



UNIVERSITAT DE  
BARCELONA

## Las pirámides etnolingüísticas

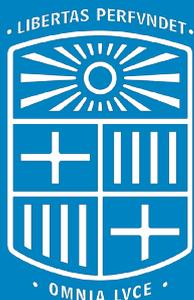
### Estudio contrastivo entre el español y el italiano basado en corpus

Francesco Spinoglio

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE  
BARCELONA

**TESIS DOCTORAL**  
**LAS PIRÁMIDES**  
**ETNOLINGÜÍSTICAS**

**Estudio contrastivo entre el español y el italiano basado en corpus**

**FACULTAT DE FILOLOGIA**  
**Tesis para optar al grado de Doctor presentada por**

**Francesco Spinoglio**

**Directora: Dra. Mar Cruz Piñol**  
**Tutora: Dra. Isabel Verdaguer Clavera**

**Barcelona, 2017**



**Programa de Doctorado: Estudis lingüístics, literaris i culturals**  
**Línea de investigación: Lèxic, comunicació lingüística i cognició**

# **LAS PIRÁMIDES ETNOLINGÜÍSTICAS**

**Estudio contrastivo entre el español y el italiano basado en corpus**

**Francesco Spinoglio**

**Directora: Dra. Mar Cruz Piñol, Universitat de Barcelona**  
**Tutora: Dra. Isabel Verdaguer Clavera, Universitat de Barcelona**



# ÍNDICE

## 1. PRESENTACIÓN DEL TRABAJO Y ESTRUCTURA DE LA TESIS

- 1.1 Trayectoria personal y motivación. **12**
- 1.2 Preguntas de partida y objetivos de la investigación. **15**
- 1.3 Organización del trabajo. **18**
- 1.4 Metodología. **20**

## PARTE I

### 2. Las huellas de la cultura en la lengua

- 2.1. Sobre el término “etnolingüística”. **28**
- 2.2. Sobre la cultura y la lengua. **30**
  - 2.2.1. Rasgos universales del lenguaje humano. **32**
  - 2.2.2. El relativismo lingüístico y la teoría de Sapir-Whorf. **34**
  - 2.2.3. El relativismo pragmático de Dell Hymes. **36**
  - 2.2.4. El idealismo lingüístico de Karl Vossler. **37**
- 2.3. Las neuronas espejo y su función en el cerebro. **39**
  - 2.3.1. Diferentes tipos de neuronas espejo. **40**
  - 2.3.2. Los medios de comunicación y la publicidad. **41**
  - 2.3.3. El poder de los memes. **43**
  - 2.3.4. Neuroeducación. **45**
- 2.4 Algunos aspectos semánticos en las lenguas neolatinas. **46**
  - 2.4.1. Extensión vs intensidad. **47**
  - 2.4.2. La intensidad. **48**
  - 2.4.3. Sinonimia y polisemia. **49**
  - 2.4.4. El *brainstorming*. **51**

### 3. Tipos de corpus y su papel en las pirámides etnolingüísticas

- 3.1. La lingüística de corpus. **56**
  - 3.1.1. Utilidad y función principal de los corpus. **57**
  - 3.1.2. Diferentes tipos de corpus. **58**
  - 3.1.3. La competencia comunicativa en los corpus. **60**
  - 3.1.4. El enfoque inductivo y la motivación. **61**
- 3.2. Principales corpus del español que se van a utilizar en esta tesis. **63**
  - 3.2.1. El CORPES. **63**
  - 3.2.2. El corpus Val.Es.Co. **65**
- 3.3. Principales corpus del italiano que se van a utilizar en esta tesis. **68**
  - 3.3.1. El CORIS/CODIS. **68**
  - 3.3.2. El corpus LIP. **71**
- 3.4. La red como corpus. **74**
- 3.5. Resumen de la PARTE I. **76**

## PARTE II

### 4. Las pirámides etnolingüísticas

- 4.1. El concepto de "pirámide etnolingüística". **82**
- 4.2. El proceso de creación de las pirámides etnolingüísticas: tres ejemplos para el español. **84**
  - 4.2.1. Fase 1: el *brainstorming*. **84**
    - 4.2.1.1. Listado obtenido tras la Fase 1 para el ámbito de la religión. **85**
    - 4.2.1.2. Listado obtenido tras la Fase 1 para el ámbito del fútbol. **87**
    - 4.2.1.3. Listado obtenido tras la Fase 1 para el ámbito de los toros. **89**
  - 4.2.2. Fase 2: la obtención del listado de trabajo: criba, metodología y criterios. **90**
    - 4.2.2.1. Listado definitivo tras la Fase 2 para el ámbito de la religión. **93**
    - 4.2.2.2. Listado definitivo tras la Fase 2 para el ámbito del fútbol. **95**
    - 4.2.2.3. Listado definitivo tras la Fase 2 para el ámbito de los toros. **96**
  - 4.2.3. Fase 3: la búsqueda en los corpus: de las ocurrencias a las expresiones. **97**
    - 4.2.3.1. Fase 3 con el listado de la religión en español. **97**
    - 4.2.3.2. Fase 3 con el listado del fútbol en español. **107**
    - 4.2.3.3. Fase 3 con el listado de los toros en español. **114**
  - 4.2.4. Fase 4: selección y ordenación de las expresiones dentro de las pirámides. **119**
    - 4.2.4.1. La pirámide etnolingüística de la religión en español. **119**
    - 4.2.4.2. La pirámide etnolingüística del fútbol en español. **126**
    - 4.2.4.3. La pirámide etnolingüística de los toros en español. **129**
- 4.3. El proceso de creación de las pirámides etnolingüísticas: tres ejemplos para el italiano. **132**
  - 4.3.1. Fase 1: el *brainstorming*. **132**
    - 4.3.1.1. Listado obtenido tras la Fase 1 para el ámbito de la religión. **132**
    - 4.3.1.2. Listado obtenido tras la Fase 1 para el ámbito del fútbol. **134**
    - 4.3.1.3. Listado obtenido tras la Fase 1 para el ámbito de los toros. **136**
  - 4.3.2. Fase 2: la obtención del listado de trabajo: criba, metodología y criterios. **138**
    - 4.3.2.1. Listado definitivo tras la Fase 2 para el ámbito de la religión. **140**
    - 4.3.2.2. Listado definitivo tras la Fase 2 para el ámbito del fútbol. **141**
    - 4.3.2.3. Listado definitivo tras la Fase 2 para el ámbito de los toros. **142**
  - 4.3.3. Fase 3: la búsqueda en los corpus: de las ocurrencias a las expresiones. **143**
    - 4.3.3.1. Fase 3 con el listado de la religión en italiano. **143**
    - 4.3.3.2. Fase 3 con el listado del fútbol en italiano. **151**
    - 4.3.3.3. Fase 3 con el listado de los toros en italiano. **157**
  - 4.3.4. Fase 4: selección y ordenación de las expresiones dentro de las pirámides. **159**
    - 4.3.4.1. La pirámide etnolingüística de la religión en italiano. **159**
    - 4.3.4.2. La pirámide etnolingüística del fútbol en italiano. **166**
    - 4.3.4.3. La pirámide etnolingüística de los toros en italiano. **168**
- 4.4. El análisis contrastivo entre el español y el italiano a partir de la visualización de datos en las pirámides etnolingüísticas. **169**

- 4.4.1. Las pirámides etnolingüísticas de la religión en contraste. **169**
- 4.4.2. Las pirámides etnolingüísticas del fútbol en contraste. **178**
- 4.4.3. Las pirámides etnolingüísticas de los toros en contraste. **181**
- 4.5. Resumen de la PARTE II. **184**

## **PARTE III**

### **5. Consideraciones finales**

- 5.1. Aportaciones y limitaciones de la investigación. **190**
- 5.2. Conclusiones. **193**
- 5.3. Líneas abiertas de investigación. **195**

## **BIBLIOGRAFÍA**

# ÍNDICE DE IMÁGENES Y GRÁFICOS

Imagen 1. Consulta de la expresión *llamar al pan pan y al vino vino* en Google (23 de septiembre de 2015). **44**

Imagen 2. Consulta de la expresión *dire pane al pane e vino al vino* en Google (23 de septiembre de 2015). **44**

Imagen 3. Interfaz del CORPES (15 de mayo de 2016). **64**

Imagen 4. Consulta de la palabra *toro* en el CORPES (15 de mayo de 2016). **65**

Imagen 5. Interfaz del corpus Val.Es.Co. (15 de mayo de 2016). **66**

Imagen 6. Conversaciones del corpus Val.Es.Co. (15 de mayo de 2016). **66**

Imagen 7. Consulta de la expresión *dios* en el corpus Val.Es.Co. (15 de mayo de 2016). **67**

Imagen 8. Interfaz del CORIS. (16 de mayo de 2016). **69**

Imagen 9. Interfaz del CODIS. (16 de mayo de 2016). **69**

Imagen 10. Marcas del CORIS/CODIS. (16 de mayo de 2016). **70**

Imagen 11. Marcas del CORIS con la palabra *toro*. (16 de mayo de 2016). **70**

Imagen 12. Página principal de BADIP. (16 de mayo de 2016). **71**

Imagen 13. Motor de búsqueda del corpus LIP. (16 de mayo de 2016). **72**

Imagen 14. Consulta de la palabra *dio* en el corpus LIP. (16 de mayo de 2016). **72**

Imagen 15. Consulta de la expresión *Dios los cría y ellos se juntan* en Google (18 de mayo de 2016). **74**

Imagen 16. Consulta de la palabra *dios* en Google Ngram Viewer. (18 de mayo de 2016). **75**

Imagen 17. Consulta de la palabra *parar* en el CORPES. (28 de septiembre de 2016). **112**

Gráfico 1. Pirámide 1/3 de la religión en español. Expresiones rentables (++) y frecuentes. **124**

Gráfico 2. Pirámide 2/3 de la religión en español. Expresiones rentables (++) pero menos frecuentes. **125**

Gráfico 3. Pirámide 3/3 de la religión en español. Expresiones coloquiales (+) y vulgares (-). **126**

Gráfico 4. Pirámide 1/1 del fútbol en español. **128**

Gráfico 5. Pirámide 1/1 de los toros en español. **130**

Gráfico 6. Pirámide 1/2 de la religión en italiano. Expresiones rentables (++) y frecuentes. **164**

Gráfico 7. Pirámide 2/2 de la religión en italiano. Expresiones rentables (++) , pero menos frecuentes y expresiones coloquiales (+). **165**

Gráfico 8. Pirámide 1/1 del fútbol en italiano. **167**

Gráfico 9. Pirámide 1/1 de los toros en italiano. **168**

Gráfico 10. Pirámide 1/3 de la religión en español. Expresiones rentables (++) y frecuentes. **170**

Gráfico 11. Pirámide 1/2 de la religión en italiano. Expresiones rentables (++) y frecuentes. **171**

Gráfico 12. Pirámide 2/3 de la religión en español. Expresiones rentables (++) pero menos frecuentes. **172**

Gráfico 13. Pirámide 2/2 de la religión en italiano. Expresiones rentables (++) , pero menos frecuentes y expresiones coloquiales (+). **173**

Gráfico 14. Pirámide 3/3 de la religión en español. Expresiones coloquiales (+) y vulgares (-). **174**

Gráfico 15. Pirámide 1/1 del fútbol en español. **178**

Gráfico 16. Pirámide 1/1 del fútbol en italiano. **179**

Gráfico 17. Pirámide 1/1 de los toros en español. **182**

Gráfico 18. Pirámide 1/1 de los toros en italiano. **182**

## Agradecimientos

A lo largo de estos tres años han sido muchas las personas que han contribuido directa o indirectamente en este trabajo y a quienes quiero expresar mi gratitud por su apoyo y sobre todo por la confianza que han depositado en mí y en este proyecto. En primer lugar, quiero agradecer a la Universitat de Barcelona que me haya ofrecido la posibilidad de realizar mi labor investigadora y me siento honrado de haber pasado estos años en una institución tan prestigiosa que me ha hecho sentir como en casa desde el primer día. Debo un gran reconocimiento a todo el Departamento de Filología Hispánica, con mención especial para la Dra. Paloma Arroyo Vega por apoyarme siempre en los momentos de dificultad y por contagiarme su pasión por la enseñanza, al Dr. Antonio Torres Torres por su disponibilidad y su gran sentido del humor y al Dr. Juan Pablo García-Borrón por sus clases magistrales de semántica que tanto me han inspirado para la realización de esta tesis. Además, les estoy muy agradecido por su tiempo y por haberse prestado voluntarios para la realización del *brainstorming* en español. En este sentido, también me gustaría agradecer a todas las personas que me han facilitado datos lingüísticos para la realización de las pirámides etnolingüísticas, tanto para el italiano como para el español.

Además, quiero dar las gracias al Dr. Joseph Hilferty por haberme introducido en el maravilloso mundo de las neuronas espejo, pilar de esta investigación, al Dr. José Daniel Barquero Cabrero, director de Eserp Business School, por sus preciados consejos y por haberme animado a realizar la tesis doctoral, a Montse Morante Vallejo (CRAI) por su apoyo en el ámbito bibliográfico y por supuesto al Dr. Aldo Ruffinatto, de quien recibí unas clases inolvidables de *Literatura española* durante mi primer año de carrera en la Università di Lettere de Turín. Gracias, Aldo, por tus recomendaciones y por enseñarme a creer a fondo en mis capacidades; sabes muy bien que tuviste una gran influencia en el viaje vital que emprendí hace ya catorce años, y jamás lo olvidaré.

Deseo expresar todo mi agradecimiento a mi familia, especialmente a María, mi inseparable compañera, por su apoyo incondicional a lo largo de este trayecto. En este sentido, también quiero dar las gracias a mis amigos y a mis compañeros de trabajo que, de un modo u otro, han respaldado este esfuerzo.

Por último, quiero expresar mi más sincero agradecimiento a la Dra. Mar Cruz Piñol y a la Dra. Isabel Verdaguer Clavera, respectivamente directora y tutora de esta tesis. Trabajar con vosotras ha sido una experiencia maravillosa que guardaré para siempre en el baúl de los recuerdos. Mar, gracias, por tu paciencia, por tus sugerencias y por estar disponible siempre que te he necesitado. Has sido una compañera de viaje formidable y sin ti todo esto no hubiera sido posible. Isabel, gracias también por tu apoyo y por tu energía positiva que me ha acompañado a lo largo de estos años. Como decía Machado, *Caminante, no hay camino, se hace camino al andar*. Y me permito añadir que “ese camino se hace mucho más ameno si caminan a tu vera personas tan profesionales como vosotras”. Espero que esto sea solo el primer paso y que sigamos colaborando en un futuro. Gracias de corazón.

## Resumen

En esta tesis de investigación, de carácter interlingüístico por ahondar en los contrastes entre lenguas e interdisciplinar por acercarse a la neurociencia, se realiza un estudio contrastivo desde la perspectiva de la etnolingüística entre el español y el italiano, y se propone una visualización de datos que hemos denominado "pirámides etnolingüísticas".

Para llevar a cabo el estudio, se han escogido tres ámbitos culturales concretos (la religión, el fútbol y los toros) y se han formulado tres hipótesis de partida: en primer lugar, **que es en las expresiones donde se puede apreciar la influencia de la cultura en la lengua**; en segundo lugar, **que es posible utilizar los corpus para demostrar que la afinidad etnolingüística entre dos lenguas aparentemente similares como el español y el italiano es solo parcial**, y, por último, **que con ámbitos culturales no compartidos, como en el caso de los toros entre el español y el italiano, la visualización de las expresiones presenta grandes diferencias**.

En lugar de buscar directamente las expresiones en diccionarios especializados, se ha optado por seguir el método inductivo Data-driven, que consiste en extraer conclusiones y demostrar las hipótesis iniciales a partir del análisis de datos, eje vertebrador de toda la investigación. Para ello, se ha utilizado una metodología mixta basada en la triangulación de datos que consta de tres pasos: la **realización de un *brainstorming*** para obtener el mayor número de palabras, la **búsqueda en los corpus** para ir de las ocurrencias a las expresiones y la **visualización de datos en las pirámides** para ordenar y analizar las expresiones seleccionadas. Con esta metodología se aspira a **encontrar el mayor número de expresiones relacionadas con el ámbito cultural en cuestión (objetivo A), ordenar y visualizar los datos obtenidos dentro de unas pirámides para realizar un primer análisis contrastivo entre el español y el italiano (objetivo B)** y, por último, **demostrar que la cultura influye de manera diferente en cada lengua, incluso con ámbitos culturales afines, ya que depende de la creatividad de los hablantes, de la imitación del propio ser humano gracias a la función de las neuronas espejo y también de la historicidad (objetivo C)**.

En definitiva, con las pirámides etnolingüísticas se pretende reflejar aspectos de la cosmovisión colectiva en español y en italiano a partir de una visualización de datos que puede ayudar a entender mejor a "los otros", facilitando la inmersión cultural y el aprendizaje de una L2.

## Abstract

In this thesis, that explores the contrast between languages with the approach to the neuroscience, we realize a contrastive study in the field of the ethnolinguistics between Spanish and Italian, offering a visualization of data that we have named "ethnolinguistic pyramids".

In order to carry out this study, three specific cultural spheres (religion, football and bulls) have been chosen and three basic hypotheses have been formulated: firstly, **there are expressions where one can appreciate the influence of culture in the language**; secondly, **that it is possible to use corpora to show that the ethnolinguistic affinity between two similar languages, such as Spanish and Italian, is only partial**, and, finally, **with non-shared cultural spheres, as is the case of bulls in Spanish and Italian, the display of expressions presents significant differences**.

Instead of looking directly for expressions in specialized dictionaries, we have chosen to follow the inductive Data-driven method by a drawing conclusions and demonstrating the initial hypotheses from data analysis, the backbone of all research. To this end, a mixed methodology based on triangulation of data has been used, consisting of three steps: **brainstorming** to obtain the highest number of words, **corpora search** to go from occurrences to expressions and **visualization of data in the pyramids** to sort and analyze the selected expressions. With this methodology, we aim **to find the greatest number of expressions related to the cultural sphere in question (objective A), to order and visualize the data obtained within a pyramid to perform a contrastive analysis between Spanish and Italian (objective B)** and, finally, **to show that culture influences each language differently, even between related cultural spheres, since it reflects the creativity of the speakers and the imitation processes of the human being thanks to the mirror neurons and also the historicity (objective C)**.

Finally, the "ethnolinguistic pyramids" are intended to reflect aspects of the collective worldview in Spanish and Italian, identifiable from a data visualization that can help to better understand "the other peoples", facilitating cultural immersion and L2 learning.



# **1. PRESENTACIÓN DEL TRABAJO Y ESTRUCTURA DE LA TESIS**

## 1.1. TRAYECTORIA PERSONAL Y MOTIVACIÓN

En junio de 2010, tras finalizar la carrera de Filología Hispánica en la Universitat de Barcelona, decidí seguir con mis estudios y, después de sopesar atentamente la propuesta académica que ofrecía la UB, me decanté por el Máster en Léxico y Comunicación Lingüística, impulsado sobre todo por la variedad temática que ofrecía. Ahora, con el paso de los años, me doy cuenta de que para mí siempre ha sido motivo de reflexión, en el embarullado mundo de la educación, el hecho de que nadie que matricule una carrera, un máster o simplemente un curso de formación sepa a ciencia cierta con qué se va a encontrar. Sabemos qué temario vamos a tener y qué argumentos en principio vamos a tratar, pero desconocemos a los profesores y es prácticamente imposible imaginarse quiénes te van a inspirar y sobre todo qué ideas pueden llegar a despertar en tu interior. Dicho de otra forma, la clase que imparte un solo profesor es la misma para todos los alumnos, pero cada uno se quedará con lo que más le emocione y despierte interés. En mi caso, hubo cinco profesores que me inspiraron muchísimo a lo largo de mi trayectoria académica y sin su aportación didáctica esta tesis no habría sido posible. Han sido la Dra. Paloma Arroyo Vega con *Historia de la Lengua española y Fonética española*, el Dr. Antonio Torres Torres con *El Español de América* y su valiosa información acerca de Dispox, el Dr. Juan Pablo García-Borrón con sus magistrales clases de *Semántica española*, la Dra. Mar Cruz Piñol con *La lingüística de corpus* y el Dr. Joseph Hilferty con *La lingüística cognitiva*. El Dr. Hilferty, en concreto, me introdujo en el mundo de las neuronas espejo, las responsables de la empatía y de la imitación, y a partir de ahí algo se activó en mi interior. Amplié mis lecturas y me topé con *Introduzione all'Etnolingüística* de Giorgio Raimondo Cardona, un manual de cabecera sobre una rama, en palabras del mismo autor, “bastante ninguneada de la lingüística que estudia cómo influye la cultura en una lengua y qué lazos establece con ella” (*Introduzione all'Etnolingüística*, 4)<sup>1</sup>; además, leí *Neuronas espejo* de Marco Iacoboni, un libro sobre estas neuronas que nos permiten ponernos en la piel del otro. En ese momento me parecieron dos temáticas curiosas que brillaban por el escaso interés que habían despertado hasta el momento en el mundo académico, así que empecé a atar cabos sueltos y me puse a explorar las ideas de la etnolingüística en paralelo con las neuronas espejo. El problema es que no conseguía aterrizar la idea, hasta que me acordé de la Lingüística de Corpus y decidí ir a ver a la Dra. Mar Cruz Piñol. Tras exponerle mis ideas y mis inquietudes, debatimos sobre si sería interesante sondear el camino para disponer de un corpus etnolingüístico español-italiano. Pese a la enorme expectativa que esta idea generó en mí, el proyecto no acabó de cuajar y además no veía cómo podía hilvanarlo con mi interés por las neuronas espejo, de modo que lo dejé apartado durante dos años, hasta que en mayo de 2014 decidí matricularme en un nuevo programa de doctorado llamado *Estudis Lingüistics, Literaris i Culturals*, siguiendo la línea de investigación *Lèxic, comunicació lingüística i cognició*. Cabe decir que, una vez terminado el Máster en 2011, empecé a ejercer como profesor de español para extranjeros (ELE) en diferentes academias y universidades y, conforme iba ganando experiencia y confianza en las clases, fui experimentando cada vez más con la etnolingüística, las neuronas espejo y los

1. La traducción del italiano al español es propia.

corpus hasta que, casi de manera accidental, nació la idea de las pirámides etnolingüísticas. A raíz de esto, cuando volví a ver a la Dra. Mar Cruz Piñol en 2014 para matricular la tesis, mis ideas estaban mucho más claras y por fin tenía la confianza necesaria para emprender el viaje.

Tres años más tarde, la tesis que aquí se presenta, de carácter interlingüístico por ahondar en los contrastes entre lenguas e interdisciplinar por acercarse a la traducción y a la neurociencia, es el resultado de todas estas ideas y de mi trayectoria, que se ha ido desgranando gracias sobre todo a las oportunidades profesionales y académicas que se me han presentado por el camino. En concreto, mediante el presente trabajo se aspira a elaborar un estudio sistemático en el campo de la etnolingüística, ahondando en las relaciones entre dos lenguas afines como son el español y el italiano, además de ofrecer, con la representación visual de las pirámides etnolingüísticas, una nueva herramienta que facilite la **visualización de datos** de cara a la realización de análisis contrastivos y a la inmersión cultural. En este mundo dominado por la tecnología, la **visualización** ha ganado mucho protagonismo y puede ser muy efectiva para mejorar y acelerar el aprendizaje, como expone Robert Spence en su libro *Information visualization: design for interaction*, dedicado enteramente a cómo se puede optimizar el diseño de cara a una mejor visualización de los datos (*Information visualization: design for interaction*, 5-11), y también como señala Colin Ware, quien hace particular hincapié en la percepción que despierta el diseño (*Information visualization: perception for design*, 4). Por eso mismo, en esta tesis se ha decidido **apostar por un sistema de visualización que denominaremos "pirámides etnolingüísticas"** a partir de un proceso de 4 fases que se detallarán de forma pormenorizada en 4.2 y 4.3.

El **marco teórico** de esta investigación, como se verá en el capítulo 2, bebe de estudios, ensayos y tesis doctorales sobre la etnolingüística, la neurociencia y la semántica; pero, por encima de todo, aborda la relación lengua-cultura para tratar de entender hasta qué punto determinados factores pueden influir en la manera de hablar de un nativo (en particular, en el uso recurrente de determinadas expresiones), de ahí que se haya escogido la etnolingüística como marco teórico principal de esta investigación. Asimismo, desde el punto de vista metodológico, el estudio no hubiera sido posible sin la aportación de los corpus (capítulo 3), que sirven, entre otras cosas, para corroborar el uso real de una lengua, saber cuáles son las expresiones más frecuentes<sup>2</sup> y conocer el alcance verdadero de una determinada palabra. Para ello, se ha decidido utilizar, como se verá en detalle en el apartado 1.4, una estrategia metodológica múltiple o mixta para corroborar las hipótesis y poder manejar tanto la información cuantitativa como la cualitativa con el fin de intentar alcanzar los objetivos principales de esta investigación, que se expondrán en el siguiente apartado.

Esta tesis parte de un estudio contrastivo entre el español y el italiano<sup>3</sup> para ofrecer un nuevo punto de vista respecto a la relación lengua-cultura mediante la visualización de las expresiones encontradas en los corpus y ordenadas en las pirámides en base a unos criterios que se expondrán a lo largo de la *Parte II*. Así pues, la creación de las pirámides etnolingüísticas

2. Nos interesan las expresiones porque tenemos la hipótesis de que en ellas es donde más se ve reflejada la influencia de la cultura, como se expondrá en el siguiente apartado.

3. Se ha escogido el italiano por ser mi lengua materna y el español por ser mi lengua de acogida que utilizo como L1 desde hace catorce años.

representa una propuesta lingüística y didáctica para comprender más a fondo la influencia cada vez más arraigada de la cultura en la lengua, las expresiones que de ello se derivan y la relación que establecen con el contexto comunicativo en el que se insertan.

## 1.2. PREGUNTAS DE PARTIDA Y OBJETIVOS DE LA INVESTIGACIÓN

La idea de esta tesis de investigación surgió, como se ha expuesto en el anterior apartado, a partir de un interés por la etnolingüística, una rama de la lingüística que se centra en la influencia de la cultura en la lengua y que, hoy en día, todavía no ha recibido la suficiente atención por parte de la comunidad investigadora. A partir de aquí, se han formulado **dos primeras preguntas**:

- 1) **Cómo se podría utilizar la etnolingüística para cotejar dos lenguas aparentemente afines como el español y el italiano;**
- 2) **Cómo se manejaría la información cuantitativa y cualitativa para visualizar los resultados de la investigación.**

Para contextualizar la primera pregunta, se escogerán tres ámbitos culturales concretos para establecer un punto de partida: **la religión, el fútbol y los toros**. Dicha elección no es aleatoria y servirá para cotejar (4.4) dos ámbitos culturales compartidos entre España e Italia y uno más idiosincrásico, en principio, solo de la cultura española (los toros). Para llegar a esta comparación, será menester acceder a una gran cantidad de datos que reflejen el uso real de la lengua (4.2 y 4.3), por eso se utilizarán los corpus lingüísticos como herramienta principal de búsqueda con el fin de crear una puerta de acceso tanto a la información cuantitativa como cualitativa.

Estas dos preguntas de partida se han materializado en **tres grandes objetivos** que servirán de guía para poder llevar a cabo el proceso investigador y ofrecer así una aportación en el ámbito etnolingüístico. El **primer gran objetivo (A)** consistirá **en encontrar el mayor número posible de expresiones relacionadas con el ámbito cultural en cuestión**, ya que partimos de la **primera hipótesis** de que **es en las expresiones donde más se puede apreciar la influencia de la cultura en la lengua**. Para ello, se necesitará primero tener un listado de palabras con el que trabajar en los corpus, y por eso se realizará un *brainstorming* con hablantes nativos. El proceso, **que representará la fase 1 de creación de las pirámides etnolingüísticas**, como se verá en 4.2.1 y 4.3.1, consistirá en reunir a cinco participantes y entregarles tres hojas en blanco. Cada hoja será un ámbito cultural y se dispondrá de veinte minutos para hacer una lluvia de ideas y escribir todas las palabras (evitando expresiones pluriverbales) que vengan a la mente y que estén relacionadas, de forma directa o indirecta, con la religión, el fútbol y los toros. Una vez hecho esto, se someterá cada listado a unos criterios de criba (**fase 2 de creación de las pirámides etnolingüísticas**, en 4.2.2 y 4.3.2) para quedarnos solo con aquellas palabras que guarden un vínculo directo con el ámbito cultural seleccionado y que formarán parte de los listados definitivos. El tercer paso consistirá en introducir todas las palabras de cada listado en los corpus (**fase 3 del proceso de creación de las pirámides etnolingüísticas**, en 4.2.3 y 4.3.3) con el fin de encontrar el mayor número posible de expresiones. Este proceso, que incluye las primeras tres fases, nos pareció útil a partir de las lecturas sobre antropología lingüística y el posterior análisis de algunas teorías que se detallarán a lo largo de

todo el apartado 2.2<sup>4</sup>.

Para poder presentar mejor las expresiones encontradas en los corpus (fase 3), se ha fijado un **segundo objetivo (B)** que consistirá en **ordenar y visualizar toda esta información cuantitativa y cualitativa dentro de unas pirámides para que luego sea posible realizar un primer análisis contrastivo entre el español y el italiano**. Hemos sido aún más conscientes de la importancia de este segundo objetivo tras la lectura del ensayo *Information Visualization: Beyond the Horizon* de Chaomei Chen, a lo largo del cual se expone cómo visualizar la información mediante gráficos o mapas mentales para poder contribuir a una mejor recepción de los conceptos y a una mejor respuesta cognitiva por parte del receptor del mensaje (*Information Visualization: Beyond the Horizon*, cap. 3). Así pues, se establecerán dos criterios fundamentales para ordenar las expresiones dentro de las pirámides (**fase 4 del proceso**, en 4.2.4 y 4.3.4): el de **frecuencia** y el de **rentabilidad comunicativa**. El primero tiene que ver con el número de ocurrencias encontradas tras la búsqueda en el corpus (fase 3), mientras que el segundo se aplicará en base a las marcas que aparecen en los diccionarios para determinar si una expresión se considera coloquial o poco apropiada en según qué contexto. Para ello, también se utilizarán los corpus orales y la red como corpus. Confiamos en que dichos criterios permitirán ordenar las expresiones de arriba abajo y la pirámide podrá ofrecer una visión clara de en qué medida la cultura influye en la lengua<sup>5</sup>.

Llegados a este punto, se ha establecido un **tercer gran objetivo (C)** que aspira a **demostrar, mediante un análisis contrastivo a partir de la representación visual de las pirámides, que la cultura influye de manera diferente en cada lengua, incluso con ámbitos culturales afines, y que la cuña de las expresiones depende de múltiples factores, como por ejemplo la creatividad de los hablantes, la imitación del propio ser humano gracias a la función de las neuronas espejo (a la que se dedicará todo el apartado 2.3), y también la historicidad, ya que acontecimientos históricos concretos pueden marcar el nacimiento de una expresión que luego perdura en el tiempo**<sup>6</sup>.

Los tres objetivos propuestos pretenden **ofrecer un nuevo enfoque en el campo de la etnolingüística, ahondando en las relaciones entre dos lenguas afines como son el español y el italiano, además de ofrecer una nueva herramienta que facilite tanto el aprendizaje de la lengua como la inmersión cultural**. Para ello, se parte de **dos hipótesis** más: por un lado, de que **es posible utilizar los corpus para demostrar que la afinidad etnolingüística entre el español y el italiano es solo parcial y que la influencia de la cultura en la lengua se puede visualizar gracias a las pirámides**; por otro lado, que **con ámbitos culturales no**

4. Somos conscientes de que existen muchos diccionarios fraseológicos que recogen las gran mayoría de expresiones tanto en español como en italiano, pero en esta tesis se optó por seguir el método “Data-Driven”, como se detallará en 1.4, que consiste principalmente en formular una teoría a partir de la recopilación de datos.

5. Para acotar la investigación, se priorizó el aspecto sincrónico de la lengua, fijándose en las expresiones de los últimos quince años (2001-2015). Además, en el caso del español, se tuvieron en cuenta solo las expresiones de España y no del resto de Hispanoamérica.

6. Queremos destacar, a este respecto, la labor del Centro de Investigación de Sociolingüística y Comunicación (CUSC) de la Universitat de Barcelona, dirigido por el Dr. Xavier Vila i Moreno, que se propone, entre otras cosas, profundizar en la realidad sociolingüística y sociocomunicativa global en un entorno multicultural mediante datos obtenidos con encuestas en colegios a niños y adolescentes.

compartidos entre dos comunidades lingüísticas, como podría ser el caso de los toros en español y en italiano, la visualización de las expresiones presenta grandes diferencias, como se verá en 4.2.3.3 y en 4.3.3.3). Además, las pirámides etnolingüísticas **pretenden reflejar aspectos de la cosmovisión colectiva; es decir, de ese conjunto de creencias y valores que maneja una comunidad cultural a la hora de interpretar el mundo**<sup>7</sup>, como se verá en el apartado 2.2.

En conclusión, con las pirámides etnolingüísticas se pretende visualizar una información cultural en español y en italiano que puede ayudar a entender mejor el pensamiento y la cosmovisión de “los otros”, especialmente a la hora de estudiar una L2.

---

7. Gracias a mis estudios en la Universitat de Barcelona he podido conocer el proyecto PremioRom sobre refranes en la Romania, dirigido por el Dr. José Enrique Gargallo, cuyas aportaciones han sido muy esclarecedoras para la realización de esta tesis.

### 1.3. ORGANIZACIÓN DEL TRABAJO

El presente trabajo está estructurado en dos partes. En la *Parte I*, nos adentraremos en el mundo de la etnolingüística, ofreciendo una panorámica detallada de esta disciplina que constituye el marco teórico principal de esta investigación y facilitará el acercamiento al primer objetivo (A) que nos hemos propuesto.

Es evidente que ha habido una tendencia bastante generalizada, especialmente en el siglo XX y en mayor medida por parte de los estructuralistas, a estudiar la lengua como un sistema independiente, sin considerar la enorme influencia que la cultura ejerce a diario sobre ella. Tras explicar el significado del término *etnolingüística* y su supuesto sinónimo *antropología lingüística*, en el capítulo 2 se analizarán los conceptos de *cultura*, *ideología* y *cosmovisión*, así como las varias teorías existentes acerca del relativismo lingüístico, *in primis* la de Sapir-Whorf, la de Harry Hoijer y la de Dell Hymes. Asimismo, se dedicará todo el apartado 2.3 a las neuronas espejo, descubiertas en 1992 por el equipo científico de Giacomo Rizzolatti. Estas neuronas se encuentran en una parte del cerebro llamada *corteza premotora*, en el lóbulo frontal izquierdo, y nos permiten leer el mundo y empatizar con las personas con las que interactuamos, además de vincularnos entre nosotros desde el punto de vista mental y emocional. Es precisamente por esta virtud que las neuronas espejo están estrechamente vinculadas con la propuesta de las pirámides etnolingüísticas, y justifican el hecho de que triunfen determinadas expresiones o muletillas, de que en una comunidad se utilice una palabra en lugar de otra o de que las expresiones procedentes de un ámbito cultural concreto cuajen o no en la lengua. Conscientes de ello, la publicidad y el *marketing* han traído a la lengua un sinfín de palabras y expresiones, en su mayoría anglicismos, que antes apenas se usaban y que ahora están en boca de todos<sup>8</sup>.

Finalmente, se dedicará el apartado 2.4 al papel fundamental que desempeña la semántica en este estudio, haciendo particular hincapié en el concepto de intensidad/extensión (2.4.1), en la intensidad (2.4.2), en la sinonimia y polisemia (2.4.3) y en los campos léxicos (2.4.4), todo estrechamente vinculado con las neuronas espejo y con la propuesta que se quiere presentar en esta tesis.

El capítulo 3 está dedicado por entero a los corpus lingüísticos, herramienta clave del primer objetivo (A) y sin los cuales la presente investigación no se hubiese podido llevar a cabo de ninguna manera. Por ello, la aportación de la Dra. Mar Cruz, directora de esta tesis, ha sido fundamental para engarzar la teoría con la práctica. Se ofrecerá una visión en conjunto de la lingüística de corpus a lo largo de todo el apartado 3.1, sopesando la utilidad de estas bases de datos lingüísticos (3.1.1) que sugieren nuevos enfoques en el ámbito de las lenguas. Se verán las diferentes tipologías de corpus que existen (3.1.2) y se abordará el tema de la *frecuencia* y de la *rentabilidad comunicativa* de una determinada palabra o expresión (3.1.3), criterios que se intentarán aplicar a la hora de ordenar las expresiones dentro de las pirámides (fase 4 del proceso de creación) para alcanzar el segundo objetivo del presente trabajo (B).

8. A este respecto, son muy esclarecedores los trabajos surgidos del proyecto LiP (LLengua i Publicitat), de la Universitat de Barcelona, coordinado por la Dra. Lúdia Pons i Griera.

En la segunda parte de este capítulo 3 se describirán los principales corpus y se analizarán a fondo los que utilizaremos para la búsqueda en español (3.2) y en italiano (3.3) puesto que serán un punto de referencia constante a lo largo de la investigación y constituirán el principal recurso metodológico para llevar a cabo la búsqueda (fase 3 del proceso de creación). Asimismo, se dedicará el apartado 3.4 al análisis del corpus más grande del mundo: la red. Los contrastes que presenta son considerables, debido sobre todo a la escasa fiabilidad de procedencia de gran parte de la información, pero en algunos casos, tras una criba realizada con tino, puede convertirse en otro recurso útil.

Con la *Parte II* se alcanza el punto crucial de la tesis, que ocupará todo el capítulo 4, en el que se procederá a explicar las pirámides etnolingüísticas y su aportación tanto en el ámbito filológico como en el lingüístico. Nos centraremos en el proceso de creación de las pirámides etnolingüísticas tanto en español (4.2) como en italiano (4.3) sobre la religión, el fútbol y los toros, para que al final se puedan cotejar con el fin de realizar un análisis contrastivo (4.4) a partir de la **visualización de datos**, lo que permitirá detectar las similitudes y diversidades entre las dos lenguas en los ámbitos culturales seleccionados, de acuerdo con el tercer objetivo de esta tesis (C). El entero proceso de creación pasará por cuatro fases bien delimitadas que se explicarán de forma pormenorizada: el *brainstorming* (fase 1, subapartados 4.2.1 y 4.3.1), la **obtención del listado de trabajo y la criba** (fase 2, subapartados 4.2.2 y 4.3.2), la **búsqueda en los corpus** (fase 3, subapartados 4.2.3 y 4.3.3) y la **selección y ordenación visual de las expresiones dentro de las pirámides** (fase 4, subapartados 4.2.4 y 4.3.4).

