto	reto	recte	droit	drech
156. to suck	chupar	chupar	xuclar	sucer	chucar
157. sun	sol	sol	sol	soleil	solelh
158. to swell	hincharse	inchar-se	inflar	se gonfler	conflar
159. to swim	nadar	nadar	nedar	nager	nadar
160. tail	cola	cauda	cua	queue	coa
161. that	ese	esse	aquest	cet	aqueste
162. there	ahí	aí	allà	là	aquí
163. they	ellos	eles	ells	ils	eles
164. thick	espeso	espesso	espès	épais	espés
165. thin	delgado	delgado	prim	mince	prim
166. to think	pensar	pensar	pensar	penser	pensar
167. this	este	este	aquest	ce	aqueste

Swadesh	Español	Port	Catalán	Fr	Occitano.
168. thou	tú	tu	tu	tu	tu
169. three	tres	três	tres	trois	tres
170. to throw	lanzar	lançar	llançar	lancer	lançar
171. to tie	atar	atar	lligar	attacher	estacar
172. tongue	lengua	língua	llengua	langue	lenga
173. tooth	diente	dente	dent	dent	dent
174. tree	árbol	árvore	arbre	arbre	arbre
175. to turn	girar	girar	girar	tourner	virar
176. two	dos	dois	dos	deux	dos
177. to vomit	vomit	vomitar	vomitar	vomir	vomir
178. to walk	caminar	caminhar	caminar	marcher	caminar
179. warm	cálido	cálido	càlid	chaud	caud
180. to wash	lavar	lavar	rentar	laver	lavar
181. water	agua	água	aigua	eau	aiga
182. we	nosotros	nós	nosaltres	nous	nosautres
183. wet	mojado	molhado	mullat	mouillé	moth
184. what?	qué	que	què	quel	qué
185. when?	cuándo	quando	quan	quand	quand
186. where?	dónde	onde	on	où	ont
187. white	blanco	branco	blanc	blanc	blanc
188. who?	quién	quem	qui	qui	qui
189. wide	ancho	largo	ample	large	larg
190. wife	esposa	esposa	dona	épouse	esposada
191. wind	viento	vento	vent	vent	vent
192. wing	ala	asa	ala	aile	ala
193. wipe	limpiar	enxugar	eixugar	essuyer	eissugar
194. with	con	com	amb	avec	amb
195. woman	mujer	mulher	dona	femme	femna
196. woods	bosque	bosque	bosc	bois	bòsc
197. worm	gusano	verme	cuc	ver	vèrme

Swadesh	Español	Port	Catalán	Fr	Occitano.
198. ye	vosotros	vocês	vosaltres	vous	vosautres
199. year	año	ano	any	année	annada
200. yellow	amarillo	amarelo	groc	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	125	63,45%
Catalán no está emparentado con ninguna	11	5,59%
Catalán-1 lengua galorrománica	8	4,06%
Catalán- 1 lengua iberorrománica	6	3,04%
Catalán-1 galorrománica y 1 iberorrománica	1	0,5%
Catalán- 2 galorrománicas	8	4,06%
Catalán – 2 iberorrománicas	19	9,64%
Catalán- 2iberorrománicas-1 galorrománica	14	7,1%
Catalán- 2 galorrománicas- 1 iberorrománica	5	2,53%
<b>TOTAL</b>	<b>197</b>	<b>99,97%</b>

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 azul y verde, que excluyen al español y al portugués, pero relacionan al catalán con francés u occitano, dan como resultado un

8,12%. La contraparte iberorrománica, es decir, la suma de las claves naranja y rosado, da un 12,7%, resultado poco diferente al del galorromance.

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, 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, 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 iberorromance, 2,4% están relacionadas únicamente con el español. De nuevo, por ende, no se ve una mayor cercanía con una de las dos lenguas con las que se suele agrupar el catalán.

Se concluye, por ende, que el método lexicoestadístico no permite llegar a un criterio 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.

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 (9<sup>a</sup>.. ed.) Disponible en: <http://www.academie-francaise.fr/dictionnaire/>
- Bec, P. (1973). *La Langue Occitane*. 3<sup>a</sup> edición. París: Presses Universitaires de France.
- 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*. 2<sup>a</sup>.edición. Barcelona.
- Lehmann, W. (1969). *Introducción a la Lingüística Histórica*. Madrid: Gredos.
- Lodge, A. (2010). The Sources of Standardisation in French – Written or Spoken? en Ingham, R. (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 Romania. *Archivo de Filología Aragonesa*. (1) 59-60.
- Ortega Caveró, D. (1966). *Diccionario español-portugués*. Barcelona: Editorial Ramón Sopena.
- Pimienta, D., Prado, D., y Blanc, A. (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., Kassian, A., Dybo, A. y Chernov V. (2010). The Swadesh wordlist. An attempt at semantic secification. Disponible en: <http://starling.rinet.ru/new100/Swadesh.htm>. Consultado el 20 noviembre de 2011.