Por último, en la *Parte III*, se dedicará el capítulo 5 a las consideraciones finales y se reflexionará sobre las aportaciones y limitaciones de la investigación (5.1). En particular, se retomarán los tres objetivos que hemos fijado al principio (A, B y C) para ver si se han alcanzado o no, viendo además hasta qué punto ha servido la metodología escogida. También se dedicará un subapartado a las conclusiones (5.2) y se dilucidará sobre las líneas que quedan abiertas (5.3), pues al escribir esta tesis se han ido abriendo muchos caminos que no hemos podido explorar, pero que abren vías de investigación para el futuro.

## 1.4. METODOLOGÍA

Una de las primeras ideas que tuve cuando este proyecto aún se encontraba en fase de gestación fue la de compilar un corpus etnolingüístico bilingüe y bidireccional español-italiano a través de grabaciones, entrevistas y una selección de textos escritos que abarcasen únicamente los últimos quince años (por ejemplo, 2001-2015), pero al poco tiempo me di cuenta de que recoger un número suficiente de datos para poder acometer una investigación exhaustiva y luego clasificarlos únicamente por su función etnolingüística era bastante inviable, así que, tras un período de reflexión, me pareció mejor opción aprovechar varios corpus ya existentes que no compilar uno desde cero, sobre todo para los fines propuestos.

Como nuestro interés se centra en las expresiones, ya que partimos de la hipótesis de que son vehículos de la cultura (1.2), al principio se barajó la posibilidad de apoyarse en los diccionarios fraseológicos, pero luego dicha idea se descartó y se prefirió utilizar el Data-Driven, un método que ofrece un enfoque mucho más inductivo con el que se pretende formular y corroborar una teoría a partir de la recopilación de datos. Según Hilary Mason y Dhanurjay “DJ” Patil, el análisis de los datos es fundamental en todo proceso investigador, y fomentar una cultura basada en los datos va a ser una de las claves del éxito en el siglo XXI (*Data Driven. Creating a data culture*, 1-5), de ahí que el Data-driven constituya uno de los puntos de partida de esta tesis<sup>9</sup>.

Desde un punto de vista estrictamente metodológico, queremos precisar que a lo largo del presente trabajo se manejará información cuantitativa con los corpus lingüísticos, pero dicha información será analizada en todo momento de manera cualitativa con unos criterios que se irán exponiendo a lo largo de esta tesis y que constituirán las columnas portantes de todo el trabajo. El manejo de esta información a menudo puede resultar complicado, por eso se optó por utilizar una metodología mixta basada en la triangulación de datos y siguiendo el modelo de Zoltan Dornyei, quien propone combinar métodos cuantitativos y cualitativos para luego aplicarlos con un plan de acción (*Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*, cap. 2). Evidentemente, los métodos mixtos implican una serie de suposiciones que guían hacia la recolección de datos y su posterior análisis con enfoques cualitativos y cuantitativos en combinación para dar respuesta a los objetivos propuestos. Para que este tipo de metodología resulte efectiva, es fundamental atenerse al concepto de “investigación-acción”, sobre todo para tener acceso a la información cuantitativa (González-Lloret, “Investigación-acción (II): La investigación cuantitativa”). En nuestro caso, **el primer paso de la triangulación de datos empieza con la realización de un *brainstorming*** (4.2.1 y 4.3.1) para obtener el mayor número posible de palabras relacionadas con los tres ámbitos culturales escogidos: la religión, el fútbol y los toros. Para ello, es menester encontrar unos informantes nativos para el español y otros para el italiano para someterlos a la prueba y conseguir una gran cantidad de palabras. Tras el *brainstorming*, se suele disponer de información cuantitativa para empezar a trabajar, pero normalmente muchas palabras apenas tienen relación con el

9. A este respecto, también ha sido muy esclarecedora la labor de Bestiario, una empresa de visualización de datos con sede en Barcelona que aplica el Data Driven y que cuenta con clientes como Telefónica, la Caixa y el Ayuntamiento de Barcelona, entre otros. [www.bestiario.org](http://www.bestiario.org).

ámbito cultural en cuestión y su aparición se debe a factores relacionados con la cosmovisión de los participantes y su capacidad de asociar ideas afines. Aquí es donde cobra importancia la metodología mixta y la combinación de enfoque cuantitativo y cualitativo. Así pues, lo primero que se llevará a cabo es una criba de los listados de palabras (fase 2 del proceso de creación de las pirámides, en 4.2.2 y 4.3.2) para descartar, en base a unos criterios fijados previamente, todas aquellas palabras que no guarden una relación directa con el ámbito cultural seleccionado. **El segundo paso de la triangulación consiste en apoyarse en los corpus lingüísticos para poder buscar todas las palabras** (fase 3 del proceso de creación, en 4.2.3 y en 4.3.3) **y pasar de las ocurrencias a las expresiones plurverbales**, lo que representaría un puente entre la información cuantitativa y la cualitativa.

Como se verá en el capítulo 3, existen muchos tipos de corpus disponibles en internet que nos pueden ser de gran ayuda para obtener información lingüística relevante, pero no se puede perder de vista el concepto de *representatividad*, fundamental para poder extraer conclusiones fiables a partir de los datos estadísticos (Data-driven). Por eso, aunque el CREA (*Corpus de Referencia del Español Actual*), un banco de datos que actualmente recoge textos en lengua española desde 1975 hasta 2004, parecía *a priori* el motor de búsqueda ideal, precisamente por aspirar a ser “de referencia”, al final se optó por el CORPES XXI, disponible actualmente en la red en versión beta y donde el reparto de pesos entre España y América pasa a ser del 30/70 (en lugar del 50/50 del CREA), con lo que resulta más representativo de la distribución de los hablantes. Según la teoría de Joan Torruella y Joaquim Llisterri, no siempre un corpus será representativo solo por ser mayor y contener muchas palabras (“Diseño de corpus textuales y orales”, 45-77).

Para la elección del corpus como motor de búsqueda se tuvo también en consideración el problema de la **lematización**; es decir, el hecho de que en el corpus se relacionen con el *lema* (por ejemplo, el verbo en infinitivo) todas sus *formas* (cada una de las posibilidades flexionadas de ese verbo). La falta de lematización dificulta bastante tanto la búsqueda en sí como las listas de frecuencia, ya que si por ejemplo buscamos el verbo *tener* solo nos aparecerá esa forma de infinitivo y no todas sus formas flexionadas (*tengo, tenía, tuve*, etc.). Nosotros compartimos la teoría de que tanto la recuperación como el manejo de la información cuantitativa son fundamentales en la mayoría de los procesos investigadores, como precisa Raquel Gómez Díaz (*La lematización en español: una aplicación para la recuperación de información*, 13) y este es otro criterio que nos ha llevado a optar por el CORPES XXI, que a diferencia del CREA sí está lematizado y nos proporciona todas las formas de un mismo lema.

Para complementar la información en el ámbito oral y a raíz de nuestro interés por el habla espontánea, se consultará principalmente el *Corpus de Conversaciones Coloquiales Val.Es.Co.*, elaborado por Antonio Briz y el Grupo Val.Es.Co. (Valencia, Español Coloquial), que recopila trescientas cuarenta horas de grabación realizadas en la comunidad lingüística de Valencia y diecinueve conversaciones coloquiales obtenidas mediante grabaciones secretas y anónimas realizadas en espacios familiares.

En cuanto al italiano, al ser una lengua con menos hablantes y menos difusión geográfica que el español, lógicamente nos hemos encontrado con una menor variedad de herramientas lingüísticas y los corpus que más se han tenido en consideración para esta tesis han sido

el CORIS/CODIS (*Corpus di Italiano Scritto*) de la Università di Bologna, el único corpus empleado que refleja integralmente la lengua escrita con cerca de ciento treinta millones de palabras, y el LIP (*Lessico dell'Italiano Parlato*), un proyecto dirigido por el lingüista Tullio de Mauro que se inició a mediados de los años noventa y que recopila conversaciones orales en los exámenes de estudiantes de italiano como lengua extranjera hasta 2006.

La triangulación de datos se cierra con la **visualización** de todas las expresiones pluriverbales que aparecen tras la búsqueda en los corpus y su clasificación dentro de las pirámides ateniéndose a dos criterios principales: el de **frecuencia** y el de **rentabilidad lingüística**, que se expondrán a lo largo de los subapartados 4.2.4 y 4.3.4. Estos criterios son fundamentales para ordenar tanto la información cuantitativa como cualitativa y ofrecer una visualización de datos clara (objetivo B) que facilitará el análisis contrastivo de las pirámides etnolingüísticas del español y del italiano con el fin de ver cómo influye la cultura en cada lengua (objetivo C). Finalmente, nos parece necesario recalcar la importancia de la red y de motores de búsqueda tan poderosos como Google, ya que a menudo permiten esclarecer algunas dudas lingüísticas, facilitando, en nuestro caso, la clasificación de las expresiones pluriverbales dentro de las pirámides. Así pues, nos apoyaremos también en la red como corpus y la utilizaremos como herramienta complementaria de la metodología mixta que se ha presentado, confiando en que permita alcanzar los tres objetivos principales de esta tesis doctoral.





# **PARTE I**

## **2. LAS HUELLAS DE LA CULTURA EN LA LENGUA**

## **3. TIPOS DE CORPUS Y SU PAPEL EN LAS PIRÁMIDES ETNOLINGÜÍSTICAS**



## **2. LAS HUELLAS DE LA CULTURA EN LA LENGUA**

## 2.1. SOBRE EL TÉRMINO “ETNOLINGÜÍSTICA”

La relación entre lengua y cultura se ha caracterizado siempre por un intercambio recíproco de información y por un enriquecimiento mutuo. El prefijo *etno-* (del griego ἔθνος *ethnos*, “pueblo” o “nación”) alude a una población humana en la cual los miembros se identifican entre ellos, normalmente con base en una real o presunta genealogía y ascendencia común, u otros lazos históricos. Las etnias están también normalmente unidas por unas prácticas culturales, de comportamiento, lingüísticas, o religiosas comunes, de ahí que se haya utilizado dicho prefijo para referirse a esta rama de la lingüística. Según Germán Fernández Guizzetti, la etnolingüística es como un puente que conecta dos mundos: el del idioma y el de la cultura (“La etnolingüística: del mundo del idioma al mundo de la cultura”, 75-93).

Hoy en día se suelen utilizar con relativa frecuencia los términos *etnolingüística*, *etnografía de la comunicación* y *antropología lingüística*. En consecuencia, antes de empezar este trabajo de investigación nos hemos planteado cuál de los tres es el más adecuado para referirse a la influencia que ejerce la cultura en una determinada lengua, sobre todo porque representa el marco teórico de la investigación. Según Giorgio Cardona, los tres términos son sinónimos y se diferencian únicamente por el hecho de que *antropología lingüística* (*anthropological linguistics*) se empezó a utilizar en Estados Unidos a mediados de los años cincuenta del siglo XX, *etnografía de la comunicación* (*ethnography of speaking*) fue introducida por Dell Hymes (1964) también en EE.UU y *etnolingüística* ganó terreno mayoritariamente en Europa (*Introduzione all'etnolingüística*, 9). El propio Cardona aboga por utilizar este último, principalmente por su posible función de adjetivo y por ser una sola palabra. Para esta tesis también creemos que *etnolingüística* es el término menos ambiguo de los tres y el que mejor refleja, sin remitirnos a la antropología y a la comunicación, la idea de relación entre lengua y cultura que constituye el eje vertebrador de este trabajo.

Manuel Casado Velarde, por otro lado, establece en primer lugar una diferencia, desde un punto de vista meramente lingüístico, entre sociolingüística y etnolingüística, definiendo la primera como una disciplina que “estudia la variedad y variación del lenguaje en relación con la estructura social de las comunidades hablantes”, mientras que la segunda se dedica al “estudio de la variedad y variación del lenguaje en relación con la civilización y la cultura”. Una vez hecha esta distinción, Casado Velarde perfila una dualidad dentro del mismo término *etnolingüística* dependiendo de si se trata de la interpretación cultural de una lengua o de la expresión idiomática de una cultura. Así pues, el autor establece la siguiente división (*Lengua y cultura. La etnolingüística*, 41):

**Lingüística etnográfica (o etnolingüística propiamente dicha):** estudia los hechos lingüísticos en cuanto determinados por nuestro conocimiento acerca de las cosas (cultura);

**Etnografía lingüística:** estudia la cultura (conocimiento acerca de las cosas) en cuanto manifestada por el lenguaje.

En otras palabras, los dos conceptos se ciñen a un análisis puramente unidireccional, el primero de la cultura al lenguaje y el segundo del lenguaje a la cultura. Evidentemente, se trata de un enfoque muy interesante y por supuesto muy innovador en cuanto a terminología se refiere, pero pensamos que la relación entre lengua y cultura se basa más bien en los conceptos de *bidireccionalidad* y *reciprocidad*; es decir, los dos se entremezclan constantemente y de manera recíproca sin que haya ninguna direccionalidad marcada, de ahí que nos hayamos decantado definitivamente por utilizar solamente el término *etnolingüística* sin ulteriores subcategorizaciones. Además, en toda lengua existe una cultura intrínseca que la alimenta, y a su vez toda cultura necesita del lenguaje para manifestarse, por eso en este trabajo se le va a otorgar un papel primordial al concepto de *reciprocidad*.

Finalmente, Eugenio Coseriu sostiene que hay dos grandes sentidos que proporcionan la justificación racional de la etnolingüística. En primer lugar, el lenguaje refleja la cultura no lingüística; en segundo lugar, no se habla solo con el lenguaje como tal (competencia lingüística), sino también con los saberes, las ideas y las creencias (competencia extralingüística) que influyen sobre la propia expresión y la determinan de alguna manera (“La socio- y la etnolingüística: sus fundamentos y sus tareas”, 5-29).

Esta teoría de Coseriu es la que mejor se ajusta a lo que se quiere proponer en esta tesis, por ello la competencia extralingüística se tendrá muy en cuenta durante el proceso de creación de las pirámides etnolingüísticas, como se verá en 4.2 y 4.3.

## 2.2. SOBRE LA CULTURA Y LA LENGUA

En nuestros primeros planteamientos, nos preguntamos qué factores determinan la evolución de determinadas expresiones y qué papel desempeña el entorno cultural en la creación y fosilización de las mismas. Teniendo en cuenta que la cultura es un conjunto acumulativo de conocimientos y creencias que nos permiten adaptarnos e integrarnos en la sociedad, y que el ser humano suele expresarlos mediante la lengua, se consideró fundamental tratar la relación entre lengua y cultura a partir de un enfoque bidireccional y recíproco.

La cultura es producto y manifestación de la interacción entre los hablantes, y constituye el fundamento de toda sociedad. Al interactuar, los grupos sociales generan cultura, y es evidente que toda sociedad simboliza en gran medida sus comportamientos a través del lenguaje, indispensable para fomentar el desarrollo de la cultura en el ámbito humano, social, económico, geográfico y político (Ariño, *Sociología de la cultura. La constitución simbólica de la sociedad*, cap. 4). Precisamente por ser un producto social, es innegable que la lengua evoluciona y cambia constantemente a raíz del uso que le dan sus hablantes mediante el habla, como mencionan Patrick Gleeson y Nancy Wakefield en su ensayo sobre la estrecha relación entre lengua y cultura (*Language and Culture*, 4). En palabras de Ferdinand de Saussure, “el habla es necesaria para que la lengua se establezca e históricamente el habla precede a la lengua y la hace evolucionar” (*Curso de lingüística general*, 13).

A principios del siglo XX, la cultura se fue asociando cada vez más a la antropología y en EEUU pasó a tener dos grandes significados: por un lado, se entendía como la evolucionada capacidad humana de clasificar y representar las experiencias con símbolos y actuar de forma imaginativa y creativa; por otro lado, se refería a las varias maneras en que la gente vive en diferentes partes del mundo, clasificando y representando sus experiencias y actuando creativamente (Levine, *Simmel: on individuality and social forms*, 6). Asimismo, resulta interesante la definición conceptual que acuñaron los autores Alfred Kroeber y Clyde Kluckhohn en 1952 con el fin de fundir en una sola las 164 definiciones anteriores de cultura que habían encontrado. Afirmaron que: “La cultura es un producto, es histórico, incluye ideas, patrones y valores, es selectiva, es aprendida, está basada en símbolos y es una abstracción de la conducta y de los productos de la conducta” (“Culture: A critical review of concepts and definitions”, 47).

Dicho esto, y de acuerdo con la intención de esta tesis (analizar las huellas de la cultura en la lengua), se contemplan dos tipologías de cultura. En primer lugar, se sugiere la denominación de **cultura adquirida**; es decir, ese conjunto de creencias, valores y tradiciones que representan la identidad de un pueblo y con los que nos moldeamos desde temprana edad. Se trata, pues, de un enorme bagaje cultural con el que entramos en contacto desde pequeños bien de manera directa bien indirecta y que nos permite caracterizarnos dentro de un perfil humano, otorgándonos unos rasgos más o menos prototípicos que nos acompañarán a lo largo de nuestra vida y que están bien asentados en nuestro subconsciente. Por otro lado, tenemos la que proponemos denominar como **cultura compartida**, originada por el contacto directo entre los pueblos y por el creciente cosmopolitismo en el que vivimos. Una sociedad cosmopolita es una sociedad viva, abierta y sobre todo en constante evolución, lo cual permite

constantes intercambios de ideas y puntos de vista para el enriquecimiento general de todos. Por lo tanto, podemos colegir que entre **cultura adquirida** y **cultura compartida** pasa algo similar a lo que ocurre entre una lengua materna y una L2. La primera la asimilamos de niños y aprendemos a utilizarla de manera innata, mientras que la segunda pasa por un proceso de aprendizaje, con todas las dificultades que eso conlleva.

Cabe señalar que tanto la **cultura adquirida** como la **cultura compartida** constituyen la base de nuestra conducta; es decir, la manera en la que actuamos e interactuamos a diario con las otras personas. Aquí desarrolla un papel fundamental la *sociedad*, concebida como *comunidad lingüística*, que se presenta como un sistema de interrelaciones que une a un cierto número de individuos por el hecho de compartir una misma lengua. Patricia Greenfield y Jerome Bruner sostienen que en esta interacción que ejercen los miembros de una sociedad se pone de manifiesto una identidad propia y de esta manera el ser humano pone en acción un conjunto de vivencias que lo distinguen de otras sociedades y que lo hacen crecer como persona. Los miembros de una sociedad, por lo tanto, tienden a identificarse con su lengua como si se tratara de una herencia de la cual tienen que apropiarse (“Culture and Cognitive Growth”, 89-107). Saussure, al hablar de la inmutabilidad del signo, menciona que: “De hecho, ninguna sociedad conoce ni jamás ha conocido la lengua de otro modo que como un producto heredado de las generaciones precedentes y que hay que tomar tal cual es” (*Curso de lingüística general*, 53).

Desde siempre, los seres humanos hemos tenido la necesidad de interactuar con nuestros semejantes por varias razones, *in primis* nuestra lucha por la supervivencia, y durante este proceso fuimos desarrollando múltiples conocimientos, hábitos y acciones que en su conjunto dieron origen a las diversas culturas, que a su vez se han vehiculado en la sociedad a lo largo de los siglos gracias a la lengua<sup>10</sup>. Por lo tanto, esta relación hombre-lengua-cultura-sociedad es fundamental para deducir los patrones etnolingüísticos de un pueblo, además de conocer su pasado cultural y poder comprender mejor los principios y causas que originan la evolución de una sociedad específica. Este aspecto permite enlazar con la **ideología**, ese conjunto sistemático y coherente de creencias, compartidas por un grupo social, que explica y controla la realidad social (*Colletti, Ideologia e società*, cap. 1). Hoy en día se habla de ideología cuando una idea determinada es compartida conscientemente por un grupo social en una sociedad. A veces es un rasgo muy identitario, de forma similar a la religión, la nación, la clase social, el sexo, los partidos políticos, las sectas o los grupos cerrados y/o abiertos, como ser partidario de un equipo de fútbol, de modo que se tendrá en gran consideración para elaborar las pirámides etnolingüísticas. Además, los miembros del grupo ideológico admiten o no que un determinado individuo pertenezca al grupo según comparta o no ciertos presupuestos ideológicos básicos. Sin embargo, es cierto que desde la época de Marx y Engels<sup>11</sup> el concepto

10. A lo largo de su obra *Etimologías*, el eclesiástico Isidoro de Sevilla (556-636) defiende la idea de que los pueblos tienen el origen en las lenguas.

11. Marx utilizó el término en *El capital* para referirse a la clase opresora (1867), definiéndolo como un proceso realizado conscientemente por el así llamado pensador, pero con una conciencia falsa cuyo fin principal es controlar y explotar a la masa obrera.

se ha rodeado de un aura peyorativa que lleva arrastrando hasta nuestros días, así que con frecuencia las ideologías se siguen considerando meros instrumentos de control social. Bajo esa concepción negativa, la ideología pretende despojar al hombre de su libertad, sumergiéndolo en una mentira y convirtiéndolo en parte de una masa que se pretende manipular y dominar. Respecto de las características principales, Terry Eagleton sostiene que la ideología es explícita, estructurada y sistemática, y se compone de creencias compartidas por los miembros de un grupo social para defender sus intereses (*Ideology. An Introduction*, 1-2). Así pues, si lengua y cultura interactúan de manera recíproca, es interesante analizar las posibles influencias que pueda tener una determinada ideología en el ámbito lingüístico y ver sus repercusiones en los hablantes.

Para entender mejor la ideología y demostrar la influencia de la cultura en la lengua una vez se hayan visualizado los datos en las pirámides etnolingüísticas (objetivo C), será determinante apoyarse en la **cosmovisión**, ese conjunto de representaciones mentales compartidas por un grupo social que pretende explicar la totalidad del universo desde un punto de vista social y natural. La cosmovisión es más amplia que la ideología por lo que a contenidos se refiere, pues sus proposiciones no solo abarcan la realidad social del hombre, sino también el universo físico. A diferencia de la cultura, que engloba todo aquel contenido cognitivo y valorativo, independientemente de que sea más o menos compartido, la ideología y la cosmovisión son compartidas por grupos sociales determinados en defensa de sus intereses y son mucho más arbitrarias.

A su vez, se puede apreciar cómo a menudo comunidades culturales y grupos sociales con ideología o cosmovisión propia coexisten dentro de una misma sociedad. Esto tiene una influencia directamente proporcional sobre el lenguaje, enriqueciéndolo con expresiones que esconden una ideología o una cosmovisión concreta del mundo (*Terracini, Conflitti di lingue e di cultura*, 252). Muchas veces una misma expresión esconde más información de la que expresa llana y superficialmente su significado, y aprender a leer entre líneas y a detectar la posible ideología o cosmovisión del “otro” nos puede ayudar a mejorar nuestra comunicación y a hacer que nos entendamos mejor en un contexto empático de respeto mutuo.

Tanto la cultura adquirida como la cosmovisión están muy asentadas en nuestro subconsciente, por eso creemos que con el *brainstorming* se puede tener acceso a un sustrato cultural mucho más profundo del cerebro, y las palabras aportadas por los informantes constituyen el primer paso de la triangulación de datos para alcanzar el primer gran objetivo de esta tesis (A), que consiste en encontrar el mayor número posible de expresiones pluriverbales relacionadas con la religión, el fútbol y los toros, tres ámbitos culturales que nos pueden dar mucha información sobre la cosmovisión general de los españoles y de los italianos.

### 2.2.1. RASGOS UNIVERSALES DEL LENGUAJE HUMANO

El ser humano siempre ha tenido la necesidad primaria de comunicarse con sus semejantes y ha encontrado en el lenguaje su mejor y más elaborado instrumento de comunicación, que destaca, según Manuel Casado Velarde (*Lengua y cultura. La etnolingüística*, 32-33), por ser **racional** (es un producto de nuestro cerebro y de nuestra facultad de pensar) y **universal**

(todos tenemos la capacidad de articular nuestro pensamiento elaborando signos). Además, el lenguaje es **innato** (todos nacemos con la capacidad de comunicarnos), **aprendido** (vamos conociendo los signos de la comunidad lingüística en la que nos criamos) y **convencional** (los signos adaptados por una determinada comunidad son aceptados por todos sus miembros). La cultura adquirida que se mencionó en 2.2 desempeña un papel determinante, y ya desde niños marca la manera de comunicarnos con los demás. Casado Velarde habla de cinco rasgos relacionados con el carácter universal del lenguaje: *la semánticidad, la creatividad, la alteridad, la materialidad y la historicidad* (29-31). En primer lugar, el lenguaje siempre significa algo, por eso hablamos de **semánticidad**, y todo hecho material que no lleve asociado un contenido significativo no se considera lenguaje. Además, el lenguaje es una producción genuina e innata del ser humano, y por consiguiente está sujeto a la creatividad del individuo, quien lo va moldeando en función de sus necesidades y bajo la influencia de la cultura, la ideología y la cosmovisión. Pero no vivimos solos, y para transmitir nuestras ideas y nuestros pensamientos necesitamos al “otro”, por eso la alteridad es quizá uno de los rasgos más indispensables del lenguaje, y se ve plasmado en el diálogo. Con el concepto de **materialidad** se hace referencia a que el lenguaje se presenta en el mundo como un hecho que podemos percibir con algún sentido, por ejemplo el oído o la vista; además, se realiza en formas históricamente determinadas que llamamos “lenguas”, de ahí el último rasgo: la **historicidad**. En este trabajo de investigación se prestará particular atención a los rasgos de **semánticidad** y **creatividad**, indispensables para analizar las pirámides etnolingüísticas y para llevar a cabo el análisis contrastivo entre el español y el italiano, lo que correspondería al segundo (B) y tercer (C) objetivo de este proyecto de investigación.

Para ejemplificar estos conceptos, vemos que en español la expresión *ni dios*, que se utiliza con significado de “nadie”, tiene dos variantes menos usadas que son *ni cristo* y *ni un alma*, como se verá en 4.2.3.1 y en 4.4.1, mientras que en italiano solo se usa una expresión y es *non esserci anima viva* (4.3.3.1), equivalente de *no hay ni un alma*. Si buscamos posibles equivalentes con *Cristo* y con *Dio* en italiano, realizando incluso búsquedas libres en la red como corpus, no encontraremos ningún resultado, lo que señalaría una sustancial diferencia entre estas dos lenguas tan aparentemente afines: la semánticidad es muy similar en las dos, pero la creatividad es diferente<sup>12</sup>. Así como en español se alternan diferentes palabras procedentes del ámbito religioso, en italiano se opta por utilizar únicamente el término *anima*, evitando mencionar tanto a *Cristo* como a *Dio*, ya que para la cosmovisión de un italiano se suele considerar blasfemia pronunciar esos nombres en vano.

Con esto se han querido introducir los rasgos principales del lenguaje humano a raíz de la clasificación ofrecida por Casado Velarde. En esta tesis se tendrán muy en cuenta la semánticidad y la creatividad como rasgos dominantes, ya que desempeñarán una función primordial en nuestro criterio de análisis contrastivo a partir de la visualización ofrecida por las pirámides; sin embargo, cabe subrayar que se tratará de corroborar las hipótesis planteadas en 1.2 de forma inductiva y a partir de la recopilación de datos (Data-driven), como se verá a lo largo de todo el capítulo 4.

12. También hay una diferencia de extensión/intensión, un concepto que se expondrá en 2.4.1.

### 2.2.2. EL RELATIVISMO LINGÜÍSTICO Y LA TEORÍA DE SAPIR-WHORF

Como ya se ha expuesto, creemos que una lengua refleja la cosmovisión de un pueblo y que interactúa constantemente y de manera recíproca con la cultura. Así pues, y de acuerdo con los objetivos que se propone en esta tesis (A, B y C, como se vio en 1.2), son especialmente relevantes las ideas de Wilhelm von Humboldt (1797-1835), considerado el padre de la etnolingüística, ya que fue el primer lingüista en hablar del espíritu de los pueblos y en plantearse si el uso de una lengua podía crear o no cultura. En concreto, llegó a afirmar que cada lengua, moldeada a partir del espíritu de una nación, constituye una imagen peculiar del mundo y contiene la visión de todos sus hablantes, defendiendo la relación intrínseca entre lengua y cosmovisión (*On language. The Diversity of Human Language-structure and its Influence on the Mental Development of Making*, cap. 4). Además, se dio cuenta de que los elementos de las lenguas se pueden intercambiar para producir oraciones infinitas, algo que tiene que ver con la productividad lingüística, pero sobre todo con el rasgo de la creatividad humana que se mencionó en el anterior apartado y que se tendrá en gran consideración a lo largo de este trabajo.

Humboldt anticipó el denominado **relativismo lingüístico**, cuyo punto de partida se sitúa en Franz Boas (1858-1942), antropólogo alemán que desarrolló esta teoría en EE.UU a principios del siglo XX. Según expuso Boas por primera vez en 1928, cada lengua es una particular clasificación de la experiencia, que se refleja tanto en la gramática como en el léxico (*Anthropology and modern life*, cap.7-8). En defensa de esta teoría, Boas puso el ejemplo de la abundancia de distinciones léxicas del campo semántico nieve en el léxico de los esquimales como reflejo de las necesidades lingüísticas de ese pueblo. Esta idea también apareció en el artículo de Laura Martin (“Eskimo words for snow: a case study in the genesis and decay of an anthropology example”, 418-423), y creemos que explica muchas de las diferencias idiomáticas entre el italiano y el español que se retomarán en ocasión del análisis contrastivo de las pirámides etnolingüísticas en el apartado 4.4.

Edward Sapir (1884-1939) y Benjamin Lee Whorf (1897-1941), discípulos de Boas, fueron más allá y elaboraron, a partir de datos empíricos y del estudio de las lenguas amerindias de Estados Unidos, la que se conoce como *hipótesis Sapir-Whorf*, ya que fue formulada de forma conjunta por ambos autores y se rige en el principio de la **relatividad lingüística**. En palabras de Whorf:

“The «linguistic relativity principle» [...] means [...] that users of marked different grammars are pointed by their toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world” (*Language, thought, and reality: Selected Writings of Benjamin Lee Whorf*, 221)

En este enunciado se afirma que los hablantes de cada lengua llegan a cosmovisiones particulares, y eso implica que una lengua refleje valores y creencias culturales. Edward Sapir sostenía la siguiente idea:

“Language is a guide to social reality [...] Human beings do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression for their society [...] the real world is to a large language unconsciously built up on the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The words in which different societies live are distinct worlds, not merely the same world with different labels attached”. (*The selected writings of Edward Sapir*, 169).

Según Sapir, pues, una lengua es reflejo de la cultura de un pueblo o de una comunidad, además de ser expresión sistemática de los valores y creencias compartidas por toda una sociedad. Con relación al problema de la traducibilidad de las lenguas, Sapir reconoció la imposibilidad de expresar todos los contenidos en todos los idiomas, pues no todas las lenguas tienen palabras para los mismos conceptos, bien por falta de necesidad bien porque los hablantes no han mostrado interés por lexicalizarlos, como vimos anteriormente con el ejemplo de los esquimales y el color blanco. **Esta idea será uno de los pilares del tercer objetivo de esta tesis (C), con el que se aspira a demostrar, mediante un análisis contrastivo a partir de la representación visual de las pirámides, que la cultura tiene una influencia diferente en español y en italiano, incluso con ámbitos culturales compartidos.** Respecto al mundo de los toros, por ejemplo, se puede afirmar que en España existe una tradición taurina enraizada en la sociedad desde hace siglos, por eso hay varias expresiones relacionadas con este ámbito cultural, como por ejemplo *dar la puntilla*, *hacer un brindis al sol* o *estar para el arrastre*, esta última utilizada con frecuencia en la lengua oral para expresar un estado de mucho cansancio. Metafóricamente, se remite a la imagen del toro muerto cuando es arrastrado fuera de la arena después de la corrida; sin embargo, en italiano no existe ningún equivalente con esa carga cultural, lo cual demuestra que el ámbito de los toros no es importante para los hablantes italianos, por eso no han tenido nunca la necesidad de lexicalizar estas expresiones ni de introducirlas en su lengua.

La lengua es, como sostenía el antropólogo Harry Hoijer (1904-1976), parte de la cultura, pero por encima de todo es la mejor clave de acceso a la mentalidad de un pueblo (*The relation of language to culture*, 554-573), y esta tesis gira alrededor de este concepto. Por eso mismo, uno de nuestros primeros planteamientos vertió sobre cuál sería la mejor forma de usar esa llave llamada lengua para acceder a la mentalidad de un pueblo, y esto nos llevó, en primer lugar, a escoger tres ámbitos culturales distintos para luego reflexionar sobre cómo se podrían extrapolar las expresiones relacionadas con cada uno de ellos.

En definitiva, la relatividad lingüística puede suponer un paso fundamental para entender mejor una lengua y ayudarnos a empatizar mejor con la cosmovisión de sus hablantes, y confiamos que este proceso se facilite mediante la representación visual de las pirámides etnolingüísticas.

### 2.2.3. EL RELATIVISMO PRAGMÁTICO DE DELL HYMES

Dell Hymes (1927-2009) es considerado como uno de los lingüistas norteamericanos más importantes del siglo XX. Según él, lo que realmente refleja las diferencias culturales entre distintas lenguas son los “modelos de interacción comunicativa”, y en 1966 llegó a afirmar lo siguiente:

“People who enact different cultures do to some extent experience distinct communicative systems, not merely the same natural communicative condition with different customs affixed. Cultural values and beliefs are in part constitutive of linguistic reality” (“Two Types of Linguistic Relativity”, 128).

Influenciado por las ideas de dos de sus predecesores, Franz Boas y Edward Sapir, llegó a la conclusión de que las diferencias culturales entre las lenguas no se dan tanto en su estructura como en su **uso**. Respecto a ello, en 1972 propuso una nueva visión etnolingüística, sosteniendo que es muy difícil comparar dos estructuras gramaticales y léxicas como vehículos portadores de una cosmovisión, lo cual supone, en parte, negar el relativismo lingüístico de Sapir-Whorf (“On communicative competence”, 269-293). Se trata, por lo tanto, de centrar el estudio etnolingüístico principalmente en el análisis del discurso y en los aspectos pragmáticos de una lengua, ciñéndose casi exclusivamente a las prácticas discursivas y dejando en un segundo plano la teoría Sapir-Whorf de la relatividad lingüística. Hymes considera que, así como la estructura fonológica, morfológica, sintáctica y semántica son intrínsecamente lingüísticas, el código pragmático de un grupo de hablantes es de naturaleza cultural, y así propuso la noción de “competencia de comunicación o competencia lingüística”, algo que los hablantes deben haber asimilado para poder emplear una lengua con éxito y propiedad en un determinado contexto comunicativo, adaptándose en todo momento a las situaciones y a las normas socioculturales que regulan el uso de la lengua dentro de una determinada comunidad. Para respaldar esta teoría, Hymes creó el modelo SPEAKING (*Models of the interaction of language and social life*, 67-68), que incluye los siguientes componentes:

- a) **La situación** (*Situation*): lugar y momento donde se produce la comunicación;
- b) **Los participantes** (*Participants*): todas las personas presentes;
- c) **Los fines perseguidos** (*Ends*): objetivo del encuentro;
- d) **Los actos realizados** (*Act sequences*): los mensajes producidos y su temática;
- e) **El tono adoptado** (*Key*): las características prosódicas de los mensajes;
- f) **Los medios empleados** (*Instrumentalities*): los canales de comunicación (oral/escrito) y los códigos y registros (lenguas, dialectos y niveles de lenguas);
- g) **Las normas** (*Norms*): las normas que regulan la interacción y los turnos de palabra a partir de inferencias socioculturales;
- h) **El género** (*Genre*): el género donde se desarrolla el evento de habla (una carta comercial, un discurso formal, un poema, un diálogo entre amigos, etc.).

En realidad, el modelo de Hymes y la teoría Sapir-Whorf de la relatividad lingüística no son incompatibles, sino que se complementan para ofrecer una visión más global y completa de la etnolingüística.

Cuando empezamos a formular la idea de las pirámides etnolingüísticas, nos encontramos con que el uso de la **frecuencia** como único baremo de categorización resultaba algo deficiente, y el modelo SPEAKING nos sugirió la idea de la **rentabilidad comunicativa**; es decir, hasta qué punto una palabra o expresión, pese a ser utilizada con frecuencia y a aparecer en muchos corpus, es rentable para un hablante en según qué contexto social. Como se ha planteado en el subapartado 2.2.1, es en el ámbito religioso donde más saltan a la vista las diferencias etnolingüísticas entre el español y el italiano, con expresiones que presentan una equivalencia parcial o nula entre las dos lenguas. Esto explicaría, como afirmó el etnólogo Bronislaw Malinowski (1884-1942), quien se ocupó de la lengua como parte del comportamiento humano, que el lenguaje es, ante todo, una *conducta*. Para el antropólogo polaco-británico, el lenguaje es un instrumento para la acción humana y una actividad social más que un medio de expresión del pensamiento. Por consiguiente, sostiene que las palabras están íntimamente ligadas a la cultura de cada comunidad, negando así su universalidad. Esto lo llevó a postular la idea de intraducibilidad de las lenguas, dado que estas, según él, son reflejo de una cultura particular y, por tanto, de una manera peculiar de ver el mundo y la realidad (*The Language of Magic and Gardening*, vol. 2, 4-65).

Según Noam Chomsky, finalmente, la competencia lingüística consiste esencialmente en la capacidad para manejar un sistema formal de reglas que produzcan un número potencialmente infinito de expresiones bien formadas de una lengua. Tal conocimiento, básicamente innato, es autónomo respecto al conocimiento del uso de la lengua (*Knowledge of language*, 3-8). La teoría de Hymes, como hemos visto, subraya la interdependencia entre la competencia lingüística y la competencia comunicativa: la lengua no se aprende (o se activa) como un sistema de reglas de combinación formal, sino como un conjunto de reglas que especifican lo que es adecuado o apropiado en una situación comunicativa. La competencia comunicativa, por tanto, es fundamentalmente un conocimiento de cómo adecuar la conducta comunicativa al contexto, por eso consideramos que el criterio de la **rentabilidad** de una expresión puede convertirse en una herramienta clave a la hora de organizar la estructura de las pirámides etnolingüísticas, como se verá en 4.2.4 para el español y 4.3.4 para el italiano.

#### 2.2.4. EL IDEALISMO LINGÜÍSTICO DE KARL VOSSLER

Para terminar esta primera parte del capítulo 2, nos parece pertinente mencionar la labor y las teorías del lingüista alemán Karl Vossler (1872-1949), considerado como el padre del idealismo lingüístico. Fue influido en gran parte por el italiano Benedetto Croce, quien defendía los aspectos creativos de lenguaje<sup>13</sup>, y también por Von Humboldt, quien consideraba el lenguaje como una actividad espiritual del ser humano. A raíz de esto, Vossler sostenía que toda lengua cambia en función de las necesidades de una comunidad y le atribuyó al hablante el papel de

13. Concebido como *enérgeia*, término griego utilizado por Aristóteles para referirse a algo que está actuando (Jaeger, Aristóteles. Bases para la historia de su desarrollo intelectual, 12).

**artista creador.** Para este autor, muchos fenómenos lingüísticos derivan de causas culturales, y pretendió demostrarlo en sus obras *Algunos caracteres de la cultura española* (1941) y *Cultura y lengua de Francia* (1955). Creía que el espíritu de un pueblo, que a menudo se ve plasmado en las obras literarias, es la causa de los cambios lingüísticos y refleja la mentalidad de un pueblo, como podemos colegir de este pasaje:

“Si consideramos, por ejemplo, el uso del subjuntivo en italiano, francés y español, se demuestra que en todas partes sirve para designar deseo, inseguridad y posibilidad, pero que, no obstante, en italiano revela una especial preferencia y capacidad para la expresión del deseo y para la manifestación del sentimiento; en francés moderno, para reflejar la inseguridad y la negación, y en español, la esfera de lo potencial y de lo futuro. Si se piensa también en los muchos medios de reflexión que el español ha encontrado y formado, a su vez, para indicar toda clase de posibilidades, condicionadas o incondicionadas, en lo que muestra una gran riqueza, si se le compara con el francés y el italiano, puede relacionarse esa predilección con los rasgos utópicos de su literatura, claro que no en relación causal, sino como fenómenos paralelos” (*Algunos caracteres de la cultura española*, 67).

En su libro *Cultura y lengua de Francia*, el romanista alemán va más allá y establece sorprendentes asociaciones entre rasgos lingüísticos y culturales, como la que relaciona el artículo partitivo en francés con el espíritu práctico y mercantil de los franceses (*Cultura y lengua de Francia*, 87), lo que le valió numerosas críticas. Sin embargo, el idealismo lingüístico de Vossler ha influenciado a muchos lingüistas, entre ellos al español Amado Alonso, quien consideraba la lengua como una expresión colectiva de las experiencias acumuladas a lo largo de las generaciones, y también a Amado de Miguel, quien ha plasmado estas ideas especialmente en relación con el ámbito religioso, como se puede apreciar en su obra *Los españoles y la religión*.

En esta tesis de investigación, la aportación de Vossler ha sido determinante y el idealismo lingüístico nos ha llevado a tomar en consideración el rasgo semántico de la **creatividad**, introducido en el subapartado 2.2.1, como aspecto característico del lenguaje humano. Como se vio con las expresiones no hay *ni dios/no hay ni cristo/non c'è un anima*, hay muchos casos en los que la semántica es similar, pero el proceso creativo ha originado algunos cambios. Apoyamos por tanto la idea de Vossler según la cual el hablante sería un **artista creador**, y pensamos que eso aporta una gran variedad lingüística que se ha de tener en cuenta, además de ofrecer una valiosa información sobre determinados patrones culturales que se verán durante el análisis contrastivo de 4.4.

Finalmente, creemos que la **creatividad** está estrechamente vinculada con la imitación y con la capacidad del ser humano de observar y de interactuar con sus similares, por eso se dedicará el próximo capítulo a las neuronas espejo, una de las piedras angulares de esta tesis de investigación.

### 2.3. LAS NEURONAS ESPEJO Y SU FUNCIÓN EN EL CEREBRO

El ser humano trata de reproducir a diario, bien por interés personal bien por la influencia que un determinado sujeto ejerce sobre él, acciones y expresiones que son procesadas inmediatamente por el cerebro. Imitación y empatía son, por lo tanto, las bases de nuestra experiencia en el mundo, nuestro crecimiento personal y nuestra madurez. Desde un punto de vista lingüístico, que es lo que se analizará en esta tesis para alcanzar el tercer objetivo (C), los hablantes tienden a reproducir expresiones que han cuajado en la lengua y que muchas veces presentan una fuerte carga cultural e identitaria, y esto se debe principalmente a las neuronas espejo, que fueron descubiertas en 1991, en Parma, por el equipo de Giacomo Rizzolatti. Estos científicos, tras numerosas pruebas, decidieron colocar electrodos en la corteza frontal inferior de un macaco con el fin de estudiar las neuronas especializadas en el control de los movimientos de la mano al realizar una determinada acción. Durante cada experimento, registraban la actividad en el cerebro del simio mientras le facilitaban tomar trozos de alimento, de manera que los investigadores pudieran medir la respuesta de la neurona a tales movimientos. Lo curioso fue que algunas de las neuronas del mono reaccionaron incluso cuando el animal no se había movido, y eso demostró la existencia de unas neuronas que reconocían una acción y la reproducían en la mente. Estas células están ubicadas en una zona del cerebro conocida como F5, situada en la corteza premotora del cerebro, en la denominada Área de Broca, responsable de la producción motora del lenguaje y también de la comprensión sintáctica de la oración. El hecho de que las neuronas espejo se encuentren justamente en el Área de Broca explica, por ejemplo, la tendencia de los seres humanos a repetir estructuras gramaticales fijas o a reproducir muletillas que escuchamos en nuestro entorno. En palabras del científico Vilayanur Ramachandran, “el descubrimiento de las neuronas espejo ha sido de tal envergadura que podría hacer por la neurociencia lo mismo que hizo el descubrimiento del ADN hizo por la biología”<sup>14</sup> (“Mirror neurons and imitation learning as the driving force behind the Great Leap Forward in human evolution”, 69). Tal vez se trate de una afirmación algo atrevida dada la enorme trascendencia del ADN, pero es la primera vez que nos encontramos frente a unas neuronas que son motoras y perceptivas al mismo tiempo; esto, además de demostrar que la percepción y la acción constituyen un proceso unificado en el cerebro, abre un panorama completamente nuevo de cara a futuras investigaciones sobre la mente humana. El mismo Ramachandran las definió como *las neuronas de la empatía*, pues facilitan la comprensión de las emociones de los otros. De alguna manera, si se activan cuando observamos una acción llevada a cabo por otro individuo, que nosotros también podríamos realizar, significa que nos encontraríamos ante una suerte de **lectura de la mente**. Respecto a ello, Vittorio Gallese y Alvin Goldman, dos neurocientíficos que formaban parte del equipo de Giacomo Rizzolatti, postularon por primera vez la teoría según la cual las neuronas espejo pueden ser el correlato neuronal de los procesos de simulación necesarios para entender otras mentes (“Mirror neurons and the simulation theory of mind-reading”, 493-501), lo que demostraría su papel fundamental en la **empatía**, justificando por tanto la afirmación

14. La traducción es propia.

de Ramachandran. Su presencia en el cerebro humano ha sido demostrada en experimentos posteriores realizados por Giacomo Rizzolatti y Marco Iacoboni, quienes estuvieron tomando registros en neuronas de pacientes humanos. Parece ser que las neuronas espejo constituyen el sustrato cerebral de la tendencia automática a imitar, permitiendo relacionar tanto las acciones propias como las ajenas y dotarlas de un significado en nuestra mente. Como exponen Rizzolatti y Sinigaglia, el cerebro que actúa es un cerebro que comprende. Somos criaturas sociales. Nuestra supervivencia depende de entender las acciones, intenciones y emociones de los demás. Las neuronas espejo nos permiten entender la mente de los demás, no solo a través de un razonamiento conceptual sino mediante la simulación directa. Sintiendo, no pensando (*Las neuronas espejo: los mecanismos de la empatía emocional*, 182).

Vittorio Gallese, compañero de Rizzolatti, fue un poco más allá y habló de un **mecanismo de simulación incorporado**, que serviría en el ser humano para comprender las intenciones de los otros con el fin de capacitarnos para la vida social (“Mirror neurons and the simulation theory of mind-reading”, 493-451). De alguna manera, gracias a este mecanismo las neuronas espejo nos ayudan a disolver la barrera que existe entre los otros y nuestro yo, y nos capacitan así para adoptar el punto de vista de otro, facilitando nuestro aprendizaje a través de la imitación intencional.

Avanzando con las investigaciones, incluso se ha llegado a sugerir que las neuronas espejo serían el mecanismo neuronal básico para el desarrollo del lenguaje, y esta teoría es sin duda la que más nos ha inspirado para la redacción de esta tesis. En concreto, el neurocientífico computacional Michael Arbib demostró que la pantomima desempeña un papel fundamental en el avance evolutivo que va desde el sistema de neuronas espejo simple del mono hasta el sistema neuronal mucho más sofisticado del lenguaje humano (“From monkeylike action recognition to human language: An evolutionary framework for neurolinguistics”, 105-167). Somos conscientes porque tenemos lenguaje y esto nos permite pensar, relacionarnos con los otros, originar cultura y entender toda la abstracción relacionada con el propio lenguaje, como podría ser, por ejemplo, el significado de una metáfora. Imitamos lo que vemos y escuchamos, y este aprendizaje por imitación constituye una de las bases de la cultura.

En definitiva, las neuronas espejo son las principales responsables de la imitación y de la creatividad (dos rasgos muy característicos de los seres humanos) y justifican gran parte de este trabajo, especialmente de cara al análisis contrastivo que se realizará en 4.4 a partir de la visualización de las pirámides etnolingüísticas.

### 2.3.1. DIFERENTES TIPOS DE NEURONAS ESPEJO

En el anterior apartado se vio que las neuronas espejo desempeñan un papel fundamental en las relaciones entre los seres humanos, pero a partir de ahora el foco de atención se centrará en el aspecto que más nos incumbe: el lenguaje. No hay duda, por ejemplo, de que no solo imitamos lo que vemos, sino que muchas veces reproducimos frases, modismos y palabras aisladas que escuchamos a diario y que antes no formaban parte de nuestro vocabulario. A modo de ejemplo, traigo a colación una breve conversación que escuché hace unos años en un bar y que me parece bastante representativa. Dos mujeres estaban hablando de jardinería y una

de ella empezó a mencionar los *tiestos* que iba a utilizar para colocar determinadas plantas. La otra, muy interesada, le dio algunos consejos y le dijo que conocía una tienda donde vendían unas *macetas* muy bonitas y a buen precio. La conversación siguió por el mismo camino, y lo interesante desde el punto de vista lingüístico es que la primera mujer dejó de utilizar la palabra *tiesto* e, influenciada por su interlocutora, empezó a decir *maceta*. Lo que realmente interesa ahora, más que entrar en detalles de sinonimia, es la influencia que ejerce un hablante sobre otro, hasta tal punto que el segundo decide adoptar palabras del primero y pasarlas de su vocabulario pasivo (palabras que conocemos, pero que no usamos, reflejo de la disponibilidad léxica) a su vocabulario activo (la base de nuestro idiolecto).

Este ejemplo no es un caso aislado y se da con mucha frecuencia, sobre todo en la lengua hablada, y a menudo se plasma también en la escrita. Para nosotros, este proceso tiene que ver con las neuronas espejo y con la capacidad de imitación del ser humano. Un hablante que delata una fuerte personalidad suele transmitir confianza y en algunos casos cierta admiración, y eso favorece sin duda el proceso de imitación. El problema, muchas veces, es que escuchamos expresiones, bien en nuestra lengua bien en una lengua extranjera que hemos aprendido, que nos resultan demasiado agresivas y nuestro cerebro se niega a reproducirlas. Según la teoría motora de la percepción del habla de Alvin Liberman (“The motor theory of speech perception revised”, 1-36), experimentada a posteriori con neuronas espejo por Luciano Fadiga, compañero de Rizzolatti, cuando escuchamos algo, las áreas cerebrales motoras correspondientes al habla se activan como si estuviéramos hablando. Pero resulta que, pese a reproducirlo en nuestra mente, a veces nos negamos a pronunciarlo. ¿Qué es lo que actúa como inhibidor lingüístico? Según Marco Iacoboni, existen dos grandes tipos de neuronas espejo: las normales y las superneuronas espejo (*Las neuronas espejo*, 196). Parece ser, en efecto, que en los sistemas cerebrales más complejos, como el de los seres humanos, existe una capa de neuronas en el lóbulo frontal que se encarga de modular y controlar la actividad de las neuronas espejo “simples”. Lo interesante es que muchas veces ese “control” y esa “modulación” están fuertemente influenciados por la **cultura adquirida** del hablante, de la que se habló en 2.2.

Gracias a la visualización de las pirámides etnolingüísticas y al análisis contrastivo entre el español y el italiano (4.4), se tratará de demostrar que muchas expresiones son aceptadas por una comunidad lingüística de acuerdo con la cosmovisión y el sistema de creencias de sus hablantes, y las neuronas espejo funcionan como un poderoso filtro de toda esa información.

### 2.3.2. LOS MEDIOS DE COMUNICACIÓN Y LA PUBLICIDAD

Sin lugar a duda, el mundo ha sufrido un cambio radical desde la llegada de internet y el consiguiente auge tecnológico, y en los últimos años se tiene la sensación palpable de que estamos mucho más “conectados” con el resto de la humanidad. Por eso mismo, la **cultura compartida** ha ganado mucho protagonismo y se ha consolidado al lado de la **cultura adquirida**, y eso gracias en gran medida a las neuronas espejo. Parece ser que nuestro cerebro tiene la posibilidad de compartir cosmovisiones y mentalidades diferentes de manera más frecuente y con mayor facilidad respecto a hace veinte años, y eso favorece muchísimo nuestro proceso

natural de imitación. En segundo lugar, este creciente “cosmopolitismo” ha fomentado el intercambio cultural entre diferentes comunidades, lo que ha llevado, siguiendo la idea que defendemos de la bidireccionalidad recíproca entre lengua y cultura, a influencias directas de la cultura en la lengua y viceversa. Además, estamos sometidos a diario a una presión mediática sin precedentes y a unas políticas de *marketing* cada vez más agresivas. A este respecto, el autor danés Martin Lindstrom explicó que las neuronas espejo influyen directamente en el proceso de compra de los consumidores, y asoció el éxito desmesurado de los aparatos de Apple con la actividad de dichas neuronas, ya que mucha gente se decantó por la compra tras ver a otras personas llevándolos por las calles de Nueva York (*Buyology, verdades y mentiras de por qué compramos*, 20). El propio Apple se encargó de generar un status en las personas que utilizaban sus dispositivos, lo que en muchos casos nos ha impulsado a querer formar parte de ese grupo elitista de personas. Sabemos que existe una disciplina llamada *Neuromarketing* que analiza el comportamiento y los hábitos de los consumidores para poder empatizar con los clientes y poder alcanzar un mayor público objetivo, y así como buscamos los productos que nos generan confianza, también reproducimos expresiones con las que nos sentimos identificados. De alguna manera, las aceptamos, luego las “compramos” y las añadimos a nuestro vocabulario activo. Para nosotros, lo asombroso de todo esto es el hecho de que **el proceso de imitación no se debe solo a una cuestión social, como se ha defendido en muchas ocasiones, sino que influye de manera determinante el aspecto biológico del cerebro**. Dicho de otra manera, estamos programados para imitar y reproducir todo lo que dictaminen nuestras neuronas espejo, y cuanto más estamos en contacto con el resto del mundo, más cultura compartida vamos a generar en nuestro cerebro. Aquí desempeñan un rol fundamental los medios de comunicación, canal vehicular de la publicidad<sup>15</sup>; esta última, bien de manera directa (anuncios en los medios), bien de manera indirecta (por ejemplo, objetos que vemos en otras personas, quienes a su vez los han comprado tras verlos en los medios), somete nuestro cerebro a una estimulación constante con imágenes, frases y temáticas específicas.

A primera vista, podría parecer que nos estamos desviando de nuestro objetivo principal, que es ceñirnos a cuestiones meramente lingüísticas, pero en realidad no lo estamos haciendo. Simplemente, se ha considerado oportuno crear un puente para poder entender mejor los capítulos que vendrán a continuación y poder justificar con mayor solidez la elección de los tres ámbitos culturales que se utilizarán como base de las pirámides etnolingüísticas a partir del *brainstorming* (4.2.1 y 4.3.1). Como decíamos, la publicidad suele aprovechar ámbitos sociales muy concretos para buscar esa complicidad y esa empatía con los receptores, y uno de los más buscados, al menos en Italia y en España, es el fútbol, una disciplina que ha trascendido los meros ámbitos deportivos y ha pasado a formar parte de nuestra vida cotidiana. Hay numerosas expresiones relacionadas con este deporte y hoy en día se escuchan con relativa frecuencia. En español tenemos, entre otras, *meter un gol*, *ganar por goleada*, *ser un maestro del regate corto*, *ser un crack*, *este partido lo tenemos que ganar entre todos*, *eso es de tarjeta roja*, *casarse de penalti* (esta última menos frecuente, ya que la gente ya no se casa tanto por obligación o presión social), etc. En italiano también hay bastantes y nos encontramos con *fare un gol*,

15. Una vez más, han sido muy esclarecedores al respecto los trabajos surgidos del proyecto LiP (LLengua i Publicitat), de la Universitat de Barcelona, coordinado por la Dra. Lúcia Pons i Griera.

*prendere un palo, salvarsi in corner, dribblare qualcosa* (del inglés *dribbling*), *essere da cartellino rosso*, etc. El *Neuromarketing*, en su intento de empatizar con los consumidores para alcanzar el mayor público objetivo, a menudo recurre a estas expresiones para buscar cierta empatía con la gente, y el resultado es que se acaban utilizando incluso con más frecuencia que antes. A veces incluso se recurre a la creatividad y se acuñan nuevas expresiones, bien apoyándose en algunas ya existentes bien creándolas desde cero, y su fosilización en el lenguaje es absolutamente impredecible y depende de innumerables factores, entre ellos de que se conviertan o no en memes, como se verá en el siguiente subapartado.

### 2.3.3. EL PODER DE LOS MEMES

El término *meme* apareció por primera vez en 1976 en el libro *The selfish gene*, de Richard Dawkins quien lo definió como “unidad de transmisión cultural o unidad de imitación” (*The selfish gene*, 5). Frente al gen, que es intrínseco al individuo, el meme tiene que ver con la mimesis, y por tanto necesita ser aprendido; es decir, está vinculado directamente con las neuronas espejo que se descubrirían en Parma quince años después de la publicación del libro de Dawkins. El diccionario de la Real Academia Española no recoge esta palabra, pero si la introducimos en la red como corpus para realizar una búsqueda libre nos saldrán muchísimos enlaces, la mayoría de ellos relacionados justamente con el mundo del fútbol, un ámbito cultural que se ha fosilizado en nuestra sociedad y que se ha convertido en un deporte de masas que acapara gran parte de la atención ciudadana. Pero, ¿qué es realmente un *meme*? Podríamos definirlo como algo que se propaga entre las personas y es compartido de manera unánime en una comunidad, sea un cuadro, un dibujo, una expresión, etc. El escritor español Daniel Ruiz García le dedica un largo párrafo en su novela *Todo está bien* y afirma que “los grandes memes son aquellos que permanecen en el tiempo y son masivos. Algunos de ellos son muy antiguos, como por ejemplo la canción de Cumpleaños Feliz, un refrán popular o el Padrenuestro” (*Todo está bien*, 197). Pero si ya existían antes, con la llegada de internet los memes se han multiplicado y en varias ocasiones se han reproducido de forma viral. En la actualidad, un *twit* de un futbolista o de un político, una determinada campaña publicitaria, un vídeo en *youtube* o una foto en *instagram* son memes potenciales, y el hecho de que se conviertan en virales depende de factores azarosos. En el mundo del fútbol, vemos que a diario se crean nuevos memes que encuentran mayor o menor fortuna en las redes sociales, pero la religión también es un caldo de cultivo importante. Por ejemplo, la expresión *llamar al pan, pan, y al vino, vino* es un meme y procede del ámbito religioso. Si realizamos una búsqueda libre en Google con esta expresión, salen alrededor de 297.000 resultados:



Imagen 1. Consulta de la expresión *llamar al pan pan y al vino vino* en Google (23 de septiembre de 2015)

En italiano existe un meme casi idéntico que es *dire pane al pane e vino al vino* (solo varía el verbo *dire* en lugar de *llamar*). Si buscamos en la red como corpus, nos salen alrededor de 900.000 resultados; es decir, un número muy superior al de la forma española:



Imagen 2. Consulta de la expresión *dire pane al pane e vino al vino* en Google (23 de septiembre de 2015)

Finalmente, **creemos que el hecho de que un meme se vuelva viral se debe muchas veces a la cultura compartida y a cómo entramos en contacto, por ejemplo, con expresiones procedentes de otros países con las que nos identificamos por razones culturales.** A partir

del grado de empatía que generen en nosotros, nuestras superneuronas espejo se encargarán de procesar la información y de decidir si son viables o no para su uso. Si a esto se le añade el rasgo semántico de la **creatividad** del que hablaba Karl Vossler y que se detalló en 2.2.4, el potencial lingüístico del ser humano parece no tener límites.

### 2.3.4. NEUROEDUCACIÓN

No se podía terminar este apartado sobre las neuronas espejo sin dedicar un breve subapartado a la **neuroeducación**, una disciplina bastante reciente que intenta aunar e integrar los procesos educativos con aspectos de la neurociencia. A este respecto, el neurocientífico Francisco Mora afirma que “repetir una y otra vez datos hasta memorizarlos no es el mejor camino para favorecer el aprendizaje y muchos estudios científicos han demostrado que la emoción, la sorpresa y la experimentación son algunos de los ingredientes necesarios para facilitar el conocimiento” (*Neuroeducación*, 21). Las emociones generan ese mecanismo básico que llamamos curiosidad, con el que se expande el abanico de conductas y el interés por descubrir lo nuevo. Son, en definitiva, el ingrediente secreto del aprendizaje, tanto para quien enseña como para quien aprende; además, la manifestación de las emociones permite la creación de lazos (familiares o de amistad) que aseguran la supervivencia tanto biológica como social (Mora, *Neurocultura*, 39). En un estudio realizado por Marco Iacoboni y Mirella Dapretto y publicado en la revista *Nature* en 2006, se demostró que cuanto mayor era la afinidad entre el profesor y los alumnos, mejor era el aprendizaje y la actitud de estos últimos (“The mirror neuron system and the consequences of its dysfunction”, 7). Se llegó a demostrar que existen dos polos opuestos de la empatía: en un primer caso, el profesor entra sonriente y lleno de energía en el aula, se pone a explicar con entusiasmo y enseguida se genera un ambiente positivo, lo que permite empatizar con los alumnos y motivarlos; en el segundo caso, el profesor entra cabizbajo y resoplando, transmite tensión y explica sin ganas, lo que genera un clima de desconfianza que produce mucha inseguridad y descontento en el alumnado. En el primer caso, las neuronas espejo reconocen a un referente, mientras que en la segunda situación lo niegan, pues la desconfianza actúa como inhibidora. Dicho de otra manera, el estado emocional del aula depende del profesor, de su grado de empatía y de sus gestos, pero también de su capacidad para leer determinadas situaciones conflictivas, impidiendo que el comportamiento de algunos alumnos resulte molesto para los otros (Ambady y Rosenthal, “Half minute: predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness”, 64). Siguiendo esa línea, las neurocientíficas Sarah-Jayne Blakemore y Uta Frith resumen muy bien este nuevo enfoque educativo al afirmar que “en todo proceso de aprendizaje, los valores, las ideas y la actitud del maestro podrían ser tan importantes como el material que está enseñando” (*Cómo aprende el cerebro: las claves para la educación*, 220). Estos comentarios enlazan con el presente trabajo de investigación porque el método que queremos proponer nació de manera fortuita durante una clase de español como L2, como se detallará más adelante en 2.4.4, y porque pensamos que la neuroeducación puede ilustrar mejor la base de nuestra propuesta y justificar la creación de las pirámides etnolingüísticas.

## 2.4. ALGUNOS ASPECTOS SEMÁNTICOS EN LAS LENGUAS NEOLATINAS

Para el desarrollo de esta tesis de investigación, el aspecto semántico desempeña un papel de primerísimo orden, puesto que, como apunta el Dr. Juan Pablo García-Borrón<sup>16</sup>, analizando el significado de un vocablo podemos extraer información cultural y etnolingüística. Todo hablante presenta tres tipos de vocabularios bastante definidos: el **activo**, el **pasivo** y el **virtual** (*Semántica de la palabra*, 9). El primero refleja nuestro idiolecto; es decir, ese conjunto de palabras que conocemos y que utilizamos cuando hablamos, reflejando nuestra propia manera de expresarnos. El vocabulario pasivo, como bien apunta el término, se refiere a todas esas palabras que conocemos y entendemos, pero que no utilizamos habitualmente. Son palabras que escuchamos a diario, bien en la calle bien en los medios de comunicación, pero que después no usamos. Por último, tenemos el vocabulario virtual, quizá el más curioso de los tres. Se trata de un conjunto de palabras que desconocemos, pero que entendemos por el contexto. Por ejemplo, durante años los aviones no aterrizaban en el mar, hasta que se inventó el hidroavión y se creó el verbo *amerizar* sobre la base del ya existente *aterrizar*. Lo mismo ocurrió con la palabra *alunizar*, cuando el hombre demostró por primera vez que sí era posible llegar a la superficie lunar. Se trata, pues, de términos que están virtualmente disponibles y que se forjan a partir de determinadas situaciones. Como recuerda el Dr. García Borrón, en enero de 2006 en España se prohibió fumar en el trabajo, y a partir de entonces en las farmacias se podía leer el eslogan *desaprende a fumar* (*Semántica de la palabra*, 10). Este sería un claro ejemplo de vocabulario virtual, ya que el verbo *desaprender* no existía antes desde el punto de vista de la norma, pero era virtualmente posible. Respecto a ello, nos parece de sumo interés la aportación de DispoLex, un proyecto panhispánico que estudia el léxico disponible en las diversas zonas del mundo hispánico con el fin de establecer comparaciones de tipo lingüístico, etnográfico y cultural<sup>17</sup>. Se entiende que el léxico disponible es ese conjunto de palabras que los hablantes tienen en su lexicón mental y cuyo uso está condicionado por la comunicación, y ayuda muchísimo a la hora de descubrir qué palabras sería capaz de utilizar un hablante en ámbitos comunicativos concretos.

El colombiano Félix Restrepo publicó en 1917 un interesante ensayo donde sostenía, entre otras cosas, que una lengua viva no consta solamente de palabras, sino también de moldes para hacer palabras (*El alma de las palabras. Diseño de semántica general*, 65). En español, por ejemplo, tenemos prefijos como en- o des- y sufijos en -zar que son muy rentables para añadir nuevas palabras al vocabulario virtual. Evidentemente, el contexto sociocultural es un caldo de cultivo perfecto para la creación de nuevas palabras y, si pensamos en la llegada de internet, veremos que desde hace unos años palabras como *chatear*, *twitear* o *googlear* son muy frecuentes en el habla oral. El hecho de que una palabra de reciente creación alcance lo que se define un “éxito lingüístico” se debe en gran medida a factores sociales, y es posible gracias

16. Tuve la suerte de tenerlo como profesor de *Semántica Española* y de *Introducción a la Historia de la Lengua Española* durante la carrera de Filología Hispánica y muchas de las ideas que aparecen en esta tesis proceden de sus clases.

17. <http://www.dispox.com>

a la labor de las neuronas espejo. Escuchamos, empatizamos tanto social como culturalmente y, finalmente, imitamos.

Además de los tres tipos de vocabularios, recuerdo que el Dr. García-Borrón hizo particular hincapié, en una de sus clases, en los enfoques sincrónicos y diacrónicos de la semántica. Para esclarecer los conceptos de **extensión** e **intensión**, que se profundizarán en el siguiente subapartado y que se tendrán muy en cuenta en el análisis contrastivo (4.4), puso como ejemplo la palabra negro, procedente del latín *nīgrum*. En latín, *niger* se opone a *ater* según negro *brillante/negro mate*, mientras que en español solo tenemos la palabra *negro*, pues parece que no hay necesidad de diferenciar. Por consiguiente, la palabra española tiene más intensión y menos extensión, mientras que en latín pasaba exactamente lo contrario. Somos conscientes de que se trata de un enfoque meramente diacrónico, pero a partir de aquí nos hemos preguntado qué ocurre desde un punto de vista sincrónico entre lenguas neolatinas, y creemos que con la visualización de las pirámides etnolingüísticas se pueden apreciar mejor algunas de esas diferencias de intensión/extensión entre el italiano y el español” (objetivo C).

### 2.4.1. EXTENSIÓN VS INTENSIÓN

Como se adelantó en el subapartado anterior, tanto la **extensión** como la **intensión** en semántica deberían ser dos aspectos relevantes en todo estudio contrastivo. Según el filósofo Rudolf Carnap, defensor del positivismo lógico, la **intensión** sirve para analizar una expresión semánticamente con el fin de entender su significado, mientras que la **extensión** nos permite ahondar en la situación a la que se refiere la expresión, para así determinar un contexto (*Meaning and necessity. A study in semantics and modal logic*, 22-26). La extensión, añade Carnap, nos proporciona una “verdad lógica” sobre el contexto comunicativo y depende de cómo sea la realidad, mientras que la intensión se ciñe únicamente al lenguaje. Para poner un ejemplo y siguiendo con los colores, pensemos en el adjetivo *blanco*. Para un español y un italiano esa palabra tiene mucha intensión, ya que su significado es muy amplio y puede referirse a muchas cosas; sin embargo, para un esquimal, acostumbrado a vivir en un ambiente donde el blanco es la tonalidad principal, dicha palabra tiene muy poco significado y se necesitan diferentes vocablos. Así pues, resulta que un esquimal tiene siete palabras para las distintas tonalidades de blanco, mientras que un español cubre sus necesidades lingüísticas con una<sup>18</sup>. Esto demuestra que intensión y extensión son inversamente proporcionales: a mayor extensión, menor intensión y viceversa.

En esta tesis se apoya esta teoría y pensamos que la **extensión** y la **intensión** están relacionadas principalmente con la **cultura adquirida** (2.2); es decir, con ese enorme bagaje cultural que recibimos en nuestros primeros años de vida, en gran medida a través del lenguaje, y que de alguna manera condiciona nuestra cosmovisión.

Estos conceptos están estrechamente vinculados con la idea de “necesidad lingüística” y serán la base de las pirámides. Es probable, por ejemplo, que si creáramos una pirámide de los

18. En español existen otras palabras para referirse al color blanco, pero son muy sectoriales y de uso menos común, de ahí que no las hayamos tenido en consideración. Un ejemplo sería el término *bayo*, que se utiliza para describir el color blanco/amarillento de los caballos.

colores e hiciéramos un contraste entre una lengua neolatina y una de los esquimales, veríamos que en la segunda predominan expresiones con la palabra *blanco*, ya que, al tener más extensión, abarca más contextos comunicativos y dispone de un mayor número de vocablos<sup>19</sup>.

Los análisis contrastivos están llenos de ejemplos parecidos y de alguna manera nos aportan una valiosa información socio y etnolingüística. En el ámbito religioso hay casos bastante curiosos, como el de la palabra *deus* en latín, que ha dado *dios* en castellano y *dio* en italiano. Por lo visto, en latín la palabra tenía mucha extensión y poca intensidad, ya que los romanos tenían una religión politeísta con muchos dioses diferentes. En países de tradición católica como Italia y España, en cambio, se supone que solo hay un Dios, con lo que el término gana muchísima intensidad y pierde extensión. En definitiva, podemos colegir que en todo análisis contrastivo nos vamos a encontrar con el fenómeno del **anisomorfismo semántico**, que tiene que ver con las diferencias que se dan en la malla semántica de cada lengua. A veces, estas diferencias incluso se dan dentro de la misma lengua, como es el caso de los vocablos *aprobado/suspense*. En un examen de conducir, por ejemplo, solo tenemos las calificaciones *aprobado/suspense*, mientras que en un examen universitario la malla es más amplia y nos encontramos con *Matrícula de Honor, Sobresaliente, Notable, Aprobado y Suspense*. Si cotejamos los dos sistemas, podemos ver que existe una equivalencia con *Suspense*; sin embargo, en la universidad hay cuatro vocablos que sirven para matizar los distintos niveles de *aprobado*, lo que indica una mayor **extensión** y una menor **intensidad** respecto al examen de conducir. Existen un sinnúmero de casos como estos, tanto en diacronía como en sincronía, pero nuestro estudio se va a centrar principalmente en las diferencias sincrónicas debido a que los datos recogidos y analizados (Data-driven) para crear las pirámides están enmarcados en los últimos quince años (2001-2015), tal y como se ha justificado en el apartado 1.2, donde se presentaron los objetivos de esta tesis. Es cierto que en algunos casos nos podemos apoyar en la evolución diacrónica para desambiguar, pero siempre dando prioridad al aspecto sincrónico, pues creemos que es el que mejor refleja las verdaderas diferencias etnolingüísticas que se dan en la actualidad entre el español y el italiano y que constituirán el fundamento del tercer objetivo (C) de este trabajo de investigación.

### 2.4.2. LA INTENSIDAD

Todas las lenguas “vivas” son herramientas indispensables que los hablantes utilizamos a diario para comunicarnos con nuestros semejantes, por eso están sujetas a un desgaste natural que se ve reflejado de manera inequívoca en las palabras. Dicho de otra manera, una lengua necesita cambiar, y se transforma justamente porque está viva. En muchas ocasiones, ese cambio se da por razones de **intensidad**, un elemento que se va a tener muy en cuenta en esta tesis. En palabras de la Dra. Josefa Martín García, “la intensidad supone una mayor carga intencional, emotiva y cuantitativa del contenido significativo de una palabra, por lo que la intensificación se enmarca dentro de la subjetividad del hablante y añade rasgos connotativos al significado denotativo de una palabra (*Los prefijos intensivos del español: caracterización*

19. Véase hipótesis sobre el ámbito de los toros al final de 1.2.

*morfo-semántica*, 103). Según la Dra. Martín García, las categorías gramaticales más sujetas a ser intensificadas, y por lo tanto más predispuestas para el cambio, son los **sustantivos, los verbos, los adjetivos y los adverbios**<sup>20</sup>.

Una lengua viva, por lo tanto, está en constante transformación, y la intensidad desempeña un rol fundamental en este proceso. Cuando una palabra o una expresión se considera “desgastada”, aparece en el hablante una necesidad de cambio que origina dos procesos: el primero de *innovación*, en el que se crea la “nueva versión”, y el segundo de *adopción*, en el que la palabra es recibida de manera favorable o no por la comunidad de hablantes. Si la adopción es favorable, la palabra cuaja en la lengua y se completa el cambio. Esto ocurre también con los memes y con muchas expresiones procedentes de ámbitos culturales específicos. Se ha mencionado en el apartado 2.2.4 el idealismo lingüístico de Karl Vossler y en el apartado 2.3 se han descrito las neuronas espejo: la relación entre ambos es mucho más estrecha de lo que pueda parecer a primera vista y enlaza directamente con este concepto de **intensidad** que se está analizando ahora. Como se vio a lo largo de 2.3, las neuronas espejo funcionan como filtro en nuestro cerebro y son las principales responsables de la imitación y de la empatía; además, hay unas neuronas superiores que Marco Iacoboni denominó superneuronas espejo (*Las neuronas espejo*, cap.7), las cuales determinan la viabilidad y la rentabilidad de todo lo que llega a nuestro cerebro. Gracias a esta función, estamos capacitados para determinar la intensidad de una expresión.

Imaginémonos por ejemplo una conversación prototípica entre un grupo de jóvenes que están intercambiando opiniones sobre un partido de fútbol. El primero de ellos dice que el partido ha sido *muy guapo* y, como el adjetivo *muy* ha sufrido un gran desgaste y ha perdido intensidad, es muy probable que otro joven, cuyo cerebro busque una mayor intensidad, continúe la conversación afirmando que el partido ha sido *superguapo*, pero tenemos un tercero, con afán de subir todavía más el grado de intensidad, que remata la conversación diciendo que *ha sido un partido de la hostia*. También es probable que alguien utilice la expresión *ha sido un partido cojonudo*, y eso demuestra que para un mismo contexto comunicativo la lengua ofrece un amplio abanico de recursos, que además se presta a la creatividad del ser humano, relacionada con el idealismo lingüístico que proponía el propio Vossler y que se analizó en 2.2.4. En definitiva, imitamos y luego creamos, muchas veces condicionados por una necesidad lingüística o por conseguir una mayor intensidad; se trata de una necesidad **biológica** del propio ser humano, y este proceso es lo que realmente mantiene viva una lengua y la hace evolucionar en el tiempo, ofreciendo además un enorme caudal de información cultural y social que puede resultar de suma importancia de cara a un estudio contrastivo.

### 2.4.3. SINONIMIA Y POLISEMIA

Todo estudio contrastivo acaba topándose, tarde o temprano, con el delicado problema de la

20. Se trata de una afirmación demostrable, sobre todo en las lenguas románicas: pensemos por ejemplo en el adjetivo *muy* delante de un adjetivo. Es evidente que ha ido perdiendo intensidad y vemos que han nacido nuevas formas tanto en italiano como en español que aportan más intensidad al lenguaje (*superguapo/strabello*). Hay casos incluso en los que ha cambiado también el adjetivo, recurriendo a un préstamo de otra lengua, como en el caso de *superheavy*.

sinonimia y la polisemia, sobre todo a la hora de buscar equivalencias de significado. Como es bien sabido, cuando un mismo significante tiene varios significados nos encontramos frente a un caso de **polisemia**, mientras que cuando dos significantes diferentes comparten un mismo significado estamos ante una **sinonimia**. Respecto a esta última, la mayoría de los lingüistas coinciden en que resulta difícil que dos palabras tengan el mismo significado en todos los contextos, especialmente en la lengua común. ¿Existen por lo tanto sinónimos plenos en una misma lengua? Según Stephen Ullmann, **muy pocas palabras son completamente sinónimas en el sentido de ser intercambiables en cualquier contexto sin la más leve alteración del significado objetivo, el tono sentimental o el valor evocativo**. Apunta que la sinonimia completa es muy rara y solamente la podemos encontrar en algunas nomenclaturas técnicas, como podría ser el caso de los dos términos médicos *cecitis* y *tiflitis* (el primero procedente del latín y el segundo del griego) para referirse a la inflamación del intestino ciego. Además, Ullmann menciona al profesor William Edward Collinson, quien realizó la siguiente clasificación de los sinónimos, en la que podemos observar los posibles matices que diferencian un término de otro y que nos parece de gran interés (*Semántica. Introducción a la ciencia del significado*, 160-170)<sup>21</sup>:

- 1) un término es más intenso que otro. Ej: *grotesco* (+) vs *ridículo* (-);
- 2) un término es más general que otro. Ej: *pelo* (+) vs *cabello* (-);
- 3) un término es más emotivo que otro. Ej: *repugnante* (+) vs *repelente* (-);
- 4) un término puede implicar aprobación o censura moral, mientras que el otro es neutro. Ej: *dilapidar* (+) vs *gastar* (-);
- 5) un término es más profesional que otro. Ej: *domicilio* (+) vs *casa* (-);
- 6) un término es más literario que otro. Ej: *amar* (+) vs *querer* (-);
- 7) un término es más coloquial que otro. Ej: *currar* (+) vs *trabajar* (-);
- 8) un término es más local o dialectal que otro. Ej: en Valencia *manises* (+) vs *azulejos* (-);
- 9) uno de los sinónimos es más infantil que otro. Ej: *papi* (+) vs *padre* (-).

John Lyons, uno de los lingüistas más importantes del siglo XX, postula dos condiciones para que dos términos sean considerados sinónimos: la primera es que sean permutables en todos los contextos; la segunda, que tengan el mismo valor cognitivo y emotivo. En el caso de que se cumpla la primera condición, hablaremos de **sinonimia total**, mientras que con la segunda tendríamos una **sinonimia completa** (*Introducción al lenguaje y a la lingüística*, 128-135). La polisemia también desempeña un rol fundamental en este trabajo de investigación y durante el análisis contrastivo nos encontraremos con muchas palabras que en una lengua tienen más significados que en la otra, lo que en algunos casos generará cierta ambigüedad y falsos amigos que pueden llevar a confusión, de ahí que hagamos particular hincapié en estos con-

21. Esta clasificación, por supuesto, es matizable, y cabe añadir que también nos podemos encontrar con sinónimos que se diferencian por su etimología, como es el caso de escorpión y alacrán; el primero procedente del latín, el segundo del árabe.

ceptos y nos explayemos recalcando su importancia en esta tesis<sup>22</sup>.

Ullmann, siempre en su libro *Semántica, introducción a la ciencia del significado*, menciona al fonetista francés Léonce Roudet (1861-1935), quien fue uno de los primeros en proponer una clasificación basada en la  **semejanza**  y la  **contigüidad**  para distinguir los cambios de significado según si se basan en una asociación entre nombre o entre sentido (*Semántica, introducción a la ciencia del significado*, 238-257). La semejanza suele producir una metáfora, mientras que la contigüidad crea una metonimia. En el caso de *falda*, por ejemplo, podemos decir que cuando se refiere a la parte baja de un monte es una metáfora originada por semejanza, mientras que la prenda femenina se relaciona con la mujer por contigüidad; se trata, pues, de una asociación mental que relaciona ambas ideas pese a que no tengan ninguna semejanza. En la palabra italiana *gonna* solamente tenemos contigüidad, sin que haya ninguna metáfora, por eso el término es monosémico.

Finalmente, la polisemia y la sinonimia favorecen enormemente la ampliación de la malla lingüística de una lengua, haciéndola mucho más elástica, y un ejemplo claro es la denominada  **irradiación sinonímica** , un recurso de la lengua que consiste en el desarrollo paralelo de dos sinónimos que amplían su significado, muchas veces para conseguir un mayor grado de intensidad<sup>23</sup>. Además, la existencia de la polisemia introduce, en palabras de Ullman, un elemento de ductilidad en el lenguaje. No hay nada definitivo en relación con el cambio semántico y una palabra puede adquirir un nuevo sentido, o veintenas de sentidos nuevos, sin perder su significado original (*Semántica, introducción a la ciencia del significado*, 218). Como ya se ha visto en 2.3.3, no solo imitamos, sino que además buscamos crear constantemente nuevas estructuras para que la lengua conserve su vitalidad; además, se considera que la cultura tiene una influencia determinante sobre muchas palabras y sobre el hecho de que sean polisémicas, por eso confiamos en que la visualización de las pirámides etnolingüísticas pueda ayudar a esclarecer algunas diferencias esenciales entre el italiano y el español (objetivo C), las dos lenguas que se van a comparar en este estudio a lo largo de 4.4.

#### 2.4.4. EL BRAINSTORMING

Algunos de los elementos principales de la semántica estructural son los  **campos léxicos** , los  **campos semánticos**  y las  **familias léxicas** . Según Eugenio Coseriu, el primero se refiere a ese conjunto de palabras que comparten (más allá de sus matices y diferencias) una parte común de significado (*Principios de semántica estructural*, 135-137). Un ejemplo de campo léxico podría ser el del aula, donde incluiríamos las palabras  *estudiar* ,  *aprobar* ,  *suspender* ,  *enseñar*  (verbos),  *profesor* ,  *alumno* ,  *pizarra* ,  *cuaderno* ,  *libro*  (sustantivos),  *intensola* ,  *fácil* ,  *difícil*  (adjetivos),

22. Por ejemplo, la palabra *falda* es polisémica en español, ya que puede referirse a la prenda femenina y también a la parte baja de un monte, pero el equivalente italiano *gonna* solo se refiere a la prenda femenina y no es polisémico.

23. Un claro ejemplo serían las palabras *punteo* y *acueducto*: ambas se refieren, con su significado principal, a un determinado tipo de construcción y, sin ser sinónimos, pueden tener cierto parecido, aunque el acueducto sea de mayor tamaño. El caso es que en español existe la locución adverbial *hacer punteo*, que se refiere a aprovechar algún día intermedio entre dos fiestas para coger vacaciones (significado 2, por semejanza), y por irradiación sinonímica tenemos *hacer un acueducto* cuando los días de vacaciones se alargan un poco más de lo normal.

etc. El campo semántico, en cambio, se caracteriza por reunir palabras que pertenecen a la misma categoría gramatical (campo semántico de las flores: *amapola*, *geranio*, *azucena*, *clavel*, *lirio*, *rosa*, etc.) y por tener un hiperónimo o archilexema y varios hipónimos (la palabra que está contenida en el hiperónimo, por ejemplo *rosa*). Por último, tenemos la **familia léxica**, constituida por palabras que comparten un mismo lexema, como es el caso del lexema mar, a cuya familia pertenecen vocablos como marino, marítimo, ultramar, marinero, etc.

Para sacar partido a estos campos, una opción es recurrir al *brainstorming* (fase 1 del proceso de creación de las pirámides etnolingüísticas), una técnica propuesta por primera vez por Alex Faickney Osborne en su libro *Applied imagination*, publicado por primera vez en 1953 y traducido al castellano en 1957. La técnica, que ganó rápidamente popularidad en los años sesenta, se basa principalmente en la idea de romper las limitaciones habituales del pensamiento y producir un conjunto de ideas entre las que poder escoger. El *brainstorming* es útil para generar una gran cantidad de ideas y el mismo Osborn propuso cuatro reglas básicas para que sea lo más efectivo posible: en primer lugar, hay que **suspender el juicio** para eliminar toda crítica consciente, ya que según el autor es muy complicado crear y juzgar al mismo tiempo y es mejor dejar que nuestro pensamiento fluya sin interferencias. Esta primera regla, en realidad, es quizá la más complicada de las cuatro, pues estamos tan acostumbrados a analizar todo lo que hacemos que no resulta fácil dejarse llevar. Esto nos conduce a la segunda regla, que consiste en **pensar libremente** y en no descartar ninguna idea *a priori*, por muy imaginable, confusa y disparatada que parezca. La tercera regla tiene que ver con la **cantidad**, puesto que uno de los objetivos principales es concentrarse para sacar el mayor número posible de ideas, aunque luego se descarten. Osborn sitúa el objetivo ideal entre 50 y 100 ideas, y el objetivo mínimo por encima de las 25 ideas. Por último, hay que reunir todas las ideas buscando el denominado **efecto multiplicador**; es decir, buscar la combinación de todas las ideas y tratar de mejorarlas. Por eso la cantidad es importante y es directamente proporcional al efecto multiplicador: cuantas más ideas tengamos, más posibilidades tenemos de pulirlas y de dar con la solución más idónea para nuestro propósito. En cuanto al grupo, se estableció que no puede superar los **doce miembros**, situándose entre **cuatro y siete el número ideal**, aunque lógicamente el *brainstorming* también se puede realizar de manera individual. Como herramientas solo necesitamos un bolígrafo y una hoja en blanco, mejor de tamaño DIN A3 que DIN A4, ya que cuanto más espacio en blanco tengamos, mayor será el alcance de nuestra creatividad<sup>24</sup>.

Además de todo esto, es fundamental que el ambiente del grupo sea ameno y que los participantes estén relajados y tengan la sensación de que están disfrutando. Y es que, como se apuntaba en el subapartado 2.3.4 sobre Neuroeducación, las emociones desempeñan un papel fundamental en la motivación y en la estimulación de la creatividad, por eso un grupo alegre tiene siempre un potencial mucho más alto respecto a un grupo tenso y cohibido. Lo más importante es que los participantes se suelten y den rienda suelta a su imaginación, así que hay que presentarles el *brainstorming* como algo lúdico y no como algo serio o competitivo que pueda generar una tensión en el cerebro. Todo tiene que fluir de manera natural y despreocupada, y

24. Se ha decidido apostar por el tamaño DIN A3 y se puede afirmar que los resultados han sido muy satisfactorios, como se verá en los apartados 4.2.1 y 4.3.1.

es importante además que haya un **moderador** para controlar el proceso, un papel que en el aula puede desempeñar el profesor. La figura del moderador es determinante en el tipo de *brainstorming* que vamos a utilizar para la creación de las pirámides etnolingüísticas, ya que es el principal responsable del ambiente que se llegue a crear entre los participantes. Necesita ser una persona empática y con herramientas suficientes para suavizar cualquier tipo de crispación que exista en el grupo; de otra manera, el resultado final de la prueba puede verse comprometido. Respecto a la duración de la sesión, Osborn recomendó que no superara los **treinta minutos** para no correr el riesgo de que los participantes perdieran el interés, razón por la que nos inclinamos a realizar sesiones incluso más breves, de veinte minutos máximos, dedicando luego entre diez y quince minutos para evaluar todas las ideas, esta vez de manera lógica y analítica.

Cabe subrayar que esta tesis se nutre de dos *brainstorming*: uno en su génesis realizado en clase con alumnos de L2 y otro en el proceso de investigación con profesores nativos (4.2.1 y 4.3.1)<sup>25</sup>, que representa el punto de partida de nuestra triangulación metodológica. El primero nació en una clase de idiomas a la que llegué con la carpeta equivocada debido a una serie de imprevistos. En lugar de resoplar y de adosarme culpas innecesarias delante de los alumnos, me inventé sobre la marcha una actividad didáctica que consistía en dividir la clase en dos grupos y ambos tenían que escribir todas las palabras que se les ocurrieran relacionadas con dos ámbitos culturales bien definidos: el fútbol y la religión. Les di veinte minutos y les ofrecí la posibilidad de consultar el diccionario<sup>26</sup>. El resultado fue satisfactorio y ambos grupos consiguieron un buen número de palabras, así que pasamos a la siguiente parte del ejercicio, que consistía en encontrar al menos una expresión pluriverbal por cada palabra que habían encontrado. Pensé que la mejor forma para acceder a esa información pasaba por consultar un corpus, así que, como todos tenían ordenadores, les enseñé a realizar búsquedas en el CORDE como ayuda complementaria a los diccionarios. Disponían de veinticinco minutos y ganaba el grupo que conseguía encontrar el mayor número de expresiones en el tiempo establecido. El resultado de este segundo ejercicio fue bueno y los dos grupos estuvieron bastante igualados, pero la nota más positiva para mí fue el hecho de que los estudiantes disfrutaron mucho y se involucraron a fondo de principio a fin en algo que consideraban “nuevo” e “inesperado”. A medida que encontraban expresiones, fue creciendo dentro de ellos cierta curiosidad que iba alimentando su interés por la búsqueda, y muchos de ellos buscaban equivalentes en su lengua materna. La buena acogida del ejercicio me dio varias ideas que luego desembocarían en esta tesis de investigación, pero sobre todo me concienciaron aún más del gran papel que desempeña la empatía durante el proceso de aprendizaje, sobre todo si existen lazos culturales afines que despiertan nuestra curiosidad y por consiguiente nuestro interés.

A partir de aquí comprobé que era posible complementar el *brainstorming* con los corpus lingüísticos, lo que representa el segundo paso de la triangulación metodológica, y así sería más fácil enlazar el flujo de ideas con unas herramientas que reflejan el uso de la lengua real y que pueden aportar una valiosa información social y cultural.

25. Para esta tesis de investigación, se decidió realizar el *brainstorming* con nativos por considerar que podrían aportar un mayor número de palabras.

26. Se trataba de un nivel intermedio/alto (B2) y la mayoría de los estudiantes procedían de países europeos como Francia, Italia, Portugal y Alemania, por lo que compartían algunos aspectos culturales, especialmente con el fútbol.



### **3. TIPOS DE CORPUS Y SU PAPEL EN LAS PIRÁMIDES ETNO- LINGÜÍSTICAS**

### 3.1. LA LINGÜÍSTICA DE CORPUS

La lingüística de corpus se basa en una serie de datos obtenidos a partir de muestras reales de la lengua y constituye por tanto una herramienta muy valiosa de cara a distintos campos de la investigación y de la enseñanza. Ateniéndonos a la definición del DRAE (vigésima segunda edición), en la lengua general se denomina “corpus” a “un conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación”, mientras que para John McHardy Sinclair, uno de los pioneros de la Lingüística de Corpus (LC):

“A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research” (“Corpus and text – Basic principles”, 12).

Para la realización de este proyecto, fue fundamental el hecho de que se trate de una recopilación de textos disponibles solamente en formato electrónico y que además sean representativos de una determinada lengua; sin embargo, aunque hace tiempo que los corpus se emplean para profundizar en la investigación lingüística, muchos docentes y especialistas en la enseñanza de una L2 siguen siendo bastante reacios y no utilizan estas herramientas en sus sistemas educativos. A este respecto, nuestra postura es bastante clara y no solo creemos que pueden complementar el material tradicional para la enseñanza de una L2, sino que además pueden brindarle al estudiante una inmersión mucho más profunda en la lengua y ofrecerle una visión del léxico en contextos reales para sacar conclusiones de forma inductiva a partir de una recopilación de datos (Data-driven), como explica Elena Alonso Pérez-Ávila (“El Corpus Lingüístico en la Didáctica del Léxico del Español como LE”, 11-25).

Las raíces de la lingüística de corpus son muy antiguas y desde hace siglos ha habido un cierto interés por reflejar el uso real de una lengua a través de los **índices de frecuencia**, como demuestra el valioso trabajo del alemán Friedrich Wilhelm Kädling, quien, a finales del siglo XIX, reunió un corpus de más de once millones de palabras del alemán con el fin de extraer la frecuencia de distribución de las letras de esa lengua<sup>27</sup>. Pese a ello, el auge de la lingüística de corpus llegó solo a partir de los años 60, favorecido también por la llegada de la informática y, por lo tanto, de nuevas herramientas que permitían procesar y gestionar un caudal de textos cada vez mayor.

En este capítulo se verán los principales tipos de corpus y explicaremos cuáles de ellos vamos a tener en cuenta y por qué razones, además de detallar su función determinante en el proyecto de las pirámides etnolingüísticas. Algunos autores, como Tony McEneary y Andrew Wilson, han defendido la idea de que, gracias a los corpus, perseguimos que los estudiantes de una L2 infieran conocimiento lingüístico de manera inductiva; es decir, utilizando ejemplos reales para mostrar conceptos, principios y reglas gramaticales (*Corpus Linguistics*, 12). Según ellos, la consulta de un corpus por parte de un estudiante de L2 potencia su conciencia

27. KÄDLING (1879). *Häufigkeitswörterbuch der deutschen Sprache*. Steglitz, publicado de manera privada.

lingüística y lo ayuda a tener una visión más global y completa de la lengua que está aprendiendo. Nosotros compartimos de pleno esta idea, y pensamos además que un corpus es una herramienta muy eficaz que nos ayuda a ahondar mucho más en una lengua, brindándonos algunos aspectos culturales que con un diccionario sería más complicado detectar, por eso en esta tesis se ha optado por demostrar las hipótesis de partida (1.2) a partir de una recopilación de datos (Data-Driven) que se consigue gracias al *brainstorming* (paso 1 de la triangulación metodológica), la búsqueda en los corpus (paso 2) y, por último, la visualización de datos en las pirámides (paso 3).

### 3.1.1. UTILIDAD Y FUNCIÓN PRINCIPAL DE LOS CORPUS

Una de las aportaciones fundamentales de los corpus es la de ofrecer bases reales para el estudio de una lengua, por eso se ha decidido recurrir a ellos como segundo paso de nuestra triangulación metodológica. Gracias a los corpus aspiramos a encontrar el mayor número posible de expresiones pluriverbales (objetivo A) para luego ordenarlas en las pirámides de acuerdo con estos parámetros: el de **frecuencia** y el de **rentabilidad comunicativa** (objetivo B). Sin embargo, toda la lingüística de corpus gira alrededor de un concepto fundamental que no es posible obviar: la **representatividad**. El Dr. Joaquim Rafel plasmó la idea de que la intención de un corpus es ofrecer una muestra representativa que permita al lingüista fundamentar su investigación en datos objetivos. Un corpus no puede identificarse con la lengua, sino que es un conjunto de datos que la representa de una manera más o menos fiable (*El procesamiento de corpus*, 41-73). Todas las muestras, por lo tanto, son de alguna manera tendenciosas, y este detalle va a ser tenido muy en cuenta para matizar el proceso final de elaboración de las pirámides etnolingüísticas (4.2.4 y 4.3.4). La Dr. Mar Cruz, a este respecto, considera la representatividad como la piedra angular de la LC, pues hace posible que podamos extraer conclusiones fiables a partir de datos estadísticos (*Lingüística de corpus y enseñanza de español como L2*, 36). Estas afirmaciones podrían llevarnos a pensar que la representatividad es directamente proporcional al tamaño del corpus, pero eso no siempre es cierto, ya que un corpus no va a ser más representativo por tener más palabras, sino por definir de manera clara y desde un principio qué se quiere representar. Si el objetivo es representar la totalidad de la lengua, entonces es fundamental que el corpus incluya muestras de diferentes registros, así como textos procedentes de todos los territorios donde se habla la lengua en cuestión en proporción con el número de hablantes. En un análisis contrastivo entre el español y el italiano, nos encontramos con dos lenguas que presentan una distribución geográfica muy diferente: el español presenta una amplia variedad geográfica, con lo que resulta más difícil componer una muestra representativa, mientras que el italiano solo es lengua oficial en Italia y en un cantón de Suiza.

Por eso mismo, a continuación se van a analizar diferentes tipos de corpus y se verá cuáles son más idóneos para la realización de este proyecto y por qué razones, ya que se manejará mucha información cuantitativa y consideramos de vital importancia hacer una buena criba y escoger las mejores herramientas lingüísticas para sacar el máximo provecho del *brainstorming*.

### 3.1.2. DIFERENTES TIPOS DE CORPUS

En la LC existen principalmente seis criterios para determinar las diversas tipologías de los corpus: **el canal de la lengua, el grado de representatividad, la especificidad de los textos, el porcentaje y distribución de los tipos de texto, los idiomas que contenga** y, por último, **el nivel de análisis o marcaje**. Se trata de unos criterios que nos ayudan a perfilar mejor la búsqueda y que hacen que nos decantemos por un corpus en lugar de otro según el objetivo final de la investigación. Ateniéndonos al canal de la lengua, tenemos los *corpus textuales* o *escritos* (cuya muestra procede únicamente de textos escritos)<sup>28</sup>, los *corpus orales*<sup>29</sup>, compuestos por conversaciones transcritas, y los corpus mixtos, que incluyen los dos tipos de muestras<sup>30</sup> y que, por lo tanto, se tendrán en consideración para redactar las pirámides etnolingüísticas. También existen los *corpus sonoros*, que recogen grabaciones orales que sirven básicamente para estudios fonéticos del habla y que, por lo tanto, no se van a tener en cuenta en este trabajo.

El segundo criterio, al que se recurrirá en diversas ocasiones, distingue los corpus en base a su representatividad y ofrece tres tipos principales: los *corpus textuales*, que incluyen documentos completos, los *corpus léxicos*, que incluyen fragmentos muy pequeños de cada documento y los *corpus de referencia*, que intentan abarcar el mayor número posible de muestras de una lengua. Estos últimos son los que más nos incumben, pues creemos que nos ofrecen unas muestras más o menos fidedignas del uso de la lengua.

El tercer criterio se refiere a la especificidad de los textos y nos brinda cuatro tipos de corpus bien definidos: los *corpus generales*, que recogen todo tipo de géneros textuales<sup>31</sup>, los *corpus genéricos*, que se centran en un único género, los *corpus canónicos*, que recogen la obra completa de un solo autor y los *corpus especializados*, centrados en un ámbito específico<sup>32</sup>. Evidentemente, para alcanzar los tres objetivos propuestos en esta tesis se utilizarán principalmente los corpus generales, ya que cuantos más géneros tengamos, mejor se reflejará el uso de la lengua real en las pirámides etnolingüísticas. Dentro de este criterio también podríamos incluir el de la temporalidad de los corpus, que se divide en sincrónicos y diacrónicos según si las muestras son de una época actual o proceden de diferentes etapas temporales. También existen los *corpus periódicos o cronológicos*, que recogen textos de unos años determinados o de

28. Un ejemplo sería el Corpus Textual Informatitzat de la Llengua Catalana (CTILC), <http://ctilc.iec.cat>

29. Por lo que se refiere al español, tenemos el Corpus Oral de Lenguaje Adolescente (COLA), el Corpus de Conversación Coloquial del Grupo Val.Es.Co. y el Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC). También hay que destacar el proyecto PRESEEA para la creación de un corpus representativo de las variedades geográficas y sociales del español.

30. Tanto el CORPES como el CREA pertenecen a este tipo de corpus, ya que el 90% de sus textos son escritos y el 10% restante, orales.

31. Por ejemplo, el CORPES y el CREA.

32. En español, por ejemplo, tenemos el proyecto ESLEE, con vocabulario de migraciones, de nueva economía, de prospectiva tecnológica y de gestión del patrimonio cultural. También hay que mencionar el GRIAL, un corpus trilingüe de registro técnico informático y el IATE, una base de datos multilingüe de la Unión Europea.

unas épocas concretas con el objetivo de estudiar la lengua producida durante ese período<sup>33</sup>. Como ya se comentó en 1.2 al exponer los objetivos principales de este trabajo de investigación, se ha decidido acotar la búsqueda del año 2001 al año 2015, por lo que nuestro estudio tiene un enfoque principalmente sincrónico.

Por lo que se refiere al porcentaje y a la distribución de los textos incluidos, tenemos los *corpus equilibrados*, que presentan la misma proporción de los diferentes tipos de texto, los *corpus piramidales*, que contienen textos distribuidos por niveles<sup>34</sup>, los *corpus cerrados*, que no se amplían ni se modifican una vez terminados y, finalmente, los *corpus monitores o abiertos*; es decir, corpus dinámicos que están en constante crecimiento y que van actualizando sus textos para mantenerse al paso de la lengua<sup>35</sup>. En cuanto a los idiomas, tenemos los *corpus monolingües*, compuestos por una sola lengua y los *corpus bilingües o multilingües*, que constan de dos o más lenguas<sup>36</sup>. Dentro de este segundo grupo, se puede distinguir entre *corpus comparables*, que presentan un conjunto de textos de contenido similar en más de una lengua o variedad lingüística y *corpus paralelos*, con muestras de textos en más de una lengua y con su respectiva traducción. Estos últimos resultan de gran utilidad para el estudio de traducciones y en entornos bilingües o multilingües, como podría ser la enseñanza de idiomas.

También cabe mencionar los *corpus de aprendices*, que recopilan textos escritos por estudiantes no nativos de L2 con diferentes grados de dominio lingüístico<sup>37</sup>. Es indudable que algunos corpus de aprendices están suponiendo unos avances considerables en la metodología de enseñanza, como es el caso del CORINÉI<sup>38</sup>, un corpus abierto cuyo objetivo principal es conseguir que los estudiantes interacciones con nativos en condiciones de autenticidad y puedan así mejorar considerablemente sus habilidades comunicativas de acuerdo con el uso real de la lengua.

Finalmente, ateniéndonos al último criterio, tenemos los *corpus simples o brutos*, que no han sido lematizados; los *corpus lematizados*, donde cada forma se ha relacionado con su lema (lo que facilita enormemente la búsqueda, especialmente en las lenguas con mucha flexión); los *corpus analizados morfológicamente*, en los que se indica la categoría gramatical de todas las palabras; los *corpus etiquetados semánticamente*, donde se especifican los marcos semánticos y las relaciones que mantienen las palabras y, por último, los *corpus analizados sintácticamente*,

33. Un claro ejemplo sería el corpus SOL, <http://spraakbanken.gu.se/konk/rom2/> o el DICCA-XV, un proyecto de la Universitat de Barcelona, <http://ghcl.ub.edu/diccxv>

34. Un nivel consta de pocas variedades temáticas pero con muchos textos para cada una; un segundo nivel, de textos más variados temáticamente, pero con menos cantidad de cada uno, etc.

35. Como por ejemplo el CORPES y el CREA.

36. Por ejemplo, el C-ORAL ROM, que incluye muestras en francés, italiano, español y portugués. <http://www.llf.uam.es/ESP/Coralrom.html>

37. Desde el nivel A1 hasta el C1 según el Marco Común Europeo de Referencia (MCER).

38. Este corpus forma parte del proyecto TELETÁNDEM, desarrollado en el ámbito del Espacio Europeo de Educación Superior (EEES), y cuyo objetivo es mejorar la competencia conversacional de los estudiantes de E/LE e I/LE, respetando las reglas de la enseñanza/aprendizaje colaborativos. El proyecto inició en 2009 entre las universidades de Alicante (UA) y Suor Orsola Benincasa de Nápoles (SOB). Sucesivamente, en 2011, se incorporó la universidad de Salerno (UNISA). <http://dti.ua.es/es/teletandem-corinei>

en los que cada palabra lleva anotada su función sintáctica (por ejemplo, S = *sujeto*)<sup>39</sup>.

Uno de los inconvenientes que se detectó en algunos corpus de referencia como el CREA fue justamente su falta de lematización, ya que nuestra búsqueda (fase 3 del proceso de creación) se centra principalmente en el lema y no en la forma, razón por la que nos decantamos primeramente por los corpus lematizados y decidimos centrar nuestras búsquedas en el CORPES. Por lo que se refiere al italiano, en cambio, la cantidad de corpus a disposición es menor y nos vimos obligados a escoger el CORIS/CODIS pese a no estar lematizado.

### 3.1.3. LA COMPETENCIA COMUNICATIVA EN LOS CORPUS

La competencia comunicativa es, básicamente, la capacidad de una persona para comportarse de manera eficaz y adecuada en una determinada comunidad de hablantes; se ve reflejada en el uso real de la lengua, y esa ha sido sin duda una de las razones principales por las que en esta tesis se ha decidido incluir los corpus en la triangulación metodológica. Tras realizar muchas búsquedas en corpus y bases de datos, aventuramos la idea de que la teoría Sapir-Whorf de la relatividad lingüística y el modelo de Hymes (2.2.2 y 2.2.3) no son incompatibles, sino que se complementan para ofrecernos una visión más global y completa de la etnolingüística. Se considera que la **frecuencia** como único baremo de categorización resultaba algo deficiente, y el modelo SPEAKING de Hymes (2.2.3) nos sugirió la idea de la **rentabilidad comunicativa**; es decir, hasta qué punto una palabra o expresión, pese a ser utilizada con frecuencia y a aparecer en muchos corpus, es rentable para un hablante en determinados contextos. Siguiendo este concepto, vale la pena recordar los cuatro postulados que Hymes considera esenciales en la competencia comunicativa y que se publicaron por primera vez en 1972 en su artículo “On communitive competence” (269-293):

1. ¿El enunciado está *bien formado*?
2. ¿Se trata de una construcción *posible* en esa lengua?
3. ¿Está formulado de un modo *adecuado al contexto* en el que se emite?
4. ¿Es una construcción que *realmente ocurre* en esa lengua y que se da en la realidad?

Así pues, la competencia comunicativa se relaciona con saber cuándo hablar, con quién, dónde y en qué forma; por eso creemos conveniente jerarquizar las expresiones extrapoladas de los corpus no solo por su frecuencia, sino también por su rentabilidad, como se detallará en 4.2.4 y en 4.3.4. Por lo tanto, nos interesa saber cuándo una determinada expresión se da en la lengua, cuántas veces y en qué contextos, ya que, como apunta Hymes en respuesta al cuarto postulado, “puede que algo resulte posible, factible y apropiado, pero que no llegue a ocurrir en la realidad” (“On Communitive competence”, 286-289).

39. Estos son cada vez más frecuentes y queremos destacar la Base de Datos Sintácticos del Español Actual (BDS) y el An-Cora, un proyecto de la Universitat de Barcelona para el español y el catalán.

### 3.1.4. EL ENFOQUE INDUCTIVO Y LA MOTIVACIÓN

En las últimas décadas ha ido ganando cada vez más protagonismo el enfoque comunicativo para la enseñanza de una L2, dejando a un lado los métodos tradicionales que se centran principalmente en los aspectos gramaticales de la lengua. Como se extrapola de la tesis doctoral de la Dra. Mónica Caballero Benavente, no siempre las estructuras gramaticales que estudiamos en el aula, por ejemplo relacionadas con las transacciones y con las peticiones, aparecen luego en los contextos comunicativos cotidianos (*La variabilidad lingüística nativa y no nativa en escenarios comunicativos. La función de transacción en una situación cotidiana: corpus y descripción para el español*, 181-183). Se enseñan diferentes maneras para pedir algo en español utilizando las varias formas del verbo querer, pero resulta que en la realidad muchos españoles acuden a un bar y simplemente dicen *un café, por favor*, omitiendo el verbo. Esto no quiere decir que la gramática desempeñe un papel secundario; más bien, significa que hay que entender y enseñar la gramática como un medio para poder comunicarse, y no como algo automático que hay que memorizar de manera mecánica. Según la teoría de Mario Gómez del Estal y Javier Zanón, la enseñanza explícita y mecánica de la gramática solo sirve para la resolución de ejercicios a corto plazo y no garantiza un uso correcto en los contextos comunicativos (“Tareas formales para la enseñanza de la gramática en la clase de español”, 73-99). Nosotros creemos en la importancia de enseñar la gramática a partir de muestras reales de la lengua para que el alumno pueda reflexionar y sacar conclusiones de carácter inductivo. Esta metodología, centrada principalmente en la forma y en el significado, contribuye de manera muy efectiva al desarrollo de la interlengua y favorece el proceso natural de adquisición, como se puede apreciar en el interesante ensayo de Catherine Doughty y Jessica Williams, en el que se hace particular hincapié en la importancia de la forma y del significado por encima de toda regla gramatical (*Focus on Form in Classroom Second Language Acquisition*, cap. 9). Todas estas teorías avalan el conocido enfoque o método inductivo en la enseñanza de una L2, que consiste en sacar, de un número acotado de casos observados, una conclusión que se considere válida para todos los casos semejantes. Dicho de otra manera, con el método inductivo creamos nuestras hipótesis a partir de la observación de datos reales (Data-driven), y hoy en día esas muestras están a nuestro alcance gracias a los corpus, con los que podemos adentrarnos en un territorio lingüístico muy extenso y explorar diferentes caminos a partir de las opciones de búsqueda.

En líneas generales, a la hora de estudiar una L2, la mayoría elaboramos hipótesis sobre el funcionamiento de determinados elementos de una lengua y nos valemos del contexto para averiguar su significado; se trata de un proceso natural en el ser humano y es evidente que estamos actuando de forma inductiva, ya que creamos nuestras propias hipótesis realizando comparaciones con nuestra lengua nativa o con otras lenguas que manejamos mejor. Pero resulta que, a veces, y pese al contexto, el significado global de la oración no resulta tan claro y necesitamos recurrir a otros medios, como por ejemplo un diccionario o un traductor automático. Sin embargo, la búsqueda en los diccionarios a veces puede resultar infructuosa y se necesita entrar en contacto con varias muestras reales de la lengua para esclarecer del todo el significado de la oración. Esas muestras nos las proporcionan, como decíamos, los corpus,

y constituyen unos *inputs* muy eficaces<sup>40</sup>. Esos *input* estimulan la curiosidad y, por ende, despiertan la motivación de los estudiantes, un elemento fundamental en todo proceso de aprendizaje. A este respecto, vale la pena mencionar, pese a no utilizar los corpus, el método IBI del Dr. Joe Barcroft para la enseñanza de vocabulario, que consiste en presentar una serie de palabras meta como *input* inicial para luego desarrollar varios aspectos de conocimiento léxico. El propio Barcroft sostiene que “la presencia de diferentes hablantes, estilos de habla y velocidades de habla produce efectos positivos en el aprendizaje de nuevas palabras” (*Input-Based Incremental Vocabulary Instruction*, 231-249) ya que favorece la participación de los estudiantes y estimula su curiosidad. Como ejemplo práctico, Barcroft propone un ejercicio para trabajar el vocabulario relacionado con la comida utilizando el método IBI con estudiantes de nivel intermedio. Escribimos en la pizarra una serie de palabras meta relacionadas con la receta del gazpacho, luego mostramos unas fotografías relacionadas con esas palabras y hacemos comentarios para que todo resulte comprensible. Finalmente, se divide la clase en grupos de 2/3 personas y cada alumno tiene que pensar en una comida típica de su país o del mundo hispano y comentarla con los integrantes de su grupo. Se trata de un ejercicio muy dinámico y sobre todo muy flexible, aunque con el *brainstorming* funciona justamente al revés.

En caso de plantear el *brainstorming* en una clase de idiomas, el docente al principio no presenta ninguna palabra meta en la pizarra y es el propio alumno quien tiene que buscarla en su cabeza. El único *input* que se proporciona al principio es el del ámbito cultural (palabras relacionadas con la religión, con el fútbol, la comida, etc.). La idea es que un *input* cultural traiga el mayor número posible de palabras meta, y que a partir de esas palabras meta se consiga recopilar, mediante los corpus, todas aquellas expresiones que aparecen en la lengua y que contienen dichas palabras, para así organizarlas y visualizarlas en una pirámide según su frecuencia y rentabilidad con el objetivo de hacernos una idea de cómo influye ese ámbito cultural en la lengua que se está estudiando.

Tras realizar el ejercicio del *brainstorming* con los alumnos, se pudo comprobar que el estudiante tiende a preguntarse para qué sirve recoger todas esas palabras, y su curiosidad aumenta al introducir el lema en los motores de búsqueda de los corpus y al encontrarse con todas las ocurrencias. El enfoque que proporciona este ejercicio es claramente inductivo y pensamos que puede ofrecer al estudiante unas herramientas bastante efectivas que lo pueden hacer reflexionar sobre el funcionamiento de la lengua y sobre el peso que ejerce la cultura en ella. A raíz de esto, consideramos que las pirámides etnolingüísticas pueden ser muy útiles para ofrecer nuevos *inputs* al alumno y facilitar su inmersión lingüística en vista de un mejor aprendizaje.

40. El Diccionario de términos clave de ELE define el *input* como “muestras de lengua meta, orales o escritas, que el aprendiz encuentra durante su proceso de aprendizaje y a partir de las cuales puede realizar ese proceso”. [http://cvc.cervantes.es/Ensenanza/biblioteca\\_ele/diccio\\_ele/diccionario/aducto.htm](http://cvc.cervantes.es/Ensenanza/biblioteca_ele/diccio_ele/diccionario/aducto.htm)

## 3.2. PRINCIPALES CORPUS DEL ESPAÑOL QUE SE VAN A UTILIZAR EN ESTA TESIS

Como se ha comentado en los apartados precedentes, para que la realización de las pirámides etnolingüísticas fuera lo más precisa posible, se ha decidido realizar las búsquedas casi exclusivamente utilizando corpus de referencia lematizados, lo cual no impide que otros corpus de referencia, como por ejemplo el CREA, se deban descartar o no sean válidos para elaborar una pirámide etnolingüística, pero es indudable que el hecho de que un corpus esté lematizado nos ha facilitado mucho la tarea.

Por lo que se refiere al español, el corpus que se va a utilizar principalmente es el CORPES, de la Real Academia Española. Pese a estar todavía en versión beta, presenta una interfaz muy buena con la que el internauta se familiariza enseguida. Después de una larga reflexión, hemos decidido no apoyarnos en el Corpus del Español pese a su gran representatividad, y esto se debe principalmente a un factor didáctico, ya que para consultarlo hay que registrarse en la web, algo que ralentizaría demasiado una posible actividad didáctica en el aula; además, se requieren unos conocimientos lingüísticos y sintácticos avanzados para aprovechar al máximo la consulta, mientras que en el CORPES resulta todo más sencillo e intuitivo. Respecto a los corpus orales, se ha tenido en cuenta el corpus Val.Es.Co. para algunas expresiones del español coloquial, y se ha complementado con búsquedas puntuales en la red para detectar posibles apariciones en redes sociales y blogs. A continuación, se analizarán brevemente estos dos corpus para ver cómo nos presentan su interfaz y cómo podemos realizar las búsquedas, lo que representa el segundo paso de la triangulación metodológica y servirá de trampolín para acercarse a los objetivos A y B.

### 3.2.1. EL CORPES

El CORPES (Corpus del Español del Siglo XXI), es considerado como uno de los proyectos más ambiciosos de la Real Academia Española y actualmente está disponible en internet en versión beta<sup>41</sup>. Se puede acceder a través de la misma página web de la RAE y es muy fácil de utilizar. Se trata, al igual que el CREA, de un corpus de referencia que incluye una gran variedad de textos, con muestras de todos los países del mundo hispanico, ateniéndose a una distribución del 70% para textos hispanoamericanos y un 30% para textos españoles. Es posible consultarlo desde diciembre de 2013 y, como se explica en la propia página:

“El objetivo final del proyecto es llegar, para el año 2018, a 400 millones de formas y palabras de la lengua común de casi 500 millones de hablantes. En noviembre de 2015 se publicó una nueva versión de consulta, llamada 0,82, que cuenta con 222.080 documentos que alcanzan los 215 millones de formas; es decir, poco más de la mitad de lo estipulado para 2018. A diferencia del CREA, es un corpus lematizado, y eso ayuda a afinar mucho más la búsqueda, sobre todo si se busca por lema

41. <http://www.rae.es/recursos/banco-de-datos/corpes-xxi>

y no por forma. Se concibe como un corpus semiabierto y la mayoría de los textos que lo integran han sido seleccionados en base al siguiente porcentaje: el 90% procedentes de la lengua escrita y el 10% de la lengua oral. Los textos escritos proceden principalmente de libros (40%) y de publicaciones periódicas (40%), mientras que el resto procede de internet (7,5%) y de miscelánea (2,5%)”.

Para este trabajo de investigación, nos vamos a valer de todos los tipos de textos, tanto de la lengua oral como escrita, pero se acotará la búsqueda a ese 30% que procede de España, como se ha señalado en 1.2. Además, se tendrá en cuenta un espacio cronológico de quince años, aproximadamente desde 2001 hasta 2015, y en este sentido la mayoría de los textos que componen el CORPES están delimitados precisamente en esa franja temporal, lo cual nos ha convencido todavía más del uso de este corpus como herramienta principal de búsqueda.

Al entrar en la página web de la Real Academia Española y al pichar en el enlace del CORPES, aparece la siguiente interfaz:



Imagen 3. Interfaz del CORPES (15 de mayo de 2016)

Como se puede apreciar, resulta muy intuitivo, y esa es una ventaja a tener muy en cuenta, sobre todo si decidimos utilizarlo en el aula con estudiantes de español como L2. Arriba nos aparecen las diferentes opciones de búsqueda, y de izquierda a derecha podemos escoger entre LEMA y FORMA, la CLASE DE PALABRA (adjetivo, sustantivo, adverbio, etc.), la GRAFÍA ORIGINAL (es una casilla para marcar), el SUBCORPUS y la PROXIMIDAD. Nuestro interés se centrará en la casilla del LEMA y la de SUBCORPUS. En la primera introduciremos el lema de la palabra que se quiere buscar, por ejemplo *toro*, mientras que al desplegar la pestaña del subcorpus se tendrá la opción de acotar la búsqueda solamente a España. Respecto al tiempo cronológico, no es necesario aplicar ninguna marca, pues la casi totalidad de los textos va desde 2001 a 2015, lo que agiliza todavía más la búsqueda, como se ve en el siguiente ejemplo:

The screenshot shows a search interface for the word 'toro'. The search criteria are: España, Bloque (Todos), Ficción (No ficción), Soporte (Internet), Tema (Actualidad, ocio y vida cotidiana), and Tipología (Académico). The results are sorted by year ascending. The concordance table shows various contexts where 'toro' is used, such as 'alegría española', 'ingenuos ingleses', 'Quiero medicina', 'de entenderla', 'Quiero medicinal', 'todas las ciudades', 'Rusias, pensó Tomás', 'paseemos juntos', 'papel de fumar', 'entre un killer y un killer', 'a un buscador', 'bien de mulettillas', 'Abajo, en la explanada', 'cortinón, degradado', 'cara a cara a Alberto', and 'había conocido a Jim Morrison'.

Clasificación, país)	CONCORDANCIA	Ordenar por:	Año asc
11 Esp	alegría española": pone paella y sangría en el menú de los lunes, televisa corridas de toros		
11 Esp	ingenuos ingleses, ojalá fuese tan fácil la fórmula de la felicidad! Paella, sangría, toros y flamenco.		
11 Esp	-Quiero medicina. Para los toros		
11 Esp	-Medicina para los toros		
11 Esp	de entenderla, ja, pero qué tonto soy, Ehhh... veamos, ¿qué problema tienen los toros ?		
11 Esp	-Los toros		
11 Esp	Quiero medicinal Tengo al chaparro fuera y sin atar. Medicina. Líquida. Para los toros		
1 Esp	todas las ciudades, aunque era mejor antes, cuando venía de soltero, venía a los toros		
11 Esp	Rusias, pensó Tomás, esa Rusia ya no existe, y tampoco digas que venías sólo a los toros		
11 Esp	paseemos juntos, mi madre nos sigue detrás, y él me coge del brazo y me habla de toros		
11 Esp	papel de fumar, y luego cigarrillos liados, y puros para las fiestas y para los toros		
11 Esp	entre un killer y un killer". Me perdí en un jardín en el que dejé que penetraran los toros		
11 Esp	a un buscador. Pero tampoco creo que nadie hubiera supuesto que el mundo de los toros		
1 Esp	bien de mulettillas, de falacias y trucos de andar por casa, con tal de esquivar el toro		
11 Esp	Abajo, en la explanada de la carretera, como toros		
11 Esp	cortinón, degradado a repugnantes tonos sepia, Júpiter rapta a Europa en forma de toro		
11 Esp	cara a cara a Alberto, al igual que ha mirado cara a cara a su "hijo muerto", agarrar el toro		
11 Esp	había conocido a Jim Morrison, y juzgó que había llegado el momento de agarrar el toro		
11 Esp	-Y a mi Manolo no le gusta que en los toros		
11 Esp	hasta que todo el mundo se empapó bien del arte de Pepe Hilio. Mi padre sabía mucho de toros		

Imagen 4. Consulta de la palabra *toro* en el CORPES (15 de mayo de 2016)

En definitiva, confiamos en que el CORPES pueda proporcionar muestras representativas de la lengua y que, como herramienta fundamental de la metodología mixta que rige este proyecto (segundo paso de la triangulación), nos permita alcanzar el primer objetivo de esta tesis de investigación (A).

### 3.2.2. EL CORPUS VALES.CO.

El proyecto Val.Es.Co. (Valencia, Español Coloquial), fue impulsado en 1990 en el Departamento de Filología Hispánica de la Universidad de Valencia<sup>42</sup>. El grupo de investigación, dirigido por el Dr. Antonio Briz Gómez, centró el foco de su estudio en el español coloquial en sus diferentes niveles de análisis, creando un corpus oral obtenido principalmente de la conversación espontánea, lo cual constituye una muestra excelente para ver algunas interacciones comunicativas en español. El propio Briz defiende la idea de que a partir de un corpus oral se puede analizar el uso informal y no planificado del lenguaje en un marco de interacción cotidiana, con fines interpersonales (*El español coloquial en la conversación: esbozo de pragmalingüística*, cap.1).

En la web es posible consultar el corpus Val.Es.Co. 2.0, que incluye 46 conversaciones en español con un total de 12.976 intervenciones, 29.205 grupos de entonación y 120.246 palabras<sup>43</sup>. Es relativamente sencillo de utilizar, y al pinchar en el apartado Corpus Val.Es.Co. 2.0 y seleccionar *consultar el corpus*, se abrirá una pestaña con cuatro opciones (ver conversaciones, consultar intervenciones, consultar grupos entonativos y búsqueda morfológica), como se ve en la siguiente imagen:

42. <http://www.valesco.es>

43. Datos recogidos en la propia página web <http://www.valesco.es/?q=es/node/13>



Imagen 5. Interfaz del corpus Val.Es.Co. (15 de mayo de 2016)

Las primeras dos opciones son las más interesantes, ya que permiten ver las conversaciones y consultar las intervenciones. La mayoría de ellas abarcan un período que va desde 1989 hasta 1996, pero hay algunas de 2007, de 2011 y de 2012, y son las que se van a tener en cuenta para ser fieles al aspecto cronológico de este proyecto.

The screenshot shows the 'Conversaciones' section of the Val.Es.Co. website. It features a search bar with the text 'Buscar' and two buttons: 'Mostrar todos' and 'Búsqueda avanzada'. Below the search bar is a table with the following columns: 'Id', 'Número de hablantes', 'Año', 'Duración (minutos)', and 'Tema'. The table contains 10 rows of data, each with a unique ID, number of speakers, year, duration, and topic. To the right of each row is a button labeled '...Consultar' followed by a number in parentheses, indicating the number of records for that entry.

Id	Número de hablantes	Año	Duración (minutos)	Tema	...Consultar
0001	2	1989	31	trabajo, ocio, ideología	(553)
0002	9	1989	46	el contestador automático, el sexo, la recuperación escolar, el fútbol, la mili	(297)
0003	5	1994	10	noticias, fútbol, familia, amistades	(202)
0004	5	1994	3	noticias, fútbol, familia, amistades	(75)
0005	5	1994	30	comida familiar, críticas juventud, zapatillas	(240)
0006	2	1994	10	relaciones familiares, regalo de cumpleaños	(227)
0007	7	1994	5	becas, fin de semana, universidad privada	(108)
0008	4	1994	2	universidad privada, relatos eróticos	(34)
0009	5	1995	12	situación familiar, temas domésticos	(380)
0010	4	1994	10	ordenadores en la facultad, anécdotas sobre un tercero	(90)

Imagen 6. Conversaciones del corpus Val.Es.Co. (15 de mayo de 2016)

También es posible realizar búsquedas por palabra en el apartado de intervenciones, lo que puede agilizar la consulta si no se quiere leer la conversación entera en busca de una determinada expresión. Por ejemplo, si se introduce la palabra *dios*, este es el resultado:

The screenshot shows a web interface for searching through a corpus. At the top, there is a navigation bar with links: INICIO, SOBRE VALESCO, CORPUS VAL.ES.CO 2.0, ENLACES, and CONTACTO. Below this is a header for 'Intervenciones'. A search bar contains the word 'dios'. Below the search bar are buttons for 'Borrar Búsqueda', 'Búsqueda avanzada', and 'Ocultar resaltado'. A pagination bar shows 'Página 1 de 1' and 'Registros 1 a 19 de 19'. The main content is a table with the following data:

Orden de la intervención	Texto en la intervención	Id conversacion	
24	M: se necesita mucho DINERO y mucha CONSTANCIA y tú con los estudios! // si vas a guitarra yaa no puedes hacerte ni los deberes	0006	...contexto
25	L: ¡ya ves que estudios!	0006	...contexto
34	M: oh chorrada! unos estudios son chorrada! pues no sé en que te vas a trabajar! si no tienes estudios! // a fregar suelos por ahí!	0006	...contexto
192	D: qué suerte! ¡ madre de dios! // yo no me lo creía cuando me lo dijisteis que que le habian denunciado por darle una patada a un perro que estaba en [(( ))]	0010	...contexto
113	A: ¡dios!	0011	...contexto
165	A: [(( ))] a mí me ha dicho que este any no eixiu? que esteu molt estudioses? Y yo sii/ molt estudioses, estem// me ha dicho eso	0012	...contexto
98	A: y- y nada / tia/ y me acuerdo un día en mis tiempos que era asi pijilla/ nos vamos ahí a crisis ↓ y yo me encané con una chaqueta que habia Chipi↓ de estas asi entalladitas vaqueras! que estaba un poco desgastada/ pero vamos que yo que sé/ y valia trece mil pelas ( la chaqueta)// TRECE MIL pesetas/ trece mil pesetas eso/	0013	...contexto

Imagen 7. Consulta de la expresión *dios* en el corpus Val.Es.Co. (15 de mayo de 2016)

En definitiva, este corpus oral servirá de apoyo puntual para perfilar mejor la estructura de las pirámides etnolingüísticas y optimizar la visualización de las expresiones pluriverbales (tercer paso de la triangulación) de cara a un primer análisis contrastivo, foco del segundo objetivo (B) de esta tesis.

### 3.3. PRINCIPALES CORPUS DEL ITALIANO QUE SE VAN A UTILIZAR EN ESTA TESIS

Tras ver los corpus que se utilizarán para la realización de las pirámides etnolingüísticas en español, vamos a centrarnos ahora en el italiano, una lengua muy inferior al español por difusión geográfica y número de hablantes. El corpus principal del que nos serviremos aquí será el CORIS/CODIS<sup>44</sup> que, pese a no estar lematizado, es actualmente la base de datos más representativa del italiano. También nos apoyaremos en el LIP, el corpus de *Lessico dell'Italiano Parlato*, cuyo proyecto abarca desde mediados de los años noventa hasta 2006, por lo que colinda bastante con el período cronológico del que bebe esta investigación; además, complementa el aspecto oral de la lengua del que carece el CORIS/CODIS, de ahí que nos apoyemos en él bastante más respecto a lo que se hará con el español.

Cabe señalar que, al principio, también se barajó la posibilidad de utilizar el corpus PAISÀ (*Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati*)<sup>45</sup>, que nació en 2009 con el objetivo principal de ser una herramienta de apoyo para la enseñanza de italiano como L2<sup>46</sup>. Este corpus selecciona textos de internet que no están sujetos al *copyright* “todos los derechos reservados”, y que simplemente tienen licencias conocidas como *Creative Commons* (CC). En total, contiene 250.000 tokens y ofrece dos tipos de búsquedas, una sencilla y otra avanzada donde se puede buscar por lema, por forma e incluso combinando palabras. Además, presenta una interfaz muy intuitiva que puede ser en italiano o en inglés, un detalle que puede resultar de gran ayuda si se trabaja con alumnos anglófonos y se decide utilizarlo en el aula para complementar alguna actividad didáctica. Teniendo en cuenta estos factores, creemos que puede ser una herramienta lingüística de gran utilidad en futuros trabajos de investigación; sin embargo, para esta tesis nos decantamos al final por el CORIS/CODIS debido a su mayor representatividad y a su mayor variedad temática.

A continuación, vamos a ver cómo se nos presentan estos dos corpus seleccionados y de qué manera se pueden realizar las búsquedas.

#### 3.3.1. EL CORIS/CODIS

El CORIS/CODIS está disponible desde septiembre de 2001<sup>47</sup>. El proyecto, dirigido por la Dra. Rema Rossini Favretti en la Università di Bologna, tiene como finalidad principal la de ser un corpus de referencia lo más representativo posible de la lengua escrita. No presenta textos del ámbito oral, pero contiene alrededor de 130 millones de palabras procedentes de textos escritos relativamente actuales (desde 1998). Tanto en el CORIS (*Corpus di Italiano Scritto*) como en el CODIS (*Corpus Dinamico dell'Italiano Scritto*), la interfaz es muy sencilla

44. [http://corpora.dslo.unibo.it/coris\\_ita.html](http://corpora.dslo.unibo.it/coris_ita.html)

45. <http://www.corpusitaliano.it/it/contents/paisa.html>

46. La duración del proyecto fue exactamente de tres años, hasta junio de 2012.

47. [http://corpora.dslo.unibo.it/coris\\_ita.html](http://corpora.dslo.unibo.it/coris_ita.html)

lla y la diferencia principal estriba en que en el CODIS podemos seleccionar los diferentes canales antes de realizar la búsqueda, además de que este último se actualiza constantemente para poder monitorizar el desarrollo de la lengua italiana. En el CORIS tenemos también la posibilidad de seleccionar el número de concordancias y las colocaciones, como vemos en la siguiente imagen. Para nuestras búsquedas, siempre seleccionaremos la opción de 1000 concordancias.

emogando il corpus CORIS/CODIS tramite questa procedura, l'utente dichiara e accetta che l'interrogazione è volta unicamente a scopi di ricerca scientifica e che non ne sarà tratto alcun beneficio economico. L'accesso al corpus è concesso esclusivamente per scopi di r

Imagen 8. Interfaz del CORIS. (16 de mayo de 2016)

Subcorpus	Size (in Mw)		
STAMPA	20	10	5
NARRATIVA	13	7	3
PROSA ACCADEMICA	5	4	2
PROSA GIURIDICO-AMM.	4	3	2
MISCELLANEA	4	3	2
EPHEMERA	2	1	1

Imagen 9. Interfaz del CODIS. (16 de mayo de 2016)

El CORIS/CODIS presenta una serie de códigos que es menester conocer antes de utilizar el motor de búsqueda, como por ejemplo las comillas:

### Corpus Query Language

**1 corpora in non-annotated form**

IMS/CWB Style	Description and Examples
"word"	Single word query <b>Inside the double quotation marks you can use regular expressions</b>
"'047"	Apostrophe (it is a special character and creates problems if you insert "'")
"word"%c	Single word query (case insensitive)
"word1" "word2"	Concatenation Operator. Example: "the" "book"
"word1"   "word2"	OR Operator. Example: "the" ("beautiful"   "red") "book"
word1" [(0,AFT) "word2"	Retrieve all the occurrences of word2 in the context following the word word2. Example: "the" [(0,4) "book" the book of the beautiful red book

**1 corpora in ANNOTATED form**

IMS/CWB Style	Description and Examples
[pos="pos-tag"]	Single pos or lemma query. Examples: [pos="NN"] -> all the common nouns (see the <a href="#">PoS-tag list</a> ) [lemma="andare"] -> all the forms corresponding to the lemma "andare"
d="word1" & pos="pos-tag"]	Search a word with a specific PoS-tag. Example: [word="amo" & pos="NN"] Or search a word that DOES NOT have a specific PoS-tag. Example: [word="amo" & pos!="NN"]

For query options please refer to sections 2.5, 2.6, 2.7, 2.8 and A.1 of [IMS/CQP Tutorial](#).

Imagen 10. Marcas del CORIS/CODIS. (16 de mayo de 2016)

Así pues, para que la búsqueda de resultados, necesitaremos incluir el lema entre comillas, como se ve en la siguiente imagen:

Corpus CORIS, annotated version (2016, 140Mw)  
- Corpus query form -

**User Authentication**  
CORIS access is now free for research purposes (Please, read the footnote carefully).

**Query**  
(Query Language Help)  
"toro"

**Subcorpus** All **Section** All

**Concordance Options**  
Show  30  100  300  1000 lines.  
Sort position: Unsorted

**Collocations**  
Log Likelihood Ratio

Imagen 11. Marcas del CORIS con la palabra *toro*. (16 de mayo de 2016)

En definitiva, el CORIS/CODIS es un corpus sincrónico y de referencia que ofrece, a nuestro modo de ver, muestras muy representativas de la lengua italiana. Es gratuito y sencillo de utilizar, con lo que puede complementar muy bien actividades de carácter inductivo en el aula enfocadas a la enseñanza de italiano como L2. Si se compara con el CORPES, el corpus de referencia principal para lo que atañe a las pirámides etnolingüísticas en español (3.2.1), vemos que la diferencia principal estriba en que el CORIS/CODIS no ofrece muestras del ámbito oral. Para subsanar esta pequeña desigualdad, se aprovechará el LIP, un corpus que está centrado por completo en la lengua oral y que se analizará en el siguiente subapartado.

### 3.3.2. EL CORPUS LIP

El corpus LIP (*Lessico di frequenza dell'Italiano Parlato*) es un proyecto que empezó en 1992 y que fue dirigido por el lingüista Tullio de Mauro en colaboración con la Fondazione IBM Italia. Es una de las bases de datos más importantes para el italiano oral y contiene 469 textos recogidos en cuatro ciudades diferentes: Milán, Florencia, Roma y Nápoles, por un total de casi 57 horas de conversación. Se trata principalmente de diálogos cara a cara, por lo que el corpus ofrece muestras muy fidedignas del italiano hablado en la calle. La procedencia de los hablantes (915 hombres frente a 726 mujeres) es muy variada, siendo en su mayoría empleados, licenciados, estudiantes, jubilados y desempleados. Curiosamente, este corpus no tiene un enlace directo ni una web propia, y solo es posible acceder a él mediante una página web gratuita de la Karl-Franzers Universität de Graz llamada BADIP, disponible en inglés y en italiano<sup>48</sup>:

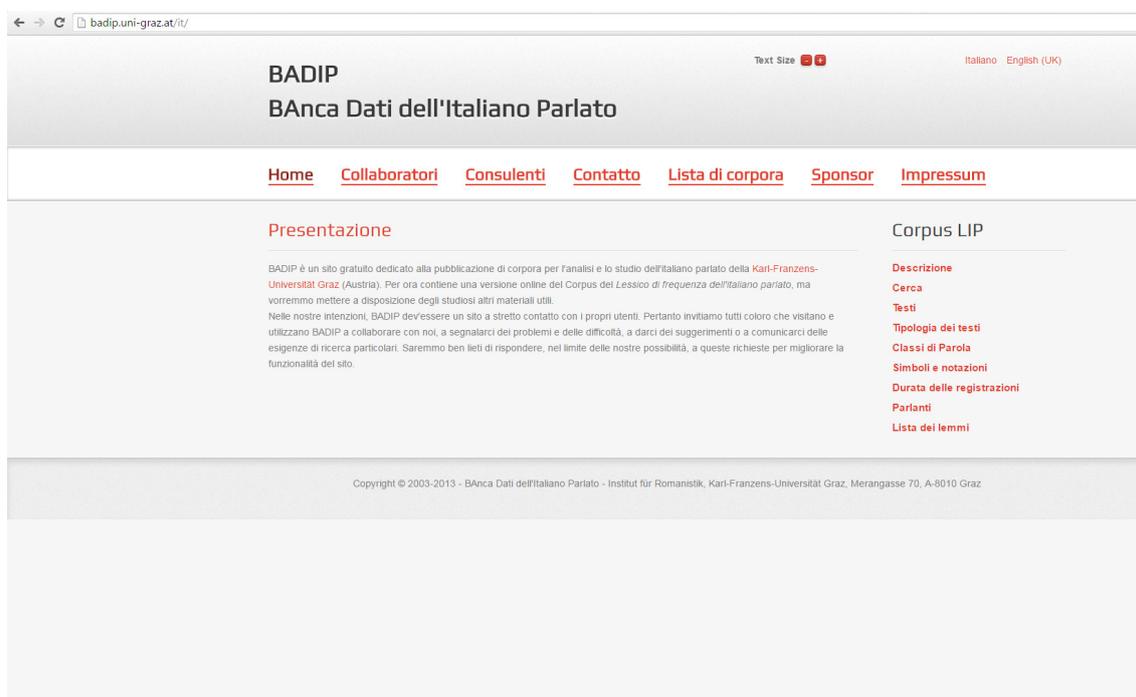


Imagen 12. Página principal de BADIP. (16 de mayo de 2016)

48. <http://badip.uni-graz.at/it>

[Home](#) [Collaboratori](#) [Consulenti](#) [Contatto](#) [Lista di corpora](#) [Sponsor](#) [Impressum](#)

Cerca tutte le sequenze che contengano (HELP)

digita la prima parola oppure [CLICK](#)  
 digita la seconda parola oppure [CLICK](#)  
 digita la terza parola oppure [CLICK](#)

e che non contengano:

digita la prima parola  
 digita la seconda parola  
 digita la terza parola

nelle città  Firenze  Milano  Napoli  Roma

**Corpus LIP**

- [Descrizione](#)
- [Cerca](#)
- [Testi](#)
- [Tipologia dei testi](#)
- [Classi di Parola](#)
- [Simboli e notazioni](#)
- [Durata delle registraz](#)
- [Parlanti](#)
- [Lista dei lemmi](#)

Imagen 13. Motor de búsqueda del corpus LIP. (16 de mayo de 2016)

occorrenze del lemma/forma trovato:146  
sul totale di parole grafiche:489178

statistiche:  
1) [dio](#)

salva i dati visualizzati:

città	tipo	con versio ne	enu cia to	parl ante	enunciato
F	A	6	21	A	si' si' eh eh gia' gia' Nino \$ e' venuto da quelle parti la' gli ho detto <a href="#">dio</a> fa una telefonata quando giri oh per lo meno una sera al ristorante tutti insieme
F	A	10	132	B	certo per carita' di <a href="#">dio</a> non ci penso per niente anzi
F	A	12	404	A	* poi vediamo # tutto quello facciamo anche un programma di cosa portiamo la' cioe' eh come agenzia come noleggi e come azienda in genere no * quindi vediamo un po' saranno tutti i soliti depliant ma si per esempio si potrebbero portare le nostre offerte l' agenzia qualche offerta che prevede i i il coso i voli da da da da si' ora piglio la pratica ora ci vuol la mano di <a href="#">dio</a> per fare il \$ ma per esempio un viaggio che prevede il volo aereo da Firenze tanto per dire capito * eh oppure altri eh altre cose proprio che tengano conto di questa di questo aeroporto insomma ecco
F	B	12	24	B	che <a href="#">Dio me la mandi buona mh</a>
F	B	14	544	A	ecco per esempio sicche' invece questa e' una cosa che proprio tutti i giorni quanti <a href="#">Dio</a> ne mette in terra e' si consuma

Imagen 14. Consulta de la palabra *dio* en el corpus LIP. (16 de mayo de 2016)

Como se aprecia en estas imágenes, realizar búsquedas en este corpus es relativamente sencillo y se pueden acotar por las ciudades mencionadas, aunque nosotros siempre marcaremos las cuatro. La única incompatibilidad que se encontró entre el LIP y el presente proyecto es de índole cronológica, pues con las pirámides etnolingüísticas se pretende reflejar la lengua de los últimos quince años, partiendo del año 2001, y las muestras del LIP son de mediados de los años noventa; es decir, hay una diferencia aproximada de entre cinco y seis años; no creemos que pueda perjudicar el resultado final, pero consideramos oportuno mencionarlo y este aspecto se retomará en el apartado 5.1, dedicado a las aportaciones y limitaciones del

presente trabajo de investigación. Como se ve en las imágenes, este corpus nos proporciona expresiones muy frecuentes y muy típicas del lenguaje oral (*che dio me la mandi buona*) que difícilmente vamos a encontrar en el lenguaje escrito, de aquí la gran importancia de este corpus para perfilar adecuadamente las pirámides etnolingüísticas, un factor determinante que permite hacer una pequeña excepción al criterio cronológico marcado desde un principio.

### 3.4. LA RED COMO CORPUS

No queríamos terminar este apartado dedicado a los corpus sin mencionar algunas fuentes complementarias que pueden representar un buen soporte para este proyecto. Es indudable que internet va a facilitar muchísimo la redacción de las pirámides etnolingüísticas, puesto que se necesitará acceder a los corpus para realizar la búsqueda (fase 2 del proceso de creación) y todos están disponibles únicamente en modalidad *online*. Sin embargo, con internet no solo tenemos acceso a los corpus, sino que podemos encontrar una cantidad de información casi ilimitada en la web para aunar el aspecto teórico con el práctico (Gatto, *The web as corpus: theory and practice*, cap. 1). Motores de búsqueda como Google, Yahoo y Bing, entre otros, permiten acceder a unos corpus gigantescos que pueden proporcionar muestras lingüísticas de todo tipo. El problema principal estriba en que resulta muy difícil clasificar toda esa información; además, a menudo se desconoce la procedencia de la muestra y no se puede estar seguro al cien por cien de si estamos frente a un canal oral o escrito. En cuanto al aspecto cronológico y geográfico, es posible acotar la búsqueda a un período concreto y, en el caso del español, seleccionar solo aquellos textos que procedan de España, por lo que el problema es menor comparado con la procedencia.

Mediante la web como corpus, se realizarán búsquedas de comprobación con algunas expresiones que puedan generar ambigüedad para cerciorarse de que su presencia es palpable en la lengua, especialmente por lo que se refiere al canal oral. Para ello, se seleccionarán únicamente las páginas en español o en italiano y se acotará la búsqueda a los últimos quince años, como se ve a continuación en el ejemplo:

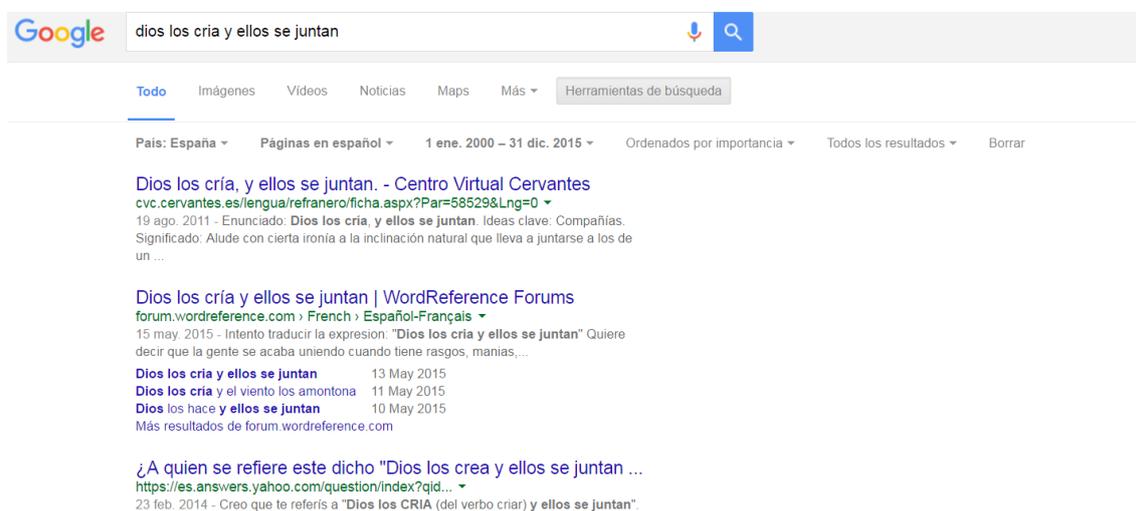


Imagen 15. Consulta de la expresión *Dios los cria y ellos se juntan* en Google (18 de mayo de 2016)

Por supuesto, otros motores de búsqueda como Yahoo o Bing pueden ser igual de eficaces, pero se ha preferido utilizar solamente el motor determinado de Google por ser el que más conocíamos y que manejamos a diario en nuestras pesquisas por internet en el período de redacción de esta tesis.

Por último, vale la pena mencionar Google Ngram Viewer<sup>49</sup>, una herramienta de Google que permite comparar las palabras que aparecen en más de 5 millones de libros publicados en Google Books desde 1500 hasta 2008 en inglés, francés, español, alemán, chino y ruso, lo que representa una considerable digitalización de buena parte de la cultura de la humanidad. Si entramos en la página principal, vemos que ofrece la posibilidad de introducir una o más palabras, seleccionar el período cronológico, el idioma y, por último, el factor de precisión (*smoothing*), entre 0 y 50, con 3 como valor por defecto.

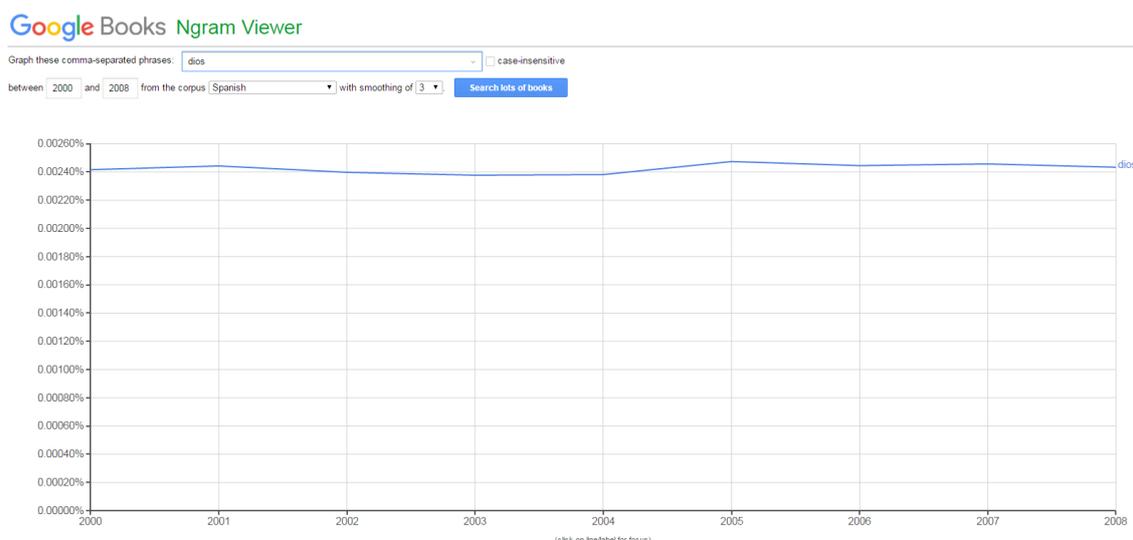


Imagen 16. Consulta de la palabra *dios* en Google Ngram Viewer. (18 de mayo de 2016)

En definitiva, es una herramienta sencilla de utilizar y puede ilustrar acerca de la mayor o menor aparición de una expresión o palabra en el ámbito escrito, por lo que funcionaría como un indicador de tendencia de la lengua escrita. En esta tesis solo se utilizará la web como corpus para complementar los resultados de búsqueda, pero pensamos que Google Ngram Viewer puede ser una herramienta interesante de cara a futuras investigaciones que pretendan estudiar los cambios de frecuencia a lo largo de los años.

49. <https://books.google.com/ngrams>

### 3.5. RESUMEN DE LA PARTE I

Hemos llegado al final de esta primera parte de la tesis, en la que, por un lado, se han presentado los tres objetivos principales (A, B y C) y el marco teórico alrededor del cual gira el presente trabajo (2.1 y 2.2) y, por otro, se han analizado los corpus que se utilizarán para crear las pirámides etnolingüísticas, aclarando las razones que nos han hecho decantar por unos en lugar de otros (3.2 y 3.3). Esta *Parte I* también ha servido para presentar y justificar la metodología mixta que sigue la idea de triangulación propuesta por Zoltan Dornyei (*Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*, cap. 2). El punto de partida es el *brainstorming* (fase 1 del proceso de creación), una técnica de creatividad que facilita información cuantitativa relacionada con el subconsciente y, por lo tanto, con la cosmovisión de los participantes. Tras realizar una criba del primer listado (fase 2 del proceso de creación) de acuerdo con unos criterios que se verán en 4.2.2 y en 4.3.2, se pasará a la búsqueda en los corpus seleccionados como segundo paso de la triangulación (fase 3 del proceso de creación)<sup>50</sup>. La idea, como se verá en la *Parte II*, es encontrar el mayor número de expresiones pluriverbales relacionadas con el ámbito cultural en cuestión (objetivo A) para luego ordenarlas en unas pirámides en base al criterio de frecuencia y de rentabilidad comunicativa, como se expondrá en 4.2.4 y 4.3.4, con el fin de visualizar la información cuantitativa y cualitativa (tercer paso y cierre de la triangulación) de cara a un análisis contrastivo entre el español y el italiano (objetivo B).

Finalmente, en esta *Parte I* se ha dedicado el apartado 2.3 a la neurociencia y a la función de las neuronas espejo en el cerebro humano. Estas neuronas presentan un sistema jerárquico en el que mandan las superneuronas espejo, quienes establecen el primer filtro de la información que percibimos; además, son responsables de la empatía y de la imitación, por eso creemos que justifican la aparición de determinadas expresiones y su aceptación en una comunidad lingüística. En 4.4 se retomarán estos conceptos para demostrar la influencia de la cultura en la lengua a raíz de la visualización de las pirámides y el análisis contrastivo, lo que representa el tercer y último objetivo de este proyecto de investigación (C).

50. Es probable que llame la atención el hecho de que no se vayan a tener en cuenta los corpus paralelos como el C-ORAL-ROM o de aprendices, ni que se les haya dedicado un apartado explicativo, pero hemos optado por no adentrarnos en otros tipos de corpus porque se quiere ver de qué manera influye la cultura en español y en italiano y por qué algunas expresiones son más frecuentes y rentables en una lengua y menos en la otra. En este sentido, y como se detallará en la segunda parte, el proceso de creación de cada pirámide es independiente y se realiza principalmente con corpus monolingües de referencia; una vez terminado dicho proceso, se comparan y se analizan los posibles equivalentes, pero se trata de un paso secundario y para eso no es indispensable recurrir a un corpus paralelo o bilingüe.





# **PARTE II**

## **4. LAS PIRÁMIDES ETNO- LINGÜÍSTICAS**



# **4. LAS PIRÁMIDES ETNOLINGÜÍS- TICAS**

## 4.1. EL CONCEPTO DE "PIRÁMIDE ETNOLINGÜÍSTICA"

El concepto de "pirámide" surgió a partir de una doble necesidad: por un lado, jerarquizar toda la información cuantitativa y cualitativa; por otro, ofrecer una imagen visual para que dicha información resulte más clara y sea más sencillo realizar *a posteriori* un análisis contrastivo entre el español y el italiano. A lo largo de esta segunda parte, se explicará todo el proceso de creación de las pirámides etnolingüísticas para intentar corroborar las hipótesis de partida y tratar de alcanzar los tres objetivos principales de esta investigación (A, B y C).

Confiamos en que las pirámides etnolingüísticas permitan representar visualmente la cosmovisión de los hablantes de una lengua en un determinado período, y que sirvan de elemento contrastivo para comparar posibles expresiones afines y equivalencias. Además, pese a la jerarquización de las expresiones, las "pirámides etnolingüísticas" son una propuesta muy flexible, ya que pueden variar constantemente en función de los parámetros cronológicos, diatópicos e incluso diastráticos que se les apliquen. Siguiendo las directrices que se expondrán a continuación, veremos que es posible realizar cualquier tipo de pirámide etnolingüística en la lengua escogida, y luego hacer lo mismo con otra lengua en caso de buscar una comparación. La única condición fundamental es que se pueda acceder al menos a un corpus de referencia de la lengua en cuestión para obtener un resultado mucho más preciso y riguroso.

Dependiendo del tipo de pirámide que se quiera crear, el proceso puede ser más o menos laborioso, aunque se ajusta a cuatro fases bien delimitadas que ejemplificaremos a lo largo de este apartado: el **brainstorming**<sup>51</sup> (4.2.1 y 4.3.1), la **obtención del listado de trabajo** (4.2.2 y 4.3.2), la **búsqueda en los corpus** (4.2.3 y 4.3.3) y la **ordenación dentro de la pirámide** (4.2.4 y 4.3.4). Todo el proceso de creación quedaría reflejado de la siguiente manera:

### PALABRAS

1. **Brainstorming (fase 1) - Resultado: listado**
2. **Criba de los resultados (fase 2) – Resultado: listado definitivo**

### EXPRESIONES

3. **Búsqueda en el corpus de ocurrencias y expresiones (fase 3) – Resultado: expresiones y frecuencia.**
4. **Selección de las expresiones y visualización (fase 4) – Resultado: pirámide etnolingüística.**

A continuación, se explicarán las cuatro fases de creación de las pirámides a partir de seis ejemplos: tres pirámides en español y tres en italiano para los ámbitos de la religión, el fútbol y los toros. Finalmente, en 4.4 se cotejará todo el material visual obtenido en español y en ita-

51. Este ejercicio se puede realizar en clase con estudiantes de español o de italiano como L2, a ser posible de nivel intermedio-alto, pero en esta tesis, más allá de la aplicación didáctica, se quieren proponer unas pirámides que ofrezcan una visualización lo más objetiva posible de la influencia de la cultura en la lengua. Por esta misma razón, se realizará el *brainstorming* con personas nativas que trabajen a diario con la lengua o la literatura y tengan un vocabulario medio superior a las 3.000 palabras.

liano para tratar de demostrar que la cultura ejerce una influencia diferente incluso en lenguas aparentemente afines (objetivo C), y eso tiene que ver con la cosmovisión de los hablantes, su capacidad creativa, el proceso de imitación y también con determinados acontecimientos históricos.

## 4.2. EL PROCESO DE CREACIÓN DE LAS PIRÁMIDES ETNOLINGÜÍSTICAS: TRES EJEMPLOS PARA EL ESPAÑOL

### 4.2.1. FASE 1: EL *BRAINSTORMING*

Como se describía en el subapartado 2.4.4, dedicado a los campos etnolingüísticos y al *brainstorming*, esta técnica sirve principalmente para dejar fluir libremente el pensamiento con el fin de generar el mayor número posible de ideas, y representa el primer paso de la triangulación metodológica de esta tesis. Es fundamental realizar un *brainstorming* como paso previo y puede resultar de gran utilidad incluso para sonsacar preciosa información etnolingüística sobre los participantes.

Así pues, la mañana del jueves 17 de diciembre de 2015 a las 11.00h realizamos el primer experimento en español con cuatro profesores del departamento de Filología Hispánica de la Universitat de Barcelona: el Dr. Antonio Torres Torres, la Dra. Paloma Arroyo Vega, el Dr. Juan Pablo García-Borrón y la Dra. Mar Cruz Piñol, a quienes agradezco su paciencia y disponibilidad<sup>52</sup>. Se optó por escoger a cinco participantes españoles porque en esta investigación la pirámide está limitada a la variedad diatópica de España, y es recomendable que los participantes sean personas nativas del país objeto de análisis. El *brainstorming* duró alrededor de veinte minutos y cada profesor lo realizó solo en su despacho. Vale la pena subrayar este detalle porque normalmente se realiza en grupo, pero esta vez pensamos que sería mejor que cada profesor no se sintiese de alguna manera influenciado por la presencia de un colega de departamento. Se les entregaron a los participantes tres hojas en formato DIN-A3, una para el fútbol, otra para la religión y la última para los toros. Como se trataba de dejar fluir las ideas, los participantes podían escribir libremente saltando de una hoja a otra o haciendo tema por tema; la idea, básicamente, era que cada uno se sintiera lo más a gusto posible, libre de tensiones, esquemas o directrices, y se entregase por completo al proceso creativo. Respecto a la categoría gramatical, podían poner tanto sustantivos como adjetivos, verbos y adverbios (intentando evitar las unidades pluriverbales), ya que el objetivo principal del *brainstorming* es buscar el efecto multiplicador (sin limitaciones) para luego realizar la criba correspondiente.

Evidentemente, se abren muchas opciones para realizar el *brainstorming*, pudiendo elegir el número y perfil de los participantes, el lugar y la duración de la prueba. Esa misma tarde se llevó a cabo el mismo ejercicio también con otro informante, quien no forma parte del mundo de la filología, para ver si podía obtener información adicional y complementaria a los otros cuatro *brainstorming*. El resultado final fue satisfactorio y con cinco personas se consiguieron un gran número de palabras que se enumerarán a continuación por ámbito cultural y por orden alfabético.

52. Se indican sus nombres porque contamos con su consentimiento para ello.

### 4.2.1.1. Listado obtenido tras la Fase 1 para el ámbito de la religión

## A

Aborto  
 Agua bendita  
 Allah  
 Alma  
 Altar  
 Amor  
 Ángel  
 Apóstoles  
 Arquitectura  
 Arte  
 Arrodillarse

## B

Banco  
 Basílica  
 Bautizo  
 Beata  
 Beato  
 Belén  
 Biblia  
 Blanco  
 Bóveda  
 Buddha  
 Budismo

## C

Capilla  
 Caridad  
 Cardenal  
 Catedral  
 Catequesis  
 Catolicismo  
 Celibato  
 Ceremonia  
 Cielo  
 Cirio  
 Comulgar  
 Conducta  
 Confesarse  
 Confesionario  
 Confirmación  
 Constantino

Convento  
 Conversión  
 Cordero  
 Credo  
 Creer  
 Creencia  
 Cristianismo  
 Cristo  
 Crucifijo  
 Cruz  
 Cruzadas  
 Cuadros  
 Culpa  
 Culto  
 Cura

## D

Demonio  
 Devoción  
 Diablo  
 Diez mandamientos  
 Dios  
 Divinidad  
 Dominicás

## E

Ecuménico  
 Ermita  
 Espabilar  
 Esperanza  
 Espíritu  
 Eucaristía  
 Existencia

## F

Fanatismo  
 Fe  
 Fieles  
 Flores  
 Franciscanos

## G

Ganesha  
 Gaudí  
 Guerras  
 Guerra santa

## H

Hagiografía  
 Hostia

## I

Iglesia  
 Imagen  
 Incienso  
 Infierno  
 Inquisición  
 Intransigencia  
 Intriga  
 Islam  
 Italia

## J

Jesucristo  
 Jesús  
 Joyas  
 Juan Pablo II  
 Judío

## L

Lutero  
 Luz

**M**

Macareno  
Mantill  
Matrimonio  
Meapilas  
Mezquita  
Mi abuela Nieves  
Misa  
Monaguillo  
Monja  
Monjes  
Monoteísmo  
Muerte

**N**

Nazarenos  
Nuevo Testamento

**O**

Obediencia  
Obispo  
Oro

**P**

Pabilo  
Palcos  
Papa  
Papa Francisco  
Pascua  
Patena  
Pecado  
Pecador  
Perdón  
Piedad  
Plata  
Plegaria  
Politeismo  
Primera Comuni3n  
Procesi3n  
Pulpito

**R**

Reproducci3n  
Reglas  
Rezar  
Rito  
Roma  
Rosario

**S**

Sacerdote  
Sacrificio  
Sacrist3n  
Sacristía  
Sagrada Familia  
Sagrado  
San Agustín  
Santa  
Santa María del Mar  
Santander  
Santiago  
Santidad  
Santo  
Santo Domingo  
Sentimiento  
Serm3n  
Sinagoga  
Sotana

**T**

Tiara  
Triana  
Trinidad  
Templo

**V**

Vaticano  
Velas  
Vida  
Vino  
Violencia  
Virgen

### 4.2.1.2. Listado obtenido tras la Fase 1 para el ámbito del fútbol

## A

Apuestas  
Árbitro  
Argentina  
Arquero  
Atlético de Madrid

## B

Balón  
Barça  
Bayern  
Balompié  
Banderas  
Banquillo  
Boca Junior  
Bola  
Borussia Dortmund  
Bota  
Botas  
Brasil  
Buffon

## C

Cabezazo  
Camiseta  
Campo  
Cantar  
Cánticos  
Capitán  
Celebrar  
Celta  
Césped  
Champions  
Chutar  
Chut  
Club

Colores  
Compañerismo  
Competición  
Copa  
Copa del Rey  
Córner  
Correr  
Crack  
Cruyff  
Culé

## D

Deporte  
Deportivo  
Despejar  
Dinero  
Diversión

## E

Ejercicio  
El Chiringuito  
El Molinón  
Entradas  
Entrenador  
Entrenamiento  
Entrevistas  
Equipo  
Escuadra  
Estadio  
Estrellas mediáticas  
Eurocopa  
Éxito

## F

Falta  
Fichajes  
Fiorentina  
Fisio  
Fondo  
Futbolista

## G

Gol  
Goleada  
Grada

## H

Himno  
Hincha  
Histeria colectiva

## I

Iker Casillas  
Inter  
Internet  
Italia

## J

Juego  
Jugador  
Juventus

## L

Larguero  
Lazio  
Lesión  
Lesiones  
Liga  
Liguilla  
Linier

**M**

Madridista  
Mediático  
Messi  
Mestalla  
Milán  
Mi padre  
Mourinho  
Multitud  
Mundial

**N**

Niño  
Norte

**P**

Países  
Palco  
Parar  
Partido  
Pase  
Pasión  
Patadas  
Patrocinador  
Pelé  
Pelota  
Penalti  
Piqué  
Portería  
Portero  
Prensa  
Presidente  
Prima  
Publicidad  
Público  
Puerta

**Q**

Quiniela

**R**

Representantes  
Radio  
Rajoy  
Real Madrid  
Red  
Reglas  
Reventa  
Riazor  
River Plate  
Rival  
Rivales  
Rivalidad  
Roma  
Ronaldo  
Rotación

**S**

Seguido  
Sergio Ramos  
Sevilla  
Silbato  
Streaming  
Suárez  
Sur

**T**

Tarjeta amarilla  
Tarjeta roja  
Televisión

**V**

Vestuario  
Violencia

**4.2.1.3. Listado obtenido tras la Fase 1 para el ámbito de los toros****A**

Abanico  
 Albero  
 Alternativa  
 Agua  
 Animal  
 Arena  
 Arrastrar  
 Arrastre

**B**

Banda de música  
 Banderillero  
 Banderilla  
 Banderillas  
 Barrera  
 Burladero

**C**

Caballo  
 Callejón  
 Camisa  
 Campo  
 Capote  
 Color  
 Corbata  
 Cornada  
 Cortar  
 Coso  
 Crítica  
 Crueldad  
 Cuadrilla  
 Cuerno

**D**

Dedicatoria  
 Dehesa

**E**

El Cordobés  
 El Guerra  
 Entrada  
 Espectáculo  
 Estampitas

**F**

Faena  
 Faenar  
 Feria de Sevilla  
 Fiesta  
 Finca  
 Fuerza  
 Furgoneta

**G**

Goya

**H**

Herida

**I**

Indulto

**L**

La Monumental  
 Las Ventas  
 Libertad  
 Lidia  
 Lorca  
 Luces  
 Lucha

**M**

Maestranza  
 Manoletilla  
 Mantilla  
 Mi abuelo Pepe  
 Miedo  
 Montera  
 Morlaco  
 Muerte  
 Mujeres  
 Mulillas

**O**

Orejas

**P**

Pasado  
 Paseíllo  
 Pasión  
 Pasodoble  
 Picador  
 Picasso  
 Plaza  
 Presidente  
 Público  
 Puerta grande  
 Puntilla  
 Pureza

**V**

Valentía  
 Vino  
 Violencia

**R**

Rabo  
 Rejoneador  
 Reventa  
 Ruedo

**S**

Sangre  
 Sangría  
 Sevilla  
 Sol  
 Sombra

**T**

Taleguilla  
 Tendido  
 Torear  
 Toro  
 Tradición  
 Traje de luces

#### 4.2.2. FASE 2: LA OBTENCIÓN DEL LISTADO DE TRABAJO: CRIBA, METODOLOGÍA Y CRITERIOS

El proceso creativo del *brainstorming* refleja, en buena medida, la información guardada en el subconsciente que está sujeta a nuestro personal sistema de creencias. Observando las palabras obtenidas, se puede notar el enorme caudal de información que contiene; por ejemplo, resulta curioso que en los tres campos aparezca la palabra *violencia*, que está relacionada con los temas de forma indirecta y refleja más bien una visión subjetiva. Otro aspecto interesante del *brainstorming* es la concatenación de ideas, que puede darse en forma de nombres propios o de objetos afines entre sí; es decir que, por ejemplo, cuando a un participante se le ocurre la palabra *Mourinho*, es muy probable que en su cerebro aparezcan nombres de otros entrenadores o de equipos de fútbol concretos y que se sucedan durante unos segundos.

El siguiente paso, antes de proceder a buscar las ocurrencias y las expresiones en los corpus, será realizar una criba para obtener un listado más reducido con el que se trabajará en la fase 3 (4.2.3 y 4.3.3). Dicho de otra manera, se trata de analizar consciente y racionalmente toda esa información cuantitativa fruto de la parte más irracional.

En primer lugar, salta a la vista la gran cantidad de **nombres propios** y **topónimos** que aparecen en los tres temas. Es algo con lo que se contaba, pero hay que decidir si se incluyen en el listado de palabras que se buscarán en los corpus. Retomando las ideas expuestas en 2.4.1 y 2.4.2 sobre intensión/extensión e intensidad, se puede apreciar que los nombres propios presentan una gran intensión y muy poca extensión, ya que el referente está muy acotado,

pero hay muchos casos en los que un nombre propio puede ampliar su significado, pasando a reflejar más personas<sup>53</sup>. Teniendo en cuenta este factor, se han establecido los siguientes criterios para todos los **nombres propios**, tratando de ser lo más objetivos y coherentes posible:

- 1) **Se excluirán aquellos nombres propios de personas que no estén directamente relacionados con la temática**, como *Constantino, Gaudí, mi abuela Nieves* (religión); *Rajoy* (fútbol); *Goya, Lorca, mi abuelo Pepe y Picasso* (toros).
- 2) **Se excluirán los topónimos por su falta de extensión**, como *Santander, Santiago, Triana* (religión); *Brasil, Italia, Roma y Sevilla* (fútbol y toros).
- 3) **Se excluirán los nombres de las iglesias, de los estadios y de las arenas (siempre por su falta de extensión)**, como *Sagrada Familia, Santa María del Mar, Vaticano* (religión); *el Molinón, Mestalla, Riazor* (fútbol); *Feria de Sevilla, La Monumental, Las Ventas y Maestranza* (toros).
- 4) **En el caso concreto del fútbol, se prescindirá de cualquier nombre de programa televisivo y de los nombres de clubes (falta de extensión)**, como *El Chiringuito, Atlético de Madrid, Barça, Bayern, Borussia Dortmund, Boca Junior, Celta, Deportivo, Juventus, Fiorentina, Inter, Lazio, Milán, Real Madrid, River Plate y Roma*.

Por lo tanto, nos decantaremos principalmente por los nombres propios que en algún momento hayan podido tener una ampliación de significado y se hayan asentado en la lengua. Para poder encontrar expresiones relacionadas con nombres de futbolistas, entrenadores, papas, santos y toreros, nos apoyaremos siempre en las búsquedas libres en *Google*, razón por la que se dedicó el apartado 3.4 a la red como corpus.

Para la criba de los **sustantivos, verbos y adjetivos** se ha decidido excluir todas aquellas palabras que sean demasiado genéricas y no presenten una relación directa con el ámbito cultural en cuestión. Por ello, se han establecido los siguientes criterios:

- 1) Con los **sustantivos o nombre comunes, se excluirán aquellos que no tengan una relación directa con el ámbito cultural y cuya aparición en el brainstorming se deba a factores emotivos, personales, etc.** Para la religión, se prescindirá de *aborto, amor, arquitectura, arte, banco, blanco, conducta, cuadros, esperanza, fanatismo, flores, guerras, imagen, intransigencia, intriga, joyas, mantilla, oro, palcos, plata, reproducción, reglas, sentimiento y violencia*. En cuanto al fútbol, se prescindirá de *cánticos, colores, compañerismo, competencia, competición, dinero, diversión, ejercicio, entradas, entrevistas, éxito, fisio, fondo, himno, internet, multitud, niño, norte, países,*

53. Pensemos, por ejemplo, en el nombre del famoso futbolista Messi, que aparece en el *brainstorming* del fútbol. La trascendencia de este jugador en el mundo es tan elevada, que con frecuencia en español se escuchan expresiones como *ser un Messi, hacer de Messi*, ser el Messi de algo para referirse a que alguien es muy bueno y que destaca sobre los demás. En este caso, el nombre propio amplía su significado, perdiendo intensidad y ganando extensión.

*palco, pasión, patrocinador, prensa, presidente, prima, publicidad, público, puerta, representantes, radio, reglas, rotación, streaming, sur, televisión y violencia. Por último, con los toros se eliminará del listado agua, animal, banda de música, camisa, campo, color, corbata, crítica, crueldad, dedicatoria, entrada, finca, fiesta, furgoneta, libertad, miedo, muerte, mujeres, pasado, pasión, plaza, presidente, tradición, valentía, vino y violencia.*

2) Respecto a los **verbos**, se aplicará un criterio muy similar al anterior, aunque vemos que su presencia en el *brainstorming* es mucho más limitada. Se ha decidido prescindir, por su falta de relación directa, de *cantar, celebrar* y *correr*, todos pertenecientes al ámbito del fútbol; en cambio, se optó por incluir el verbo *despejar* porque está muy vinculado al ámbito futbolístico.

3) En cuanto a los **adjetivos**, pasa algo similar a los verbos, ya que su presencia no es muy elevada. **Una vez más, se seguirá el mismo criterio aplicado anteriormente** y en este caso solo se prescindirá de *mediático* (fútbol).

4) Por último, vamos a ver cómo se tratarán los **compuestos pluriverbales** formados por un sustantivo y un adjetivo, como por ejemplo *agua bendita, diez mandamientos, guerra santa* (religión); *tarjeta amarilla, tarjeta roja* (fútbol) y *puerta grande* (toros). **Se ha decidido incluirlos por su relación directa con el ámbito cultural**, ya que a menudo el adjetivo suele aportar una mayor intensión al nombre común que acompaña y, en los casos mencionados, acota su significado, proporcionándole una mayor intensión directamente relacionada con el tema cultural. De los compuestos pluriverbales aparecidos en el *brainstorming*, solo se excluirán *estrellas mediáticas* e *histeria colectiva* (fútbol) por no ajustarse a los criterios propuestos.

Ahora, antes de ver cuál sería el resultado de los listados tras aplicar la criba (fase 2), nos gustaría hacer algunos incisos respecto a los sustantivos o nombres comunes, puesto que representan lo grueso del *brainstorming*. Por lo que se refiere a la **religión**, se ha decidido, pese a su gran extensión semántica, mantener las palabras *vida* y *muerte*, ya que es una dicotomía que está muy relacionada con toda expresión de espiritualidad y que podría proporcionar valiosas muestras. Lo mismo ocurre con el vocablo *luz*, que se incluirá por razones similares. Al ser palabras muy frecuentes en la lengua y con una gran extensión, obtendremos una gran cantidad de ocurrencias en los corpus, pero consideramos oportuno sacrificar un poco más de tiempo con el fin de no dejarnos ninguna expresión por el camino (objetivo A). Somos conscientes de que su relación es indirecta y de que la decisión de incluir estas palabras podría ser calificada de arbitraria, pero asumimos ese riesgo. En cuanto al **fútbol**, el asunto es más delicado, ya que la mayoría de los sustantivos, si bien es cierto que presentan una relación directa, se caracterizan por una gran extensión y podrían pertenecer a otros deportes, como es el caso de *apuestas, banderas, camiseta, campo, capitán* y un largo etcétera. La idea de incluirlas se debe al hecho de que constituyen gran parte del *brainstorming* del fútbol, y quitarlas sería muy limitante de cara a las búsquedas en los corpus. Está claro, por ejemplo, que un *seguidor*

puede ser también de un equipo de baloncesto o de cualquier otro deporte, pero se quiere ver si en el corpus aparecen expresiones que contengan *seguidor* y que se refieran solamente al ámbito futbolístico, de ahí la decisión de incluirla. Por último, la criba de los **toros** ha sido un poco más sencilla y nos encontramos con muchas palabras técnicas y específicas del mundo taurino que no dejaban lugar a dudas. Solo cabría mencionar la inclusión de las palabras *herida* y *lucha* que, pese a su gran extensión, se consideró que podrían ser de ayuda en la búsqueda en los corpus (fase 3).

Tras aplicar estos criterios, así es como quedarían las tres listas.

#### 4.2.2.1. Listado definitivo tras la Fase 2 para el ámbito de la religión

### A

Agua bendita  
Allah  
Alma  
Altar  
Ángel  
Apóstoles  
Arrodillarse

### B

Banco  
Basílica  
Bautizo  
Beata  
Beato  
Belén  
Biblia  
Bóveda  
Buddha  
Budismo

### C

Capilla  
Caridad  
Cardenal  
Catedral  
Catequesis  
Catolicismo  
Celibato  
Ceremonia  
Cielo  
Cirio  
Comulgar  
Confesarse  
Confesionario  
Confirmación  
Convento  
Conversion  
Cordero  
Credo  
Creer  
Creencia  
Cristianismo  
Cristo  
Crucifijo  
Cruz  
Cruzadas  
Culpa  
Culto  
Cura

### D

Demonio  
Devoción  
Diablo  
Diez mandamientos  
Dios  
Divinidad  
Dominicas

### E

Ecuménico  
Ermita  
Espabilar  
Espíritu  
Eucaristía  
Existencia

### F

Fe  
Fieles  
Franciscanos

### G

Ganesha  
Guerra santa

**H**

Hagiografía  
Hostia

**M**

Macareno  
Matrimonio  
Meapilas  
Mezquita  
Misa  
Monaguillo  
Monja  
Monjes  
Monoteísmo  
Muerte

**R**

Rezar  
Rito  
Rosario

**I**

Iglesia  
Incienso  
Infierno  
Inquisición  
Islam

**N**

Nazarenos  
Nuevo Testamento

**S**

Sacerdote  
Sacrificio  
Sacristán  
Sacristía  
Sagrado  
San Agustín  
Santa  
Santidad  
Santo  
Santo Domingo  
Sermón  
Sinagoga  
Sotana

**J**

Jesucristo  
Jesús  
Juan Pablo II  
Judío

**O**

Obediencia  
Obispo

**T**

Tiara  
Trinidad  
Templo

**L**

Lutero  
Luz

**P**

Pabito  
Papa  
Papa Francisco  
Pascua  
Patena  
Pecado  
Perdón  
Piedad  
Plegaria  
Politeísmo  
Primera Comunión  
Procesión  
Púlpito

**V**

Velas  
Vida  
Vino  
Virgen

**4.2.2.2. Listado definitivo tras la Fase 2 para el ámbito del fútbol****A**

Apuestas  
Árbitro  
Arquero

**B**

Balón  
Benítez  
Balompié  
Banderas  
Banquillo  
Bola  
Bota  
Botas  
Buffon

**C**

Cabezazo  
Camiseta  
Campo  
Capitán  
Césped  
Champions  
Chutar  
Chut  
Club  
Copa  
Copa del Rey  
Córner  
Crack  
Cruyff  
Culé

**D**

Deporte  
Despejar

**E**

Entrenador  
Entrenamiento  
Equipo  
Escuadra  
Estadio  
Eurocopa

**F**

Falta  
Fichajes  
Futbolistas

**G**

Gol  
Goleada  
Grada

**H**

Hincha

**I**

Iker Casillas

**J**

Juego  
Jugador

**L**

Larguero  
Lesión  
Lesiones  
Liga  
Liguilla  
Linier

**M**

Madridista  
Messi  
Mourinho  
Mundial

**P**

Parar  
Partido  
Pase  
Patadas  
Pelé  
Pelota  
Penalti  
Piqué  
Portería  
Portero

**Q**

Quiniela

**R**

Red  
Reventa  
Rival  
Rivales  
Rivalidad  
Ronaldo

**S**

Seguidor  
Sergio Ramos  
Silbato  
Suárez

**T**

Tarjeta amarilla  
Tarjeta roja

**V**

Vestuario

**4.2.2.3. Listado definitivo tras la Fase 2 para el ámbito de los toros****A**

Abanico  
Albero  
Alternativa  
Arena  
Arrastrar  
Arrastre

**B**

Banderillero  
Banderilla  
Banderillas  
Barrera  
Burladero

**C**

Caballo  
Callejón  
Capote  
Cornada  
Cortar  
Coso  
Cuadrilla  
Cuerno

**D**

Dehesa

**E**

El Cordobés  
El Guerra  
Estampitas

**F**

Faena  
Faenar

**H**

Herida

**I**

Indulto

**L**

Lidia  
Luces  
Lucha

**M**

Manoletilla  
Mantilla  
Montera  
Morlaco  
Mulillas

**O**

Orejas

**P**

Paseíllo  
Pasodoble  
Picador  
Puerta Grande  
Puntilla

**R**

Rabo  
Rejoneador  
Ruedo

**S**

Sangre  
Sangría  
Sol  
Sombra

**T**

Taleguilla  
Tendido  
Torear  
Torero  
Toro  
Traje de luces

### 4.2.3. FASE 3: LA BÚSQUEDA EN LOS CORPUS: DE LAS OCURRENCIAS A LAS EXPRESIONES

Para poder alcanzar el primer objetivo (A) y realizar después la pirámide etnolingüística (fase 4 del proceso de creación), es necesario buscar en los corpus una por una las palabras del listado para analizar todas las ocurrencias en busca de las expresiones pluriverbales, lo que representaría la fase 3 del proceso de creación de las pirámides. Evidentemente, se trata de un trabajo laborioso a la vez que indispensable si se quiere seguir el método Data-driven y demostrar las hipótesis de partida a partir del análisis de datos, como se vio en 1.4. En este caso, se ha optado por seguir el orden alfabético de cada listado, ya que resulta más sencillo. Así pues, hay que entrar en el CORPES, el corpus de referencia que se ha seleccionado para el español (3.2.1) y se marcará en la pestaña *subcorpus* que busque solo las ocurrencias de España.

En las páginas siguientes se recogen todos los resultados obtenidos con la búsqueda para el ámbito cultural de la religión, el fútbol y los toros<sup>54</sup>, y se anotarán todas las expresiones pluriverbales. El paso siguiente será buscar esas expresiones para determinar el número de ocurrencias y poder luego ordenarlas de forma visual en las pirámides (fase 4), como se verá en 4.2.4 para el español y en 4.3.4 para el italiano.

#### 4.2.3.1. Fase 3 con el listado de la religión en español

Si nos fijamos en los tres listados obtenidos tras la criba (fase 2 del proceso de creación), salta a la vista que la religión es el apartado que reúne el mayor número de palabras, y eso se debe a la influencia que ejerce en la cultura y, en consecuencia, en la lengua. Se analizarán ahora una por una buscando por lema (siempre que sea posible) para ver cuántas ocurrencias (casos) proporciona el CORPES. El método consiste en analizarlas una por una, apuntando todas las expresiones pluriverbales destacables. El único inconveniente es que las expresiones no están lematizadas en el CORPES, por lo que hay que introducir una por una todas las formas flexionadas.

54. Las búsquedas se realizaron en septiembre de 2016, por lo que es posible que en la actualidad los resultados obtenidos no se ajusten por completo debido a los cambios y a las actualizaciones que se van introduciendo en el CORPES.

## A

**Agua bendita:** al ser un compuesto pluriverbal, hubo que buscarla directamente como *forma* y no como *lemā*<sup>55</sup> y se obtuvieron 81 casos en 62 documentos en el corpus. No se encontró ninguna expresión relevante, y solo vale la pena mencionar la aparición, en la ocurrencia 65, de la expresión *ser agua bendita* para referirse a que algo nos viene muy bien en un momento concreto.

**Allah:** 12 casos en 10 documentos. Ninguna expresión significativa en el corpus.

**Alma:** 6.456 casos en 2.143 documentos. Destacan en el corpus las expresiones *ser un alma en pena* (53 ocurrencias, para referirse a alguien inquieto/atormentado), *ser el alma de [algo]* (28 ocurrencias, con significado de “ser la cabeza pensante y el pilar de algo”), *ni un alma* (27 ocurrencias, con significado de “nadie”)<sup>56</sup>, *salir [algo] del alma* (21 ocurrencias, expresión utilizada para referirse a que algo sale desde lo profundo de nuestro ser y es muy sincero), *ir/salir como alma que lleva el diablo* (20 ocurrencias, con significado de “salir deprisa y corriendo”) *tener/no tener alma* (20 ocurrencias), *ser/buscar el alma gemela* (20 ocurrencias), *caérsele [a alguien] el alma a los pies* (18 ocurrencias, para indicar que alguien está triste y abatido) y *vender el alma al diablo* (4 ocurrencias, para referirse a alguien que está dispuesto a todo con tal de conseguir sus objetivos)<sup>57</sup>.

**Altar:** 1.086 casos en 528 documentos. Destacan en el corpus las expresiones *tener/poner [a alguien] en un altar* (21 ocurrencias, con significado metafórico de “idolstrar a alguien”) y *llevar [a alguien] al altar* (12 ocurrencias, con significado de “casarse con alguien”).

**Ángel:** 4.592 casos en 1.086 documentos. Destacan en el corpus las expresiones *hacer [algo] como los ángeles* (20 ocurrencias, para decir que “alguien hace algo muy bien”) y *ser un ángel* (18 ocurrencias, con significado metafórico de “ser una persona muy buena y disponible”).

**Apóstol:** 403 casos en 207 documentos. Ninguna expresión relevante en el corpus.

**Arrodillarse:** 597 casos en 337 documentos. No hay ninguna expresión relevante en el corpus, aunque el propio verbo procede del ámbito religioso, al igual que *espabilar y confesarse*.

## B

**Banco:** 9.235 casos en 3.751 documentos, pero no se encontró ninguna expresión relevante en el corpus.

**Basílica:** 328 casos en 202 documentos. Ninguna expresión en el corpus.

**Bautizo:** 230 casos en 158 documentos. Ninguna expresión en el corpus.

**Beato/a:** 404 casos en 160 documentos. Ninguna expresión relevante en el corpus.

55. Es algo que pasará con todas las expresiones pluriverbales, ya que en el CORPES no están lematizadas.

56. Véase 2.2.2.

57. En el DRAE aparecen muchas más expresiones, pero estas son las que se encontraron tras la búsqueda en el CORPES y las que se van a tener en cuenta para las pirámides (Data-driven).

**Belén:** 585 casos en 227 documentos<sup>58</sup>. Ninguna expresión en el corpus.

**Biblia:** 761 casos en 435 documentos. Destaca en el corpus la expresión *ser la biblia* (13 ocurrencias, para referirse a algo que es fundamental y muy importante en un ámbito determinado).

**Bóveda:** 498 casos en 242 documentos. Ninguna expresión relevante en el corpus.

**Buddha:** 3 casos en 3 documentos. Ninguna expresión en el corpus.

**Budismo:** 166 casos en 83 documentos. Ninguna expresión relevante en el corpus.

## C

**Capilla:** 1.203 casos en 518 documentos. Ninguna expresión relevante en el corpus.

**Caridad:** 482 casos en 343 documentos. Destacan en el corpus las expresiones *hacer [algo] por caridad* (27 ocurrencias, con significado de “hacer algo por pena”) y *vivir de la caridad* (18 ocurrencias). *¡Por caridad!* se considera poco relevante por ser una interjección; además, solo aparecen 2 ocurrencias en el corpus.

**Cardenal:** 1.053 casos en 469 documentos. Curiosamente, no aparece en el corpus la expresión *hacerse/salirse/tener un cardenal* para referirse a un moratón.

**Catedral:** 2.073 casos en 883 documentos. Solo se encontró en el corpus la expresión *ser [algo] como una catedral*, usada para denotar importancia o grandeza (9 ocurrencias)<sup>59</sup>.

**Catequesis:** 94 casos en 71 documentos. Ninguna expresión en el corpus.

**Catolicismo:** 377 casos en 171 documentos. Ninguna expresión en el corpus.

**Celibato:** 49 casos en 32 documentos. Ninguna expresión en el corpus.

**Ceremonia:** 2.235 casos en 1.292 documentos. Ninguna expresión relevante en el corpus.

**Cielo:** 7.293 casos en 2.293 documentos. Es una de las palabras con más ocurrencias de este listado y nos brinda varias expresiones en el corpus, de las que cabe destacar, por número de ocurrencias: *tocar el cielo* (35 ocurrencias, con el significado metafórico de “llegar a lo más alto”), *ser un cielo* (31 ocurrencias, con significado metafórico de “ser muy buena persona”), *írsele [a alguien] el santo al cielo* (18 ocurrencias, con el significado de “olvidarse de algo/despistarse”), *estar en el séptimo cielo* (16 ocurrencias, con el significado metafórico de “sentir una gran alegría y estar muy contento”) y *remover cielo y tierra* (13 ocurrencias, con el significado de “buscar algo por todas partes”).

**Cirio:** 119 casos en 86 documentos. Ninguna expresión relevante en el corpus.

**Comulgar:** 230 casos en 179 documentos. Destaca en el corpus la expresión *comulgar con ruedas de molino* (9 ocurrencias), un refrán muy presente en la lengua que alude de forma metafórica a la pretensión o imposición de algo inverosímil o absurdo.

**Confesarse**<sup>60</sup>: 4.751 casos en 2.451 documentos. Aquí pasa algo similar a lo que ocurre con el verbo *arrodillarse*; no hay expresiones relevantes, pero el mismo verbo procede del ámbito religioso. Aun así, no se tendrá en consideración para las pirámides, puesto que el foco del primer objetivo (A) son las expresiones pluriverbales.

58. La mayoría de los casos son del nombre propio Belén, bastante frecuente en España.

59. Por ejemplo, tenemos es una mentira como una catedral.

60. Se buscó el lema confesar.

**Confesionario:** 112 casos en 75 documentos. Ninguna expresión relevante en el corpus.

**Confirmación:** 809 casos en 625 documentos. Ninguna expresión relevante en el corpus, y cabe comentar que el término aparece muchas veces con su acepción principal de *corroborar algo*, sin referirse al sacramento de la Iglesia Católica.

**Convento:** 1.311 casos en 464 documentos. Solo aparece en dos ocasiones en el corpus la expresión coloquial *para lo que me queda en el convento, me cago dentro*, pronunciada para justificar una actuación perjudicial contra un sitio u organización con el que queda poco tiempo de relacionarse, ya que se considera que las consecuencias no afectarán. Se tendrá en cuenta para la pirámide de la religión (4.2.4.1)

**Conversión:** 753 casos en 453 documentos. Ninguna expresión relevante en el corpus.

**Cordero:** 860 casos en 439 documentos. Destaca en el corpus la expresión *ser la madre del cordero* para referirse a que algo es el origen del problema (18 ocurrencias)<sup>61</sup>.

**Credo:** 230 casos en 182 documentos. Ninguna expresión relevante en el corpus.

**Crear:** 50.070 casos en 13.334 documentos. Evidentemente, se trata de un verbo comodín muy frecuente en la lengua, y esto queda claramente reflejado en el número de ocurrencias. Sin embargo, no se ha encontrado en el corpus ninguna expresión relacionada directamente con el ámbito religioso.

**Creencia:** 2.171 casos en 968 documentos. Ninguna expresión relevante en el corpus relacionada directamente con la religión.

**Cristianismo:** 626 casos en 291 documentos. Ninguna expresión relevante en el corpus.

**Cristo:** 1.535 casos en 730 documentos. Destacan en el corpus las expresiones *ni cristo* (10 ocurrencias, con significado de “nadie”<sup>62</sup>), *por los clavos de cristo* (10 ocurrencias, exclamación para delatar asombro), *armar/liar la de dios es cristo* (5 ocurrencias, para referirse al hecho de montar un lío muy grande), *ni cristo* (2 ocurrencias, con significado de “nadie”<sup>63</sup>) e *ir/estar hecho un cristo* (2 ocurrencias, con significado de “estar muy maltratado y en mal estado”).

**Crucifijo:** 253 casos en 139 documentos. Ninguna expresión relevante en el corpus.

**Cruz:** 2.011 casos en 986 documentos. La cruz es uno de los símbolos más importantes del cristianismo, por eso aparece con tanta frecuencia en la lengua. La expresión que más destaca en el corpus es sin duda *ser una cruz* (27 ocurrencias), para referirse a una carga pesada moralmente<sup>64</sup>.

**Cruzadas:** 108 casos en 60 documentos. Ninguna expresión en el corpus.

**Culpa:** 4.107 casos en 1.683 documentos. Ninguna expresión en el corpus.

**Culto:** 2.024 casos en 1.055 documentos. Ninguna expresión en el corpus.

61. Tiene una clara carga religiosa, ya que, según la Biblia, el cordero de Dios es Jesús y, por lo tanto, su madre es la responsable de su misión redentora.

62. Véanse los ejemplos del subapartado 2.2.2.

63. Véase 2.2.2.

64. A menudo aparece en el corpus sin el verbo en interjecciones como ¡Qué cruz!.

**Cura**<sup>65</sup>: 2.569 casos en 854 documentos. Ninguna expresión en el corpus relacionada con el ámbito religioso.

## D

**Demonio**: 1.566 casos en 639 documentos. Destacan en el corpus las interjecciones *qué demonios* (242 ocurrencias) y *cómo demonios* (47 ocurrencias), además de *ser un demonio* (20 ocurrencias, con significado de “persona mala”) y *ser más feo que el demonio* (1 ocurrencia).

**Devoción**: 648 casos en 426 documentos. Destaca en el corpus la expresión *ser o no ser santo de la/mi/tu/nuestra devoción* (20 ocurrencias), usada para delatar agrado o desagrado frente a algo.

**Diablo**: 1.433 casos en 694 documentos. Como se puede apreciar, ocurre algo similar a lo que pasa con la palabra demonio, también por lo que se refiere al número de ocurrencias. Destacan en el corpus las expresiones *ser un pobre diablo* (45 ocurrencias, para referirse a alguien desgraciado), *ir/mandar [a alguien] al diablo* (22 ocurrencias, para expresar rechazo o falta de interés respecto a algo o a alguien), las interjecciones *qué diablos* (102 ocurrencias, solo 1 con el término en singular) y *cómo diablos* (14 ocurrencias, solo 1 con el término en singular), además de *ser el diablo* (3 ocurrencias, para referirse a alguien muy malo o muy astuto) y *ser más malo que el diablo* (1 ocurrencia, para referirse a una persona muy mala).

**Diez mandamientos**: 46 casos en 42 documentos. Ninguna expresión en el corpus.

**Dios**: 13.389 en 3.017 documentos. Se encontraron muchas expresiones en el corpus, in primis las interjecciones *¡por Dios/vaya, por Dios!* (763 ocurrencias), *¡Dios mío/Dios santo!* (respectivamente 639 y 79 ocurrencias, por un total de 718) y *¡por el amor de Dios!* (108 ocurrencias) y *¡Ay Dios!* (30 ocurrencias); luego tenemos ser/hacer algo como *Dios manda* (144 ocurrencias, con significado de “como debe ser/como es debido”), *gracias a Dios* (117 ocurrencias, con significado de “menos mal”), *ni Dios* (92 ocurrencias, con significado de “nadie”)<sup>66</sup>, *ser Dios* (33 ocurrencias, con significado metafórico de “ser alguien muy bueno en lo que hace”) *todo Dios* (32 ocurrencias, con significado de “todo el mundo”), *si Dios quiere* (32 ocurrencias, para remitir a la voluntad de Dios), *cagarse alguien en Dios* (24 ocurrencias, expresión vulgar para expresar sumo desagrado), *Dios dirá* (17 ocurrencias, para remitir a la voluntad de Dios de cara al porvenir), *encomendarse a Dios* (13 ocurrencias, con significado de “confiar en la ayuda divina”), *a la buena de Dios* (10 ocurrencias, con significado de “de cualquier manera”), *Dios te oiga* (9 ocurrencias, con significado de “ojalá”<sup>67</sup>), *Dios aprieta pero no ahoga* (9 ocurrencias, para indicar que siempre hay una salida), *Dios los cría y ellos se juntan* (6 ocurrencias, refrán popular que alude con cierta ironía a la inclinación natural que lleva a juntarse a las personas de un mismo carácter y temperamento), *a Dios rogando y con el mazo dando* (6 ocurrencias, para indicar que no solo hay que encomendarse a la ayuda

65. Estamos frente a un claro caso de homógrafos, ya que un mismo significante presenta dos acepciones principales muy diferentes: por un lado se refiere al sacerdote, y por otro es sinónimo de curación.

66. Véase 2.2.2.

67. El término ojalá también procede del ámbito religioso, en este caso islámico, y su significado es “si Allah quiere”.

divina, sino que es importante trabajar duro), *armar/montar la de Dios es Cristo* (5 ocurrencias, con significado de “montarse un gran escándalo/caos”), *el hombre propone y Dios dispone* (5 ocurrencias, refrán que se utiliza cuando surge un obstáculo o un imprevisto que trastoca o destruye nuestros planes y expectativas) y *Dios le da pan a quien no tiene dientes* (1 ocurrencia, refrán que se utiliza cuando nos encontramos frente a una injusticia)<sup>68</sup>.

**Divinidad:** 502 casos en 209 documentos. Destaca en el corpus la expresión ser una divinidad (30 ocurrencias) para referirse a alguien muy importante que supuestamente merece nuestra admiración.

**Dominicas:** 19 casos en 16 documentos. Ninguna expresión en el corpus.

## E

**Ecuménico:** 61 casos en 51 documentos. Ninguna expresión relevante en el corpus.

**Ermita:** 587 casos en 229 documentos. Ninguna expresión relevante en el corpus.

**Espabilar:** 220 casos en 153 documentos. También existe la variante *despabilar*, de la que se encontraron 45 casos en 35 documentos, con lo que el total de ocurrencias ascendería a 265 en 188 documentos<sup>69</sup>. Destaca en el corpus la expresión ser una persona espabilada (71 ocurrencias) para referirse a alguien despierto y avisado, además del propio verbo espabilar, que en la mayoría de los casos se utiliza con el significado de quitar el sueño, despertar y darse prisa.

**Espíritu:** 5.169 casos en 2.443 documentos. Se encontraron varias expresiones interesantes en el corpus, como *tener espíritu de lucha* (37 ocurrencias; “ánimo de batalla”), *tener espíritu de sacrificio* (27 ocurrencias; “ánimo de esfuerzo”), tener espíritu de equipo (26 ocurrencias; “actitud solidaria con el grupo”), *tener espíritu de superación* (24 ocurrencias; “ganas de superarse”), *tener espíritu de empresa* (6 ocurrencias; “actitud proactiva”<sup>70</sup>), *ser pobre de espíritu* (9 ocurrencias; “persona llana y humilde”), *mantener vivo el espíritu* (6 ocurrencias; “seguir creyendo en algo”), y *fortalecer el espíritu* (2 ocurrencias; “reforzar el alma”).

**Eucaristía:** 159 casos en 99 documentos. Ninguna expresión en el corpus.

**Existencia:** 8.072 casos en 3.175 documentos. Es una palabra común en la lengua (pero no con el valor metafórico), por eso aparece con tanta frecuencia; sin embargo, no se encontró ninguna expresión en el corpus que guarde una relación directa con la religión.

68. Cabe subrayar que en el diccionario de la Real Academia aparecen varias expresiones más (<http://dle.rae.es/?id=DpH1nYv>), pero solo estas que hemos mencionado aparecen el CORPES y se tendrán en cuenta como datos (Data-driven).

69. Cabe subrayar que en el diccionario de la Real Academia aparecen varias expresiones más (<http://dle.rae.es/?id=-DpH1nYv>), pero solo estas que hemos mencionado aparecen el CORPES y se tendrán en cuenta como datos (Data-driven).

70. De este grupo de expresiones, para las pirámides solo se tendrá en cuenta tener espíritu de sacrificio por ser la única que guarda cierta relación con la religión católica.

**F**

**Fe**<sup>71</sup>: 3.047 casos en 1.607 documentos. Se encontraron en el corpus las expresiones *tener/no tener fe [en algo/alguien]* (129 ocurrencias), *dar fe [de algo]* (90 ocurrencias) y *de/con buena/mala fe* (150 ocurrencias, 115 con buena y 35 con mala) para indicar algo que se hace con sinceridad o con malicia.

**Fieles**: 1.135 casos en 778 documentos. Ninguna expresión relevante en el corpus.

**Franciscanos**: 199 casos en 114 documentos. Ninguna expresión en el corpus.

**G**

**Ganesha**: 2 casos en 2 documentos. Ninguna expresión en el corpus.

**Guerra santa**: 90 casos en 59 documentos buscando por forma, ya que se trata de un compuesto pluriverbal. Aparece en dos ocasiones en el corpus la expresión *ser una guerra santa* con sentido figurado para indicar que hay una lucha a brazo partido entre dos bandos enfrentados, pero en la mayoría de las ocurrencias mantiene su significado principal relativo a la guerra religiosa entre moros y cristianos.

**H**

**Hagiografía**: 72 casos en 33 documentos. Ninguna expresión en el corpus.

**Hostia**<sup>72</sup>: 777 casos en 294 documentos. Se trata de una palabra bastante frecuente en España, especialmente en el canal oral/coloquial. De la búsqueda en el CORPES, destacan las expresiones *ser la hostia/de la hostia* (179 ocurrencias, para referirse a algo increíble) *dar/pegar de hostias* (65 documentos, con significado de “pegar”), la interjección *hostias!* (alrededor de 200 ocurrencias, para delatar estupor) y *tener/poner de mala hostia* (42 casos, con significado de “poner/estar de mal humor”).

**I**

**Iglesia**: 7.949 casos en 2.497 documentos. Solo se encontró en el corpus la expresión *cajarse por la iglesia* (alrededor de 22 ocurrencias)<sup>73</sup>.

**Incienso**: 263 casos en 168 documentos. Ninguna expresión en el corpus.

**Infierno**: 1.808 casos en 873 documentos. Destacan en el corpus las expresiones *irse/mandar [algo/a alguien] al infierno* (alrededor de 50 ocurrencias, con significado de “rechazar algo o a alguien que no es de nuestro agrado”).

**Inquisición**: 280 casos en 188 documentos. Ninguna expresión en el corpus.

**Islam**: 741 casos en 387 documentos. Ninguna expresión en el corpus.

71. Esta podría ser una palabra controvertida porque se ha convertido en un término común muy presente en la lengua, pero creemos que no por ello pierde su carga religiosa.

72. Cabe subrayar que en español también existe ostia sin h- inicial, pero tiene otro significado y la búsqueda solo proporcionó 35 casos en 20 documentos, por lo que se decidió tener en cuenta solo la grafía con h-.

73. Al buscar el término en el diccionario de la Real Academia Española, se listan varias expresiones con iglesia (<http://dle.rae.es/?id=KvQlZU1>), pero no aparecieron en la búsqueda en el CORPES, lo que indica que o bien no se utilizan o bien no son frecuentes en la lengua escrita de los últimos quince años en España (Data-driven).

**J**

**Jesucristo:** 432 casos en 236 documentos. Ninguna expresión en el corpus<sup>74</sup>.

**Jesús:** 1.675 casos en 651 documentos<sup>75</sup>. Destaca en el corpus únicamente la interjección *¡Jesús!* (35 ocurrencias).

**Juan Pablo II:** 485 casos en 248 documentos buscando directamente por *forma* al tratarse de un nombre propio. Ninguna expresión en el corpus.

**Judío:** 3.263 casos en 1.070 documentos. Ninguna expresión relevante en el corpus.

**L**

**Lutero:** 97 casos en 42 documentos buscando directamente por *forma* al tratarse de un nombre propio. Ninguna expresión en el corpus.

**Luz:** 20.650 casos en 6.275 documentos<sup>76</sup>. Hay muchas expresiones en el corpus, pero ninguna de ellas está relacionada directamente con el ámbito religioso.

**M**

**Macareno:** 4 casos en 4 documentos buscando por *forma* para evitar que aparezcan todos los casos con el nombre propio femenino Macarena. Ninguna expresión en el corpus.

**Matrimonio:** 4.371 casos en 1.841 documentos. Solo vale la pena mencionar la unidad pluriverbal *cama de matrimonio* (66 ocurrencias), que no se considera relevante para la pirámide de la religión.

**Meapilas:** 40 casos en 33 documentos, y todos para referirse a la expresión coloquial *ser un meapilas* para referirse a un falso devoto.

**Mezquita:** 665 casos en 361 documentos. Ninguna expresión en el corpus.

**Misa:** 1.660 casos en 742 documentos. Destacan en el corpus las expresiones decir misa (20 ocurrencias; expresión coloquial para desestimar lo que alguien pueda decir), *ir a misa [algo] que se dice* (10 ocurrencias, para señalar que algo es indiscutible) y *hablar como en misa* (1 ocurrencia, con significado de “en silencio”).

**Monaguillo:** 119 casos en 67 documentos. Ninguna expresión en el corpus.

**Monje:** 2.185 casos en 691 documentos. Se encontró en el corpus la expresión *el hábito no hace al monje* (4 ocurrencias), para indicar que no hay que juzgar a las personas por su apariencia.

**Monoteísmo:** 34 casos en 25 documentos. Ninguna expresión en el corpus.

**Muerte:** 20.105 casos en 6.989 documentos<sup>77</sup>. No se encontró ninguna expresión relevante en el corpus.

74. Son mucho más frecuentes y originan más expresiones los términos Cristo y Jesús.

75. En español, Jesús es un nombre propio bastante frecuente y aparece en más de la mitad de los casos.

76. Término muy común en la lengua, lo que justifica el número de ocurrencias.

77. Se trata de una palabra muy común que se ha decidido incluir en la lista por la dicotomía que tiene con vida y que de alguna manera se ve plasmada en la religión.

## N

**Nazareno:** 110 casos en 68 documentos. El diccionario de la Real Academia recoge las expresiones *cuando vengan los nazarenos* y *estar alguien hecho un nazareno*<sup>78</sup>, pero no aparecieron en la búsqueda en el CORPES, por lo que no se van a tener en cuenta.

**Nuevo Testamento:** 70 casos en 40 documentos buscando por *forma*, ya que se trata de un compuesto pluriverbal. Ninguna expresión en el corpus.

## O

**Obediencia:** 388 casos en 198 documentos. Ninguna expresión en el corpus relacionada directamente con la religión.

**Obispo:** 1.991 casos en 716 documentos. Ninguna expresión relevante en el corpus.

## P

**Pabilo:** 17 casos en 17 documentos. Ninguna expresión en el corpus.

**Papa Francisco:** 0 casos.

**Pascua:** 429 casos en 221 documentos. Se encontró en el corpus la expresión *estar más contento que unas pascuas* (9 ocurrencias) para indicar un estado de ánimo muy alegre.

**Patena:** 40 casos en 39 documentos. Destaca en el corpus la expresión *dejar/tener/estar/ quedar [algo] como una patena* (24 ocurrencias, con significado de “muy limpio”).

**Pecado:** 1.485 casos en 713 documentos. Pese a la importancia que tiene esta palabra en la religión católica y al gran número de expresiones que brinda el diccionario de la RAE<sup>79</sup>, solo se encontró en el corpus *ser pecado* (76 ocurrencias, para referirse a algo grave o tabú) y *quien esté libre de pecado, que tire la primera piedra* (7 ocurrencias).

**Perdón:** 2.029 casos en 1.018 documentos<sup>80</sup>: Se utiliza constantemente para pedir disculpas y en todas y cada una de las ocurrencias queda manifiesto ese significado; sin embargo, no se encontró ninguna expresión pluriverbal en el corpus, por lo que se descartará por no entrar dentro del primer objetivo (A).

**Piedad:** 1.044 casos en 484 documentos<sup>81</sup>. Destacan en el corpus las expresiones *sin piedad* (221 ocurrencias) y *tener piedad* (36 ocurrencias, la mayoría de ellas en la forma de imperativo con *ten/tenga*).

**Plegaria:** 211 casos en 151 documentos. Destaca en el corpus la expresión *hacer/escuchar las plegarias [de alguien]* (104 ocurrencias, usada cuando suplicamos a alguien para que nos conceda un determinado favor).

**Politeísmo:** 15 casos en 12 documentos. Ninguna expresión en el corpus.

**Primera comunión:** 177 casos en 113 documentos buscando por *forma* al tratarse de un compuesto pluriverbal. Ninguna expresión en el corpus.

78. <http://dle.rae.es/?id=QJUFCAL>

79. <http://dle.rae.es/?id=SFVbedk>

80. Se trata de otra palabra clave en el ámbito religioso, ya que la propia idea del perdón es uno de los cimientos del catolicismo.

81. Otro término estrechamente vinculado con la religión católica.

**Procesión:** 884 casos en 400 documentos. Ninguna expresión en el corpus.

**Púlpito:** 150 casos en 101 documentos. Ninguna expresión en el corpus.

## R

**Rezar:** 1.902 casos en 1.054 documentos. Pese a tratarse de un verbo muy relacionado con la religión, no se ha encontrado ninguna expresión significativa en el corpus.

**Rito:** 951 casos en 545 documentos. Ninguna expresión en el corpus.

**Rosario:** 1.341 casos en 485 documentos. Destaca en el corpus la expresión *acabar/terminar como el rosario de la aurora* (25 ocurrencias, con significado de “terminar muy mal una relación”)<sup>82</sup>.

## S

**Sacerdote:** 1.889 casos en 712 documentos. Ninguna expresión relevante en el corpus.

**Sacrificio:** 1.530 casos en 901 documentos. Ninguna expresión relevante en el corpus.

**Sacristán:** 176 casos en 86 documentos. Ninguna expresión en el corpus.

**Sagrado:** 2.194 casos en 994 documentos. Destaca en el corpus la expresión *ser sagrado* (116 ocurrencias, con significado metafórico de “intocable”).

**San Agustín:** 224 casos en 128 documentos buscando por forma al tratarse de un nombre propio. Ninguna expresión en el corpus.

**Santo:** 5.899 casos en 2.161 documentos. Se encontraron varias expresiones en el corpus, como *ser un santola* (75 ocurrencias, respectivamente 53 para el masculino y 22 para el femenino, con el significado metafórico de “ser una persona muy buena”), *santo y seña* (49 ocurrencias, con significado de “contraseña”), *¿a santo de qué?/¿a qué santo?* (respectivamente 26 y 7 ocurrencias, utilizado normalmente en preguntas para delatar estupor y cuyo significado es “¿con qué motivo?”), *ser o no ser santo de la/mi/tu/nuestra devoción* (20 ocurrencias, como ya se vio en devoción), *írsele [a alguien] el santo al cielo* (18 ocurrencias, como ya se vio en cielo), la interjección *¡por todos los santos!* (12 ocurrencias), quedarse una mujer para vestir santos (9 ocurrencias, con el significado de “quedarse soltera”) *llegar y besar el santo* (6 ocurrencias, utilizada cuando se logra algo con mucha brevedad), *alzarse/quedarse con el santo y la limosna* (3 ocurrencias, con el significado de “apropiarse alguien de todo”), *desnudar a un santo para vestir a otro* (1 ocurrencia, con el significado de “arreglar algo estropeando otra cosa”) y *santo y bueno* (1 ocurrencia, para dar nuestra aprobación a un asunto)<sup>83</sup>.

**Santidad:** 275 casos en 164 documentos. Ninguna expresión en el corpus.

**Santo Domingo:** 259 casos en 168 documentos buscando por forma al tratarse de un nombre propio. Ninguna expresión en el corpus.

**Sermón:** 296 casos en 197 documentos. Destaca en el corpus la expresión *darle/echarle/*

82. En 4.4. se analizará esta expresión y su procedencia, y luego se verá si existe o no un equivalente en italiano, lo que formaría parte del tercer objetivo (C).

83. Al realizar una búsqueda en el Diccionario de la Real Academia Española, aparecen varias expresiones más (<http://dle.rae.es/?id=XGGB4k6>), pero, gracias a la búsqueda en el CORPES es posible saber cuáles de estas se utilizan realmente en el período y ámbito geográfico seleccionado (Data-driven).

**soltarle un sermón [a alguien]** (9 ocurrencias, con significado de “amonestar o reprender a alguna persona sobre algo”).

**Sinagoga:** 173 casos en 87 documentos. Ninguna expresión en el corpus.

**Sotana:** 199 casos en 92 documentos. Ninguna expresión en el corpus.

## T

**Tiara:** 27 casos en 20 documentos. Ninguna expresión en el corpus.

**Trinidad:** 292 casos en 157 documentos. Ninguna expresión en el corpus.

**Templo:** 2.435 casos en 928 documentos. Destaca en el corpus la expresión **ser una verdad como un templo** (11 ocurrencias, con significado metafórico de “ser algo muy cierto y no cuestionable”).

## V

**Vela:** 1.802 casos en 820 documentos. Destacan en el corpus las expresiones **pasar la noche en vela** (171 ocurrencias, “sin dormir”), **quedarse/estar a dos velas** (15 ocurrencias, “sin entender algo/sin dinero”), **no darle [a alguien] vela en este entierro** (11 ocurrencias, “no darle autoridad a alguien para que intervenga”), y **poner una vela a Dios y otra al diablo** (2 ocurrencias, “intentar sacar provecho de varias cosas al mismo tiempo”).

**Vida:** 68.448 casos en 16.908 documentos. Se trata de un vocablo muy común en la lengua y se encontraron numerosas expresiones en el corpus, pero las únicas que están realmente vinculadas con la religión son **pasar a mejor vida** (70 ocurrencias, con significado de “morir”) y **enterrarse en vida** (5 ocurrencias, con significado metafórico de “aislarse de los demás”).

**Vino:** 9.524 casos en 1.866 documentos. La única expresión significativa encontrada en el corpus es **llamar al pan pan y al vino vino** (6 ocurrencias, con significado de “llamar las cosas por su nombre”).

**Virgen:** 2.811 casos en 1.215 documentos. Destacan en el corpus las expresiones **ser un viva la virgen** (5 ocurrencias, para referirse a una persona despreocupada) y **fiate de la virgen y no corras** (1 ocurrencia, para expresar que hay que aplicarse y no confiar demasiado).

Estas son todas las expresiones que se han encontrado en el CORPES a partir del listado filtrado del *brainstorming*, lo que correspondería al primer gran objetivo (A) de esta tesis. La cuarta y última fase en el proceso de realización de las pirámides etnolingüísticas consistirá en ordenar estas expresiones por frecuencia y en aplicar el criterio de rentabilidad comunicativa, como se detallará en 4.2.4. Antes, veamos los resultados de las búsquedas en el CORPES con el listado del fútbol y de los toros.

### 4.2.3.2. Fase 3 con el listado del fútbol en español

En este subapartado se analizarán una por una las palabras del listado del fútbol obtenidas tras la criba para encontrar el mayor número posible de expresiones pluriverbales relacionadas con

este ámbito (objetivo A)<sup>84</sup>. Pese al gran impacto social que tiene este deporte en la sociedad actual, su interacción con la lengua es mucho más reciente respecto a la religión, por lo que el caudal de expresiones en principio será inferior. Veamos los resultados.

## A

**Apuestas:** 2.661 casos en 1.914 documentos. En el corpus, aparecen ocurrencias tanto con el sustantivo como con el verbo *apostar*. Se han analizado una por una, pero no se ha encontrado ninguna expresión significativa relacionada con el fútbol.

**Árbitro:** 1.880 casos en 1.107 documentos. Ninguna expresión en el corpus.

**Arquero:** 101 casos en 74 documentos. Ninguna expresión relacionada con el fútbol en el corpus.

## B

**Balón:** 4.174 casos en 2.148 documentos. Destaca en el corpus la expresión *echar/lanzar/despejar*<sup>85</sup> *balones fuera* (54 ocurrencias), utilizada cuando alguien trata de evitar una situación comprometida.

**Benítez:** 261 casos en 141 documentos buscando por *forma* al tratarse de un nombre propio. Aparecen en el corpus diferentes personas con este apellido, no solo el conocido entrenador de fútbol. De todas formas, no hay ninguna expresión ni casos de ampliación de significado.

**Balompie:** 99 casos en 59 documentos. Ninguna expresión en el corpus.

**Bandera:** 9.955 casos en 1.625 documentos<sup>86</sup>. Ninguna expresión en el corpus relacionada directamente con el fútbol.

**Banquillo:** 1.316 casos en 1.032 documentos. Solo encontramos la expresión *calentar banquillo* (3 ocurrencias), utilizada cuando alguien se queda mucho tiempo de suplente<sup>87</sup>. Cabe mencionar también la variante más coloquial *chupar banquillo* (<http://dle.rae.es/?id=4yZ4djp>), de la que no se encontró ninguna ocurrencia en el CORPES.

**Bola:** 2.519 casos en 1.031 documentos. Se trata de una palabra polisémica con muchas expresiones idiomáticas, pero la mayoría de ellas pertenecen al español de América<sup>88</sup> (<http://dle.rae.es/?id=5mIvKnQ>) y en España son menos frecuentes, como indica el resultado de la búsqueda en el CORPES, donde aparecen únicamente las expresiones *no dar pie con bola* (24 ocurrencias, con significado de “no acertar/no tener un buen día”) y *dar bola [a alguien]*

84. Las búsquedas se realizaron en septiembre de 2016, por lo que es posible que en la actualidad los resultados obtenidos no se ajusten por completo debido a los cambios y a las actualizaciones que se van introduciendo en el CORPES.

85. Véase más adelante también la entrada *despejar*.

86. La mayoría de las expresiones que se encontraron están relacionadas con el ámbito marítimo y militar.

87. Se trata de una expresión muy acotada al ámbito futbolístico, pero se puede utilizar también con significado metafórico para indicar que alguien se queda en segundo plano, sin tener la posibilidad de actuar y demostrar su valor.

88. Respecto al español de América, han sido muy esclarecedoras las clases que recibí del Dr. Antonio Torres Torres, así como la lectura de su libro *El español de América*, publicado por la Universitat de Barcelona.

(8 ocurrencias, con significado de “prestar atención”)<sup>89</sup>.

**Bota:** 1.477 casos en 729 documentos. Destaca en el corpus la expresión *colgar las botas* (19 ocurrencias) con significado principal de “dejar de jugar al fútbol”, pero utilizada metafóricamente con el significado de “dejar una actividad”. También se encontraron muchas ocurrencias de *ponerse alguien las botas* con significado de “comer mucho” o “aprovecharse de algo”, que se descarta por no estar vinculada con el fútbol.

**Buffon:** 71 casos en 41 documentos buscando por *forma* al tratarse de un nombre propio. Ninguna expresión en el corpus.

## C

**Cabezazo:** 297 casos en 248 documentos. Ninguna expresión en el corpus.

**Camiseta:** 2.207 casos en 1.177 documentos. Destaca en el corpus la expresión *sudar la camiseta* (9 ocurrencias) con significado de “esforzarse y poner mucho empeño en algo”.

**Campo:** 17.692 casos en 7.597 documentos. Se trata de una palabra muy común, lo que justifica el número de casos. Nos apoyamos también en el Diccionario de la Real Academia Española y se encontraron muchas expresiones (<http://dle.rae.es/?id=711mEYU>), pero ninguna relacionada con el ámbito futbolístico.

**Capitán:** 4.017 casos en 1.361 documentos. Ninguna expresión relacionada directamente con el fútbol.

**Césped:** 1.045 casos en 661 documentos. Ninguna expresión en el corpus.

**Champions:** 862 casos en 642 documentos buscando por *forma* al tratarse del nombre de una competición deportiva. Ninguna expresión en el corpus.

**Chutar:** 105 casos en 86 documentos. Se encontró en el corpus la expresión *ir [algo] que chuta* (5 ocurrencias), usada para indicar que algo va muy bien<sup>90</sup>.

**Club:** 9.015 casos en 3.522 documentos. Ninguna expresión en el corpus.

**Copa:** 8.577 casos en 3.202 documentos. Ninguna expresión en el corpus.

**Copa del Rey:** 772 casos en 531 documentos buscando por *forma* al tratarse del nombre de una competición deportiva. Ninguna expresión en el corpus.

**Córner:** 378 casos en 290 documentos. Ninguna expresión en el corpus.

**Crack:** 401 casos en 256 documentos. Con más de 200 ocurrencias, destaca en el corpus la expresión *ser un crack* para indicar que alguien es tan bueno en algo que literalmente “rompe moldes”, de ahí la onomatopeya con la palabra inglesa crack<sup>91</sup>.

**Cruyff:** 93 casos en 62 documentos buscando por *forma* al tratarse de un nombre propio. Ninguna expresión en el corpus.

89. Según afirma Sergio Cordero Monge en su tesis doctoral, esta expresión se utiliza en Costa Rica, Ecuador y Bolivia con el significado de “coquetear con alguien” (*Unidades pluriverbales en los diccionarios de español de Costa Rica: análisis y propuesta de descripción*, 443).

90. El verbo proviene de la palabra inglesa *shoot*, que significa “tiro o disparo fuerte de pelota”.

91. Esta expresión se empezó a utilizar en el ámbito deportivo, especialmente en el fútbol, para designar jugadores que sobresalían sobre los demás, pero actualmente ha ampliado su significado y puede representar a cualquier persona que destaque por sus capacidades. Parece un claro ejemplo de la trascendencia que tiene el fútbol en la sociedad moderna, especialmente en España.

**Culé:** 221 casos en 171 documentos. Ninguna expresión en el corpus.

## D

**Deporte:** 5.969 casos en 2.746 documentos. Se trata de un hiperónimo de *fútbol*, por lo que su extensión es muy grande. No se encontró ninguna expresión destacable en el corpus.

**Despejar:** 1.079 casos en 821 documentos. Aparece en el corpus la expresión *despejar balones fuera* (1 ocurrencia), como ya se ha visto en la entrada balón<sup>92</sup>.

## E

**Entrenador:** 4.419 casos en 2.260 ocurrencias. Ninguna expresión relevante en el corpus.

**Entrenamiento:** 3.183 casos en 1.669 documentos. Ninguna expresión en el corpus.

**Equipo:** 33.180 casos 13.120 documentos. Destacar en el corpus la expresión *ser un equipo* (65 ocurrencias), para indicar que hay que colaborar y trabajar juntos por el bien del grupo<sup>93</sup>.

**Escuadra:** 440 casos en 345 documentos. Ninguna expresión en el corpus.

**Estadio:** 2.883 casos en 1.754 documentos. Se trata de una palabra polisémica con diferentes acepciones; sin embargo, la búsqueda en el corpus no desveló ninguna expresión relacionada con el fútbol.

**Eurocopa:** 640 casos en 345 documentos. Ninguna expresión en el corpus.

## F

**Falta:** 14.571 casos en 7.571 documentos. Se trata de una palabra muy frecuente, pero no se encontró en el corpus ninguna expresión relacionada con el fútbol.

**Fichaje:** 1.169 casos en 750 documentos. Se encontró en el corpus la expresión *ser un buen fichaje* (6 ocurrencias) para indicar que una persona ha sido una incorporación positiva en un determinado ámbito. También aparece la interjección *¡menudo fichaje!* (3 ocurrencias), usada para delatar cierta decepción frente a la incorporación de alguien nuevo.

**Futbolista:** 2.771 casos en 1.726 documentos. Ninguna expresión en el corpus.

## G

**Gol:** 7.923 casos en 3.193 documentos. Destaca en el corpus la expresión *meter/colar un gol [a alguien]* (6 ocurrencias) con el significado metafórico de que alguna persona es engañada por otra a causa de un descuido.

**Goleada:** 354 casos en 319 documentos. Destaca en el corpus la expresión *ganar por goleada* (60 ocurrencias), utilizada a menudo con significado metafórico cuando la victoria o la diferencia entre dos conceptos que se comparan es muy abultada.

**Grada:** 1.019 casos en 726 documentos. Ninguna expresión en el corpus.

92. Según el resultado de la búsqueda, son más frecuentes las versiones con *echar y lanzar*.

93. La mayoría de las veces en el corpus aparece en primera persona plural del presente de indicativo (56 ocurrencias, somos un equipo).

**H**

**Hincha:** 197 casos en 134 documentos. Ninguna expresión en el corpus.

**I**

**Iker Casillas:** 160 casos en 134 documentos buscando directamente por *forma* al tratarse de un nombre propio. Ninguna expresión en el corpus.

**J**

**Juego:** 20.063 casos en 7.889 documentos. Es una palabra muy frecuente en la lengua, pero solo se encontró una expresión en el corpus relacionada con el fútbol y *es pillar [a alguien] en fuera de juego* (13 ocurrencias), para indicar que una persona no se encuentra en una posición correcta para contestar una determinada pregunta o resolver un problema<sup>94</sup>.

**Jugador:** 14.323 casos en 5.378 documentos. Ninguna expresión en el corpus relacionada directamente con el fútbol.

**L**

**Larguero:** 276 casos en 239 documentos. Ninguna expresión en el corpus.

**Lesión:** 4.411 casos en 2.278 documentos. Ninguna expresión en el corpus.

**Liga:** 6.796 casos en 3.413 documentos. Se trata de otra palabra polisémica, y no se encontró en el corpus ninguna expresión relacionada con el fútbol a excepción de *haber liga [en algún lugar]* (1 ocurrencia), con el significado de que “en ese sitio también hacen lo mismo”.

**Liguilla:** 129 casos en 96 documentos. Ninguna expresión en el corpus.

**Linier:** 36 casos en 16 documentos. Ninguna expresión en el corpus.

**M**

**Madridista:** 886 casos en 586 documentos. Ninguna expresión en el corpus.

**Messi:** 738 casos en 347 documentos buscando por *forma* al tratarse de un nombre propio. No se encontró ningún caso de ampliación de significado en el corpus.

**Mourinho:** 654 casos en 269 documentos buscando directamente por *forma* al tratarse de un nombre propio. Ningún caso de ampliación de significado en el corpus.

**Mundial:** 12.161 casos en 6.275 documentos. Ninguna expresión relevante en el corpus.

**P**

**Parar:** el problema con esta palabra es que al buscar por *lema*, el CORPES en principio debería buscar todas las formas lematizadas del verbo, pero resulta que aparece también la preposición *para*, con lo que se obtendrían 501.629 casos en 65.734 documentos, algo que

94. También puede significar que alguien ha sido pillado con las manos en la masa.

entorpecería demasiado nuestra búsqueda, como se puede observar en la siguiente foto<sup>95</sup>:

The screenshot shows the CORPES search interface. At the top, there's a search bar with 'parar' entered. Below it, there are various filters and a table of search results. The table has columns for 'REF (Clasificación, país)', 'CONCORDANCIA', and 'Ordenar por'. The first result is: '1 2001 Esp es gastronómica, un "burr" de siglos que se llama el "Stagg her" antigua posada para las diligencias, que ha hecho historia como el primero que recibe una estrella'. The interface also shows a navigation bar with 'Inicio', 'Ayuda', and 'Sugerencias'.

Imagen 17. Consulta de la palabra *parar* en el CORPES. (28 de septiembre de 2016)

**Partido:** 32.066 casos en 11.631 documentos. Relacionada con el fútbol, solo se encontró en el corpus la expresión *hacerse dueño del partido* (9 ocurrencias), para indicar que alguien se ha hecho con el control de una determinada situación.

**Pase:** 2.281 casos en 1.606 documentos<sup>96</sup>. Relacionada con el fútbol, solo se encontró en el corpus la expresión *hacer un pase a un periodista* (1 ocurrencia), con el significado de “pasar una noticia a la prensa, normalmente de forma exclusiva”.

**Patadas:** 865 casos en 510 documentos. Ninguna de las expresiones que aparecen en el corpus (a patadas/dar la patada a alguien) está relacionada directamente con el fútbol y se consideran demasiado generales como para incluirlas en la pirámide.

**Pelé:** 115 casos en 69 documentos buscando por *forma* al tratarse de un nombre propio. Curiosamente, aparecen 5 ocurrencias en el corpus con la expresión *ser el Pelé [de algo]*, un claro caso de ampliación de significado para indicar, una vez más, que alguien es muy bueno en lo que hace. Sin embargo, hay que subrayar que todas aparecen en el mismo documento, un detalle que se tendrá en cuenta para ordenar esta palabra en la pirámide, ya que en este caso la frecuencia está influenciada por el idiolecto del autor<sup>97</sup>.

**Pelota:** 2.348 casos en 1.145 documentos<sup>98</sup>. Se encontraron en el corpus algunas expre-

95. Se trata de un error llamativo a la vez que inesperado, aunque desde el principio se ha tenido en cuenta el hecho de que fuera una versión beta y es posible que en la actualidad el problema se haya resuelto. Pese a estos pequeños imprevistos, seguimos confiando en nuestra decisión de utilizar la versión beta del CORPES en lugar del Corpus del español por tener una interfaz más sencilla y sobre todo por no necesitar registro. Para agilizar el proceso, se ha decidido buscar en el CORPES el verbo *parar* como *forma* (introduciendo las flexiones) y no se encontró ninguna expresión relacionada directamente con el fútbol.

96. Además de ser una palabra polisémica, aquí hay casos de homógrafos (*pase* sustantivo/*pase* forma flexionada del verbo pasar), por lo que la búsqueda ha resultado más complicada de lo previsto.

97. Según el índice en el CORPES, los ejemplos proceden del libro de Pradera, Máximo: *¿De qué me suena eso? Paseo informal por la música clásica*. Madrid: Santillana, 2005.

98. Palabra polisémica con diferentes acepciones, como indica el DRAE <http://dle.rae.es/?id=SPDvCpX|SPeziQe>.

siones idiomáticas como *devolver la pelota* [a alguien], (15 ocurrencias, con significado de “rechazar algo y devolvérselo al remitente”) y *dejar la pelota en el tejado* [de alguien/algo], (6 ocurrencias, con significado de que “el éxito de un negocio es todavía incierto”). No tienen una relación directa con el fútbol, sino con el juego de la pelota en general, por lo que no se van a tener en cuenta.

**Penalti:** 1.454 casos en 830 documentos. Se encontró en el corpus la expresión *casarse de penalti* (3 ocurrencias), con significado de “casarse como obligación por haberse quedado embarazada la mujer”<sup>99</sup>.

**Piqué:** 574 casos en 284 documentos buscando por *forma* al tratarse de un nombre propio<sup>100</sup>. No se encontró en el corpus ninguna ampliación de significado con el nombre del futbolista.

**Portería:** 1.085 casos en 815 documentos. Ninguna expresión en el corpus.

**Portero:** 2.217 casos en 1.223 documentos. Ninguna expresión en el corpus relacionada con el fútbol.

## Q

**Quiniela:** 214 casos en 156 documentos. Ninguna expresión en el corpus.

## R

**Red:** 17.389 casos en 6.536<sup>101</sup>. No se encontró ninguna expresión en el corpus relacionada con el fútbol.

**Reventa:** 59 casos en 42 documentos. Ninguna expresión en el corpus.

**Rival:** 5.455 casos en 3.657 documentos. No se encontró ninguna expresión en el corpus relacionada directamente con el fútbol.

**Rivalidad:** 509 casos en 326 documentos. Ninguna expresión en el corpus.

**Ronaldo:** 1.237 casos en 557 documentos buscando por *forma* al tratarse de un nombre propio. La mayoría de los casos se refieren al famoso jugador brasileño y los otros al más joven jugador portugués. Se encontró un solo caso de ampliación de significado en la expresión *ser el Ronaldo [de algo]* para indicar que alguien es muy bueno en lo que hace, como ya se vio con la entrada Pelé<sup>102</sup>.

## S

**Seguidor:** 2.190 casos en 1.398 documentos. Ninguna expresión en el corpus.

**Sergio Ramos:** 210 casos en 163 documentos buscando por *forma* al tratarse de un nom-

99. Evidentemente, hoy en día esta expresión es mucho menos frecuente, como indica la búsqueda en el CORPES, por razones socio-culturales evidentes. Aventuramos la teoría de que vaya desapareciendo en los próximos años y deje de utilizarse.

100. En la búsqueda aparece también el sustantivo llano *pique*, ya que el CORPES no distingue las palabras con acento.

101. Se trata de otra palabra polisémica muy frecuente en la lengua, lo que justifica el gran número de casos.

102. Según el CORPES, aparece en un artículo de prensa de José Comas titulado “Escándalo en Alemania por el paso de Schröder al sector privado”. *El País.com*. Madrid: elpais.com, 14/12/2005.

bre propio<sup>103</sup>. Ningún caso de ampliación de significado en el corpus.

**Silbato:** 109 casos en 83 documentos. Ninguna expresión en el corpus.

**Suárez:** 443 casos en 197 documentos buscando por *forma* al tratarse de un nombre propio. No se encontró ningún caso de ampliación de significado relacionado con el futbolista.

## T

**Tarjeta amarilla:** 129 casos en 123 documentos buscando por *forma* al tratarse de un compuesto pluriverbal. Se encontró en el corpus la expresión idiomática *sacar tarjeta amarilla* (2 ocurrencias), con el significado de “amonestar a alguien por algo considerado poco correcto”.

**Tarjeta roja:** 74 casos en 68 documentos buscando por *forma* al tratarse de un compuesto pluriverbal. Se encontró en el corpus la expresión *sacar/mostrar tarjeta roja* (8 ocurrencias), con significado de “aplicar un castigo severo por una falta grave”.

## V

**Vestuario:** 1.987 casos en 1.299 documentos. Ninguna expresión en el corpus.

Tras la búsqueda en el corpus, se puede colegir que el fútbol tiene una discreta influencia en la lengua española, y eso se ve plasmado en un buen número de expresiones pluriverbales de uso común. Además, y pese a que en este trabajo se sigue la metodología Data-driven (1.4), se han consultado, *a posteriori*, el *Diccionario del fútbol*, de Antoni Nomdedeu Rull, y la tesis de licenciatura *Terminología futbolística*, de Joaquim M<sup>a</sup>. Puyal Ortiga, actual comentarista de fútbol en la radio catalana, y se ha podido corroborar que la mayoría de las expresiones citadas coinciden con las que aparecen en la búsqueda realizada en el CORPES, lo que sugeriría que tanto el *brainstorming* (fase 1 del proceso de creación) como la **búsqueda en el corpus** (fase 3 del proceso de creación), primeros dos pasos de la triangulación metodológica mixta que se utiliza en este proyecto, han dado resultados hasta el momento satisfactorios para alcanzar el primer objetivo (A).

La cuarta y última fase en el proceso de creación de las pirámides etnolingüísticas consistirá en ordenar estas expresiones por frecuencia y por rentabilidad comunicativa para visualizar los datos, como se detallará en 4.2.4. Sin embargo, antes falta por ver el listado de los toros en español.

### 4.2.3.3. Fase 3 con el listado de los toros en español

En el tercer y último ejemplo de la fase 3 se procederá a analizar un ámbito cultural cuya presencia en España está muy consolidada, mientras que en Italia es casi nula, lo que haría suponer una diferencia evidente entre las dos lenguas en cuanto al número de expresiones y

103. Se trata de un apellido muy común y aparecen varios personajes públicos en el resultado de la búsqueda.

su visualización en las pirámides<sup>104</sup>. Pese a ello, el mundo taurino ya no es tan popular como antaño y puede que este factor tenga repercusiones en la lengua española, con expresiones que de cara al futuro tengan la tendencia a desaparecer. Vamos a centrarnos ahora en el listado obtenido tras la criba de la fase 2 y ver qué resultados proporciona<sup>105</sup>.

## A

**Abanico:** 871 casos en 659 documentos. Ninguna expresión en el corpus.

**Albero:** 69 casos en 56 documentos. Ninguna expresión en el corpus.

**Alternativa:** 4.144 casos en 2.558 documentos. Se encontró en el corpus la expresión *tomar la alternativa* (2 ocurrencias) con significado metafórico de “titularse o graduarse en algo”). Es una expresión que procede del mundo taurino, ya que se refiere a la ceremonia en la que el novillero pasa a ser matador de toros, y para ello es autorizado a alternar en un festejo con el torero que lo apadrina.

**Arena:** 3.401 casos en 1.548 documentos. No se encontró en el corpus ninguna expresión relacionada con el mundo taurino.

**Arrastrar:** 4.164 casos en 1.888 documentos. Ninguna expresión en el corpus.

**Arrastre:** 259 casos en 188 documentos. Destaca en el corpus la expresión *estar para el arrastre* (21 ocurrencias), para referirse a alguien que se encuentra al límite de sus fuerzas y no puede más. La expresión viene de la imagen del toro que, tras ser estoqueado, es arrastrado por el ruedo hacia el desolladero.

## B

**Banderillero:** 63 casos en 36 documentos. Ninguna expresión en el corpus.

**Banderilla:** 134 casos en 87 documentos. Ninguna expresión en el corpus.

**Barrera:** 2.408 casos en 1.508 documentos. Ninguna expresión en el corpus relacionada con el mundo taurino.

**Burladero:** 45 casos en 37 documentos. Ninguna expresión en el corpus.

## C

**Caballo:** 4.720 casos en 1.632 documentos. Ninguna expresión en el corpus relacionada con los toros<sup>106</sup>.

**Callejón:** 660 casos en 392 documentos. Ninguna expresión en el corpus.

**Capote:** 318 casos en 179 documentos. Se encontró en el corpus la expresión *echarle/ lanzarle un capote [a alguien]* (28 ocurrencias) con el significado de “ayudar a alguien”.

**Cornada:** 94 casos en 63 documentos. Ninguna expresión en el corpus.

104. Esta es una de nuestras hipótesis de partida (1.2) que se aspira a demostrar mediante el presente trabajo.

105. Las búsquedas se realizaron en septiembre de 2016, por lo que es posible que en la actualidad los resultados obtenidos no se ajusten por completo debido a los cambios y a las actualizaciones que se van introduciendo en el CORPES.

106. Según el Diccionario de la Real Academia Española, existe una expresión procedente del mundo de la tauromaquia y es *sacar bien/limpio el caballo* para indicar que sale bien de una disputa o un peligro (<http://dle.rae.es/?id=6OPHOdx>). Sin embargo, en la búsqueda realizada en el CORPES no aparece ninguna ocurrencia de esta expresión, por lo que se ha decidido no tenerla en cuenta para la pirámide de los toros (Data-driven).

**Cortar:** 7.583 casos en 2.820 documentos. Se encontró en el corpus la expresión *cortarse la coleta* (12 ocurrencias), expresión metafórica utilizada cuando alguien abandona una profesión, ya que era costumbre entre los toreros lucir una coletilla de pelo natural que se cortaban cuando se retiraban.

**Coso:** 168 casos en 107 documentos. Ninguna expresión en el corpus.

**Cuadrilla:** 371 casos en 223 documentos. Ninguna expresión en el corpus.

**Cuerno:** 726 casos en 403 documentos. Relacionada con los toros, se encontró en el corpus la expresión *coger/tomar/agarrar el toro por los cuernos* (51 ocurrencias) con significado de “enfrentarse resueltamente a una dificultad”<sup>107</sup>.

## D

**Dehesa:** 333 casos en 165 documentos. Ninguna expresión en el corpus.

## E

**El Cordobés:** 137 casos en 57 documentos buscando por *forma* al tratarse de un nombre propio. Ningún caso de ampliación de significado en el corpus.

**El Guerra:** 11 casos en 4 documentos buscando por *forma* al tratarse de un nombre propio. Ningún caso de ampliación de significado en el corpus.

**Estampitas:** 34 casos en 29 documentos. Ninguna expresión en el corpus.

## F

**Faena:** 862 casos en 492 documentos. Destaca la expresión en el corpus *rematar/terminar/acabar la faena* (73 ocurrencias, la mayoría con rematar), con significado de “llevar a cabo un trabajo de forma brillante”. Procede del momento final en el que el torero tiene que matar al toro para coronar su actuación.

**Faenar:** 261 casos en 172 documentos. Se encontró en el corpus la expresión *salir/volver a faenar* (51 ocurrencias), pero no procede directamente del mundo taurino y su significado es más general<sup>108</sup>, por lo que no se tendrá en cuenta de cara a la pirámide de los toros.

## H

**Herida:** 3.909 casos en 2.117 documentos. Ninguna expresión en el corpus relacionada con los toros.

## I

**Indulto:** 259 casos en 126 documentos. Ninguna expresión en el corpus.

107. Se trata de una expresión curiosa, ya que nadie durante una corrida de toros suele coger al animal por los cuernos, y es posible que su origen tenga que ver con la tradición de los *forcados* en Portugal, un grupo de ocho personas que al final de la corrida intentan inmovilizar el toro agarrándolo por los cuernos y resistiendo la embestida.

108. Como se puede ver en el DRAE <http://dle.rae.es/?id=HUBDq0R>.

**L**

**Lidia:** 270 casos en 56 documentos<sup>109</sup>. Con referencia al mundo taurino, solo se encontró en el corpus la expresión *estar cabreado como un toro de lidia* (1 ocurrencia), para indicar un gran enfado.

**Luces:** 2.802 casos en 1.332 documentos buscando por forma para evitar el singular luz. Ninguna expresión en el corpus.

**Lucha:** 7.043 casos en 4.094 documentos. Ninguna expresión en el corpus relacionada con los toros.

**M**

**Manoletina:** 43 casos en 27 documentos. Ninguna expresión en el corpus.

**Mantilla:** 180 casos en 89 documentos. Ninguna expresión en el corpus.

**Montera:** 129 casos en 88 documentos. Destaca en el corpus la expresión *ponerse el mundo por montera* (28 ocurrencias), con significado de “llevar a cabo algo sin reparar en las consecuencias”.

**Morlaco:** 31 casos en 23 documentos. Ninguna expresión en el corpus.

**Mulilla:** 9 casos en 5 documentos. Ninguna expresión en el corpus.

**O**

**Oreja:** 3.028 casos en 1.174 documentos. Ninguna expresión en el corpus relacionada con los toros.

**P**

**Paseílo:** 101 casos en 68 documentos. Ninguna expresión en el corpus.

**Pasodoble:** 170 casos en 102 documentos. Ninguna expresión en el corpus.

**Picador:** 198 casos en 74 documentos. Ninguna expresión en el corpus.

**Puerta Grande:** 118 casos en 108 documentos buscando directamente por *forma* al tratarse de un nombre propio. Destaca en el corpus la expresión *salir por la puerta grande* (25 ocurrencias), utilizada para designar a quien obtiene un gran triunfo.

**Puntilla:** 539 casos en 394 documentos. Se encontraron en el corpus las expresiones de *puntillas*<sup>110</sup> (317 ocurrencias, con significado de “sigilosamente, pisando con la punta de los pies”) y *dar/poner la puntilla* (128 ocurrencias, con significado de “rematar/acabar algo”).

**R**

**Rabo:** 532 casos en 259 documentos. Se encontró en el corpus la expresión *hasta el rabo todo es toro* (4 ocurrencias) para indicar que no se puede bajar la guardia hasta que haya finalizado la tarea.

**Rejoneador:** 69 casos en 37 documentos. Ninguna expresión en el corpus.

**Ruedo:** 432 casos en 248 documentos. Destaca en el corpus la expresión *salir al ruedo* (9

109. La mayoría de las veces aparece el nombre propio femenino.

110. Esta expresión no se tendrá en cuenta para la pirámide porque no procede directamente del mundo taurino.

ocurrencias)<sup>111</sup>, con significado metafórico de “entrar en escena”.

## S

**Sangre:** 8.649 casos en 2.871 documentos. Ninguna expresión en el corpus relacionada directamente con el mundo taurino.

**Sangría:** 251 casos en 177 documentos. Ninguna expresión en el corpus.

**Sol:** 9.757 casos en 2.956 documentos. Destaca en el corpus la expresión *hacer un brindis al sol* (23 ocurrencias), utilizada a menudo para indicar que alguien hace falsas promesas para mantener contento al público.

**Sombra:** 4.348 casos en 1.757 documentos. Ninguna expresión en el corpus relacionada con los toros.

## T

**Taleguilla:** 13 casos en 12 documentos. Ninguna expresión en el corpus.

**Tendido:** 285 casos en 197 documentos. Ninguna expresión en el corpus.

**Torear**<sup>112</sup>: 471 casos en 221 documentos. No se encontró ninguna expresión pluriverbal en el corpus.

**Torero:** 1.535 casos en 580 documentos. Se encontró en el corpus la expresión *saltarse [algo] a la torera* (31 ocurrencias) con significado de “soslayar una obligación”.

**Toro:** 4.216 casos en 1.239 documentos<sup>113</sup>. Destacan en el corpus las expresiones *coger/tomar/agarrar* el toro por los cuernos (51 ocurrencias, como ya se vio en la entrada cuerno), *a toro pasado* (36 ocurrencias, con significado de “tras dejar pasar la oportunidad”), la interjección *¡al toro!* (9 ocurrencias, utilizada como incitación para enfrentarse a un asunto), *pillar el toro [a alguien]* (5 ocurrencias, con significado de “no conseguir terminar algo en el plazo fijado” o “llegar tarde”) *ver los toros desde la barrera* (3 ocurrencias, con significado de “presenciar algo sin correr peligro”) y *estar cabreado como un toro de lidia* (1 ocurrencia, como ya se vio en *lidia*)<sup>114</sup>.

**Traje de luces:** 26 casos en 21 documentos buscando directamente por *forma* al tratarse de un compuesto pluriverbal. Ninguna expresión en el corpus.

A primera vista, parece que este tercer listado ha proporcionado un buen número de expresiones pluriverbales, incluso más que con el fútbol, y eso tal vez se debe a que el mundo taurino

111. En español también existen las variantes tirarse al ruedo y lanzarse al ruedo, pero no han aparecido en el corpus, por lo que no se tendrán en cuenta (Data-driven).

112. El mismo verbo procede del ámbito taurino y puede ser utilizado metafóricamente con el significado de lidiar con alguien o con alguna dificultad. No se tendrá en cuenta para la pirámide de los toros al no tratarse de una expresión pluriverbal.

113. Como cabía suponer desde un principio, se trata de la palabra más importante del listado y es la que nos brinda más expresiones.

114. El Diccionario de la Real Academia Española muestra varias expresiones más (<http://dle.rae.es/?id=a6ZyZYZ|a6aLY-nG>), pero estas son las únicas que se han encontrado en el CORPES y que por tanto se tendrán en cuenta en la pirámide española de los toros (Data-driven).

es un ámbito cultural que se remonta al siglo XII, como señala Bartolomé Benassar (*Historia de la tauromaquia: una sociedad del espectáculo*, cap. 1) y que desde entonces ha sido muy popular en España, con lo que su impacto en la lengua ha sido mayor. Al final de la búsqueda, se han cotejado estos resultados con lo que aparece en algunos escritos sobre el lenguaje taurino, como el ensayo *¡Derecho al toro! El lenguaje taurino y su influencia en lo cotidiano de Carlos Abella* y el interesante artículo de Francisco Reus Boyd-Swan (“El léxico taurino en la vida cotidiana”), y la gran mayoría de las expresiones coinciden con las que se encontraron en esta fase 3, por lo que se considera que tanto el *brainstorming* como la búsqueda en el corpus, primeros dos pasos de la triangulación metodológica mixta utilizada en este proyecto, has servido para alcanzar el primer objetivo (A) que se planteó al principio de esta tesis en 1.2. A continuación, se verá cómo manejar y ordenar todas estas expresiones en las pirámides (fase 4 del proceso de creación) para ofrecer una visualización de toda la información cuantitativa (tercer paso de la triangulación metodológica mixta) en base a los criterios de **frecuencia y rentabilidad comunicativa** (objetivo B).

#### 4.2.4. FASE 4: SELECCIÓN Y ORDENACIÓN DE LAS EXPRESIONES DENTRO DE LAS PIRÁMIDES

La fase 4 representa el último paso dentro del proceso de creación de las pirámides etnolingüísticas y consiste en listar todas las expresiones encontradas con la búsqueda en el corpus (fase 3) para luego visualizarlas en las pirámides de acuerdo con una serie de criterios que se detallarán en 4.2.4.1, 4.2.4.2 y 4.2.4.3.

En primer lugar, las expresiones se ordenarán **según el número de ocurrencias**; es decir, ateniéndose únicamente a su **frecuencia**. En segundo lugar, se aplicará el **criterio de la rentabilidad comunicativa** para destacar aquellas expresiones que se pueden utilizar en la mayoría de los contextos comunicativos sin que resulten vulgares o demasiado coloquiales. También se realizará, en algunos casos, una segunda búsqueda en los corpus orales que se han seleccionado para corroborar algunos aspectos lingüísticos de uso, como se verá a continuación.

Esta fase 4 completa el proceso de creación de las pirámides etnolingüísticas y es fundamental para presentar de forma visual los datos obtenidos tras la búsqueda en los corpus (fase 3). Dicha visualización representa el tercer paso de la triangulación metodológica escogida para el presente trabajo de investigación y servirá para alcanzar el segundo objetivo (B) de esta tesis.

##### 4.2.4.1. La pirámide etnolingüística de la religión en español

Teniendo en cuenta el gran número de expresiones que proporcionó la búsqueda en el CORPES (fase 3 del proceso de creación), es evidente que esta pirámide va a ser la más grande y compleja de las tres, y la que, por consiguiente, más aclaraciones va a necesitar desde el punto de vista metodológico. Para empezar, se van a listar todas las expresiones según el número de ocurrencias:

**> 90 ocurrencias**

- ¡Por Dios/vaya, por Dios! (763)
- ¡Dios mío/Dios santo! (639)
- ¡Qué demonio/s! (242)
- Sin piedad (221)
- Ser la hostia (179)
- Pasar la noche en vela (171)
- De/con buena/mala fe (150)
- Ser/hacer [algo] como Dios manda (144)
- Tener fe (129)
- Ni Dios/ni un alma/ni cristo (121)<sup>115</sup>
- Gracias a Dios (117)
- Ser sagrado (116)
- ¡Por el amor de Dios! (108)
- Hacer/escuchar las plegarias [de alguien] (104)
- ¡Qué diablo/s! (102)
- Dar fe [de algo] (90)

**Entre 20 y 89 ocurrencias**

- Ser pecado (76)
- Ser un santo (76)
- Pasar a mejor vida (70)
- Dar/pegar de hostias (65)
- Ser un alma en pena (53)
- Irse/mandar al infierno (50)
- Santo y seña (49)
- ¡Cómo demonio/s! (47)
- Tener/poner de mala hostia (42)
- Ser un meapilas (40)
- Tener piedad (36)
- Tocar el cielo (35)
- ¡A santo de qué!/¡A qué santo! (33)
- Ser Dios (33)
- Todo dios (32)
- Si Dios quiere (31)
- Ser un cielo (31)
- Ser una divinidad (30)
- Ser el alma [de algo] (28)
- Ser una cruz (27)
- Hacer [algo] por caridad (27)

115. Se ha decidido juntar las tres expresiones, ya que son tres variantes con un mismo significado (2.2.2).

Tener espíritu de sacrificio (27)  
 Acabar/terminar como el rosario de la aurora (25)  
 Cagarse en Dios (24)  
 Dejar/tener/estar/quedar como una patena (24)  
 Ir/mandar al diablo (22)  
 Casarse por la iglesia (22)  
 Tener/poner [a alguien] en un altar (21)  
 Salir del alma (21)  
 Ir/salir como alma que lleva el diablo (21)  
 Decir misa (20)  
 No tener alma (20)  
 Ser/buscar el alma gemela (20)  
 Ser un demonio (20)  
 Ser santo de la/mi/tu/nuestra devoción (20)  
 Hacer [algo] como los ángeles (20)

### Entre 10 y 19 ocurrencias

Ser un ángel (18)  
 Ser la madre del cordero (18)  
 Írsele [a alguien] el santo al cielo (18)  
 Caérsele [a alguien] el alma a los pies (18)  
 Vivir de la caridad (18)  
 Dios dirá (17)  
 Estar en el séptimo cielo (16)  
 Quedarse/estar a dos velas (15)  
 ¡Cómo diablo/s! (14)  
 Encomendarse a Dios (13)  
 Remover cielo y tierra (13)  
 Ser la biblia (13)  
 ¡Por todos los santos! (12)  
 Llevar al altar (12)  
 Ser una verdad como un templo (11)  
 No darle [a alguien] vela en este entierro (11)  
 Ir a misa [algo] que se dice (10)  
 A la buena de Dios (10)  
 Ni cristo (10)  
 Por los clavos de cristo (10)

### < 10 ocurrencias

Dios te oiga (9)  
 Dios aprieta pero no ahoga (9)  
 Quedarse una mujer para vestir santos (9)

Estar más contento que unas pascuas (9)  
 Comulgar con ruedas de molino (9)  
 Ser como una catedral (9)  
 Ser pobre de espíritu (9)  
 Darle/echarle/soltarle un sermón [a alguien] (9)  
 Quien esté libre de pecado, que tire la primera piedra (7)  
 Dios los cría y ellos se juntan (6)  
 A dios rogando y con el mazo dando (6)  
 El hombre propone y Dios dispone (6)  
 Llamar al pan pan y al vino vino (6)  
 Llegar y besar el santo (6)  
 Mantener vivo el espíritu (6)  
 Armar la de Dios es Cristo (5)  
 Enterrarse en vida (5)  
 Ser un viva la virgen (5)  
 Vender el alma al diablo (4)  
 El hábito no hace al monje (4)  
 Ser el diablo (3)  
 Alzarse/quedarse con el santo y la limosna (3)  
 Poner una vela a Dios y otra al diablo (2)  
 Ser una guerra santa (2)  
 Fortalecer el espíritu (2)  
 Ir/estar hecho un cristo (2)  
 Para lo que me queda en el convento, me cago dentro (2)  
 Dios le da pan a quien no tiene dientes (1)  
 Desnudar a un santo para vestir a otro (1)  
 Fíate de la virgen y no corras (1)  
 Santo y bueno (1)  
 Hablar como en misa (1)  
 Ser más malo que el diablo (1)  
 Ser más feo que el demonio (1)

Tenemos un total de **106 expresiones** que presentamos ordenadas por frecuencia. A diferencia de lo que se verá en español con la pirámide del fútbol (4.2.4.2) y de los toros (4.2.4.3), aquí nos encontramos con mucha información desde el punto de vista cuantitativo, por lo que es menester organizar bien estas expresiones dentro de la pirámide ateniéndose a **tres criterios principales** que servirán para ofrecer una mejor visualización (tercer paso de la triangulación metodológica):

**1) PRIMER CRITERIO: rentabilidad comunicativa**, con el que se usarán los símbolos ++ en caso de que la expresión sea rentable en la mayoría de los contextos, + en caso de que se considere coloquial y – con expresiones vulgares. Evidentemente, las

expresiones con el símbolo ++ son las que más interesan y que merecen ocupar un nivel superior en la pirámide, independientemente de su frecuencia. Por el contrario, expresiones vulgares o coloquiales ocuparán niveles inferiores. Para asignar los valores se ha buscado apoyo en el Diccionario de la Real Academia Española (DRAE), que suele utilizar la marca coloq. (símbolo + en la pirámide) y también se ha recurrido al corpus Val.Es.Co. y a la web como corpus, aunque somos conscientes de que la red cambia de forma vertiginosa y por ello la búsqueda realizada en septiembre de 2016 dará unos resultados diferentes en la actualidad, de ahí que se recurra a la web como corpus solo de manera puntual. Las expresiones vulgares (símbolo –) solo son dos: *cagarse en Dios* y *para lo que me queda en el convento*, me cago dentro, pero solo aparecerá la primera en la pirámide porque consideramos que, si bien va a ocupar el nivel más bajo, no se puede obviar debido a su alta frecuencia. La segunda, en cambio, solo presenta dos ocurrencias en el CORPES y no se considera relevante. Cabe subrayar que se incluirán en la pirámide y se marcarán todos los refranes con ++ por tratarse de expresiones que aportan vitalidad a la lengua, además de preciada información etnolingüística<sup>116</sup>.

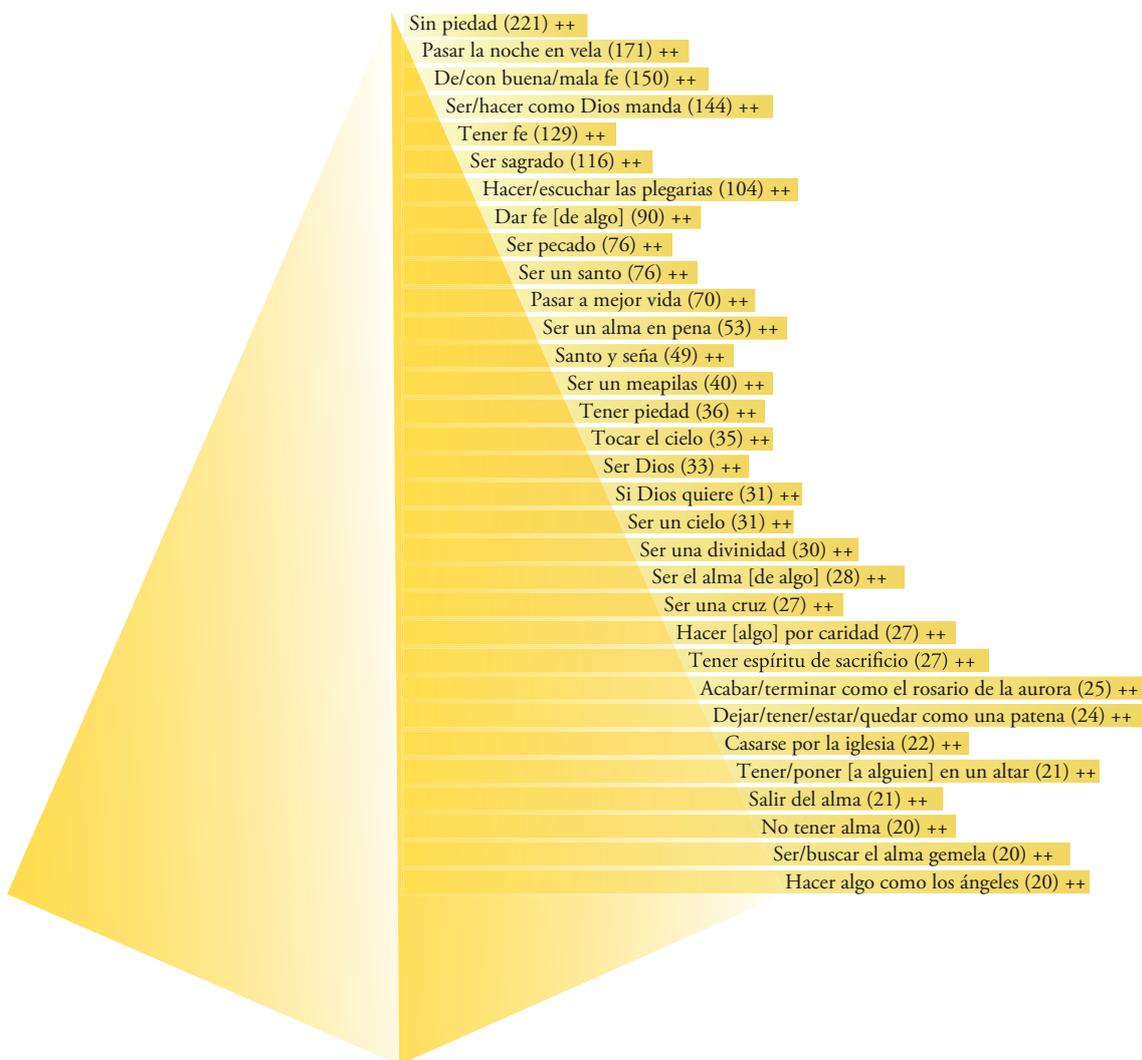
**2) SEGUNDO CRITERIO:** se ha decidido **prescindir de las interjecciones** pese a su frecuencia por considerar que tampoco son demasiado relevantes para extraer información etnolingüística.

**3) TERCER CRITERIO:** se ha decidido **excluir aquellas expresiones que en el CORPES tengan menos de 5 ocurrencias, que no aparezcan en el Val.Es.Co. y que den pocos resultados en la web como corpus**, siempre por su escasa relevancia.

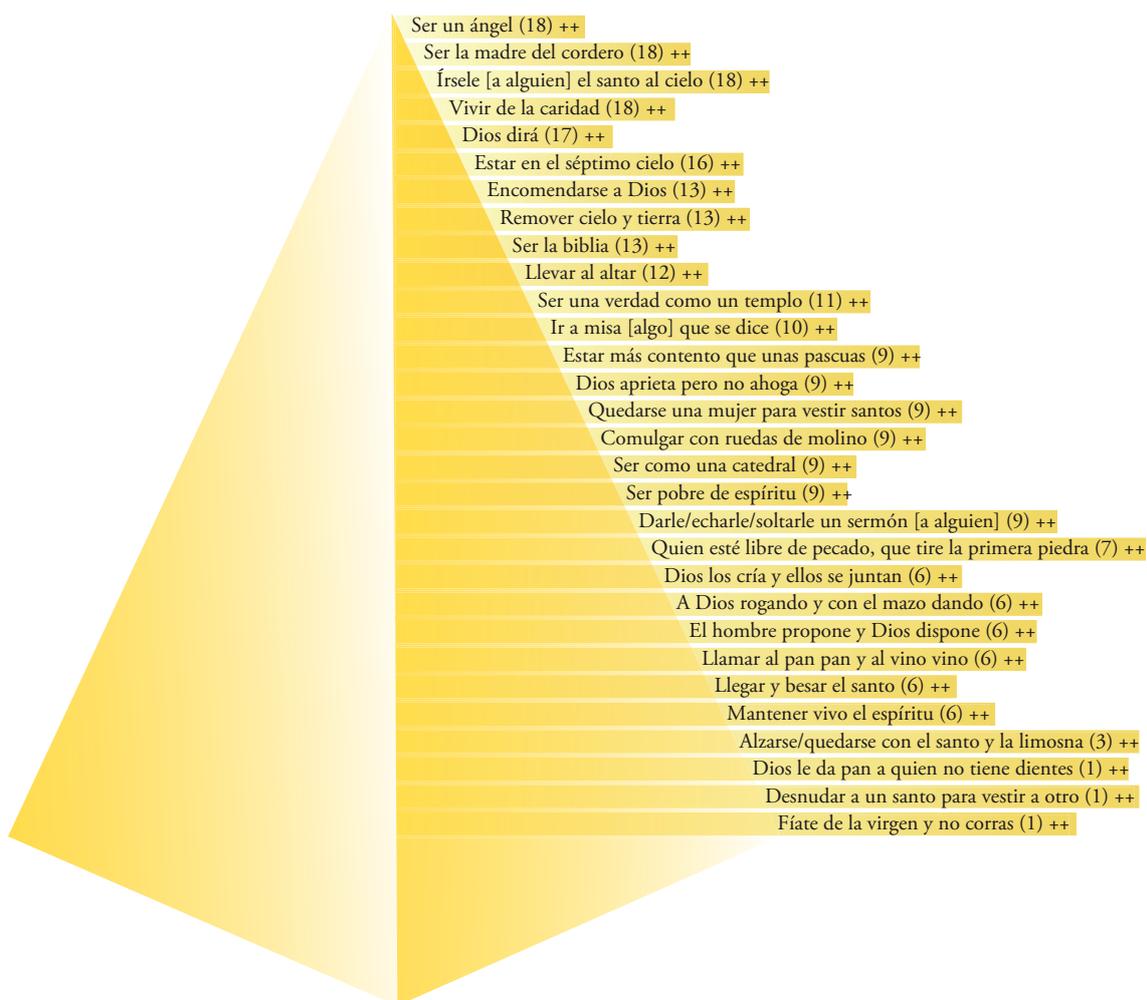
Aplicando estos criterios, se ha decidido repartir las expresiones en tres pirámides para que la visualización sea más clara y ordenada de cara a un análisis contrastivo entre el español y el italiano (objetivo B). En la primera pirámide aparecen las expresiones más rentables y frecuentes, en la segunda las expresiones rentables, pero con una frecuencia inferior a 20 y en la tercera las palabras menos rentables. Teniendo en cuenta estos criterios, así quedarían ordenadas las expresiones<sup>117</sup>:

116. Coincidimos con la teoría de M<sup>a</sup> Amparo Del Campo Martínez (“Los refranes en la enseñanza de español como lengua extranjera”), quien afirma que son vehículos de cultura y que favorecen una mayor inmersión cultural por parte del estudiante.

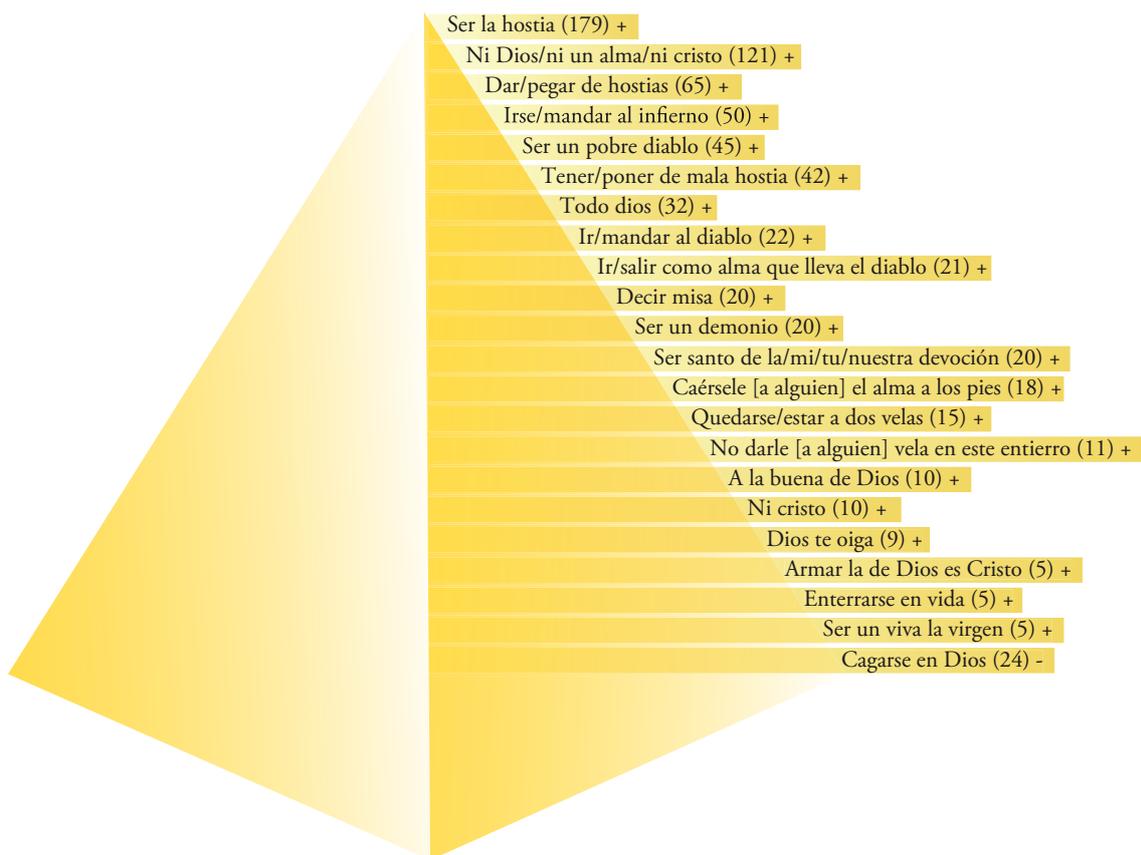
117. Debido al auge de internet, las redes sociales, los memes y el alud informativo al que estamos sometidos a diario, es posible que una expresión varíe su frecuencia y su rentabilidad comunicativa en tan solo unos años, razón por la que se ha decidido enmarcar las pirámides etnolingüísticas en un período limitado de tiempo (2001-2015).



**Gráfico 1. Pirámide 1/3 de la religión en español. Expresiones rentables (++) y frecuentes**



**Gráfico 2.** Pirámide 2/3 de la religión en español. Expresiones rentables (++), pero menos frecuentes



**Gráfico 3. Pirámide 3/3 de la religión en español. Expresiones coloquiales (+) y vulgares (-)**

Somos conscientes de la extensión de estas tres pirámides y de que tal vez se podrían aplicar más criterios para simplificarlas y reducir el tamaño, pero al mismo tiempo pensamos que al hacerlo se perdería algo de información cualitativa, lo que iría en detrimento del segundo objetivo (B) de este proyecto de investigación, alcanzado mediante la visualización de datos en las pirámides. De todos modos, se retomará este concepto en el apartado final dedicado a las aportaciones y limitaciones de la investigación (5.1), así como en las líneas abiertas de investigación (5.3).

#### 4.2.4.2. La pirámide etnolingüística del fútbol en español

A diferencia de la religión, el fútbol es un ámbito cultural que, si bien es muy popular hoy en día y ocupa gran parte del abanico informativo diario, lleva mucho menos tiempo en contacto con la lengua, por eso el número de expresiones es bastante inferior, hecho que facilita la tarea de clasificación.

A continuación, se listan todas las expresiones encontradas con la búsqueda en el CORPES (fase 3 del proceso de creación):

Ser un crack (200)

Ser un equipo (65)

Ganar por goleada (60)  
 Echar/lanzar/despejar balones fuera (54)  
 No dar pie con bola (24)  
 Colgar las botas (19)  
 Devolver la pelota (15)  
 Pillar [a alguien] en fuera de juego (13)  
 Sudar la camiseta (9)  
 Hacerse dueño del partido (9)  
 Dar bola [a alguien] (8)  
 Sacar/mostrar tarjeta roja (8)  
 Meter/colar un gol (6)  
 Ser un buen fichaje (6)  
 Ser el Pelé [de algo] (5)  
 Ir [algo] que chuta (5)  
 Calentar banquillo (3)  
 Casarse de penalti (3)  
 ¡Menudo fichaje! (3)  
 Sacar tarjeta amarilla (2)  
 Hacer un pase a un periodista (1)  
 Haber liga [en algún lugar] (1)  
 Ser el Ronaldo [de algo] (1)

Salta a la vista que el listado es mucho más reducido y tenemos 23 expresiones frente a las 100 que se enumeraron con la religión en un primer momento, una cantidad con la que será relativamente sencillo aplicar los mismos tres criterios que se han aplicado en 4.2.4.1:

**PRIMER CRITERIO:** es el de la **rentabilidad comunicativa** con los tres símbolos que se propusieron en el subapartado anterior (++)/+/(-); sin embargo, en este listado no aparecen expresiones vulgares, por lo que solo se utilizarán las marcas ++ y +<sup>118</sup>.

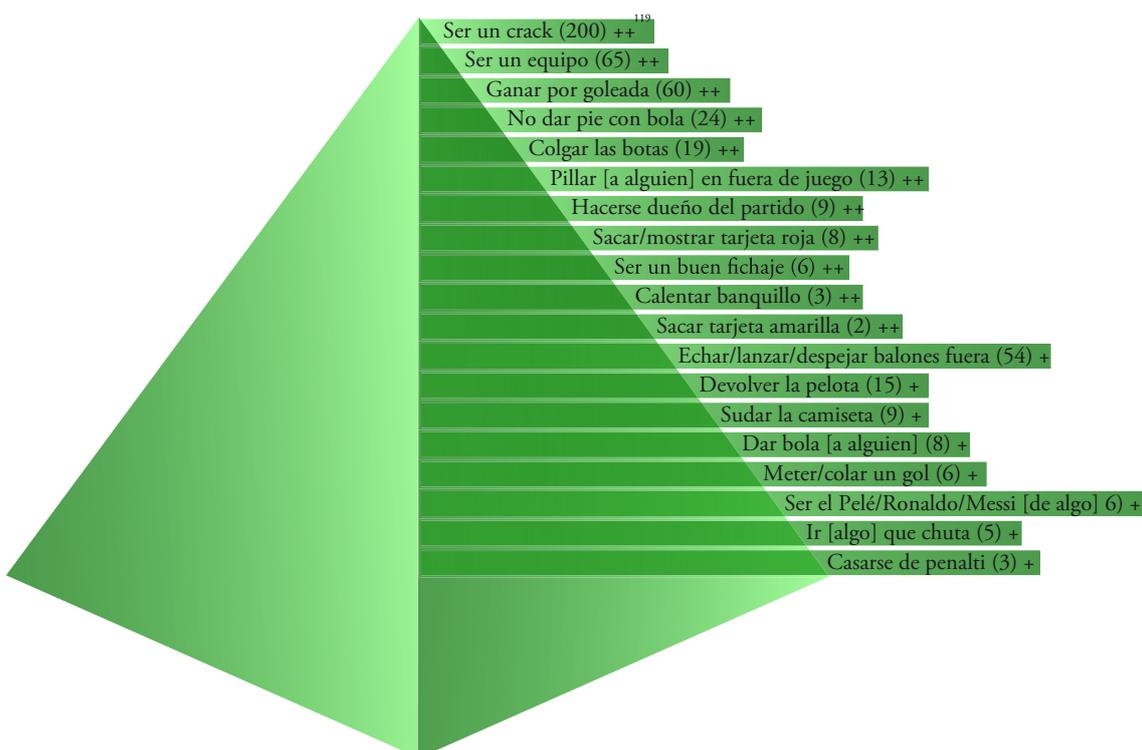
**SEGUNDO CRITERIO:** en el caso de las **interjecciones**, se eliminará *¡Menudo fichaje!* para ser coherentes con el segundo criterio que se aplicó en la pirámide de la religión.

**TERCER CRITERIO:** se decidió **prescindir únicamente de aquellas expresiones que tuviesen una sola ocurrencia y poca presencia en la red** por considerarlas poco relevantes.

Por último, queremos señalar las expresiones ser el *Ronaldo/Pelé [de algo]* a la que se podría añadir más nombres ilustres que de hecho aparecen con frecuencia en la red, como *ser el Mes-*

118. Si el DRAE marca la expresión como *coloq.*

*si [de algo]*, y que se prestan a la popularidad de un jugador en un determinado momento histórico. Para evitar problemas, hemos pensado que sería mejor unificar esta expresión en una sola, ya que al final el significado sigue siendo el mismo, independientemente de quien sea el protagonista de la metáfora. Tras aplicar estos criterios, así es como quedaría la pirámide del fútbol:



#### Gráfico 4. Pirámide 1/1 del fútbol en español

El resultado final es una pirámide con **19 expresiones**, un número suficientemente reducido que ofrece una visualización clara y concisa (tercer paso de la triangulación metodológica) de cómo influye el fútbol en la lengua real, mostrando qué expresiones se utilizan en España en el período seleccionado (2001-2015), lo que corresponde al segundo objetivo (B) de este proyecto de investigación.

Finalmente, es posible que esta pirámide sufra cambios de cara a los próximos años, puesto que el fútbol tiene un impacto social cada vez más fuerte y no es de extrañar que se incorporen a la lengua nuevas expresiones que se vuelvan virales en muy poco tiempo y que incluso reemplacen algunas de las que tenemos ahora. Un ejemplo claro es la expresión **ser un crack**, reina indiscutible de la pirámide. Solo en la web como corpus aparecen más de 17 millones de resultados, y todo apunta a que estos números van a subir a lo largo de los próximos años,

119. No aparece en el DRAE, pero sí con frecuencia en el Val.Es.Co. y sobre todo en la red.

pues se trata de una expresión relativamente joven, pero que ha entrado con fuerza en la lengua. Sería interesante, por lo tanto, realizar otra pirámide dentro de unos años para cotejarla con esta en busca de posibles cambios etnolingüísticos<sup>120</sup>. Una lengua viva, al fin y al cabo, cambia constantemente, y eso es uno de los fundamentos de su riqueza.

#### 4.2.4.3. La pirámide etnolingüística de los toros en español

Para terminar la fase 4, falta visualizar los datos en la pirámide etnolingüística de los toros, un ámbito cultural que hoy en día despierta cierta controversia y que cada vez tiene menos impacto en la sociedad, con una parábola descendiente del todo opuesta a la del fútbol. Pese a ello, tenemos expresiones fosilizadas y de uso común que se siguen utilizando con bastante frecuencia independientemente de que el mundo de los toros esté perdiendo la popularidad que tenía antaño. Esta es la lista de expresiones que extrapolamos del CORPES en la fase 3:

- Dar/poner la puntilla (128)
- Rematar/terminar/acabar la faena (73)
- Coger/tomar/agarrar el toro por los cuernos (51)
- A toro pasado (36)
- Saltarse [algo] a la torera (31)
- Echarle/lanzarle un capote [a alguien] (28)
- Ponerse el mundo por montera (28)
- Salir por la puerta grande (25)
- Hacer un brindis al sol (23)
- Estar para el arrastre (21)
- Cortarse la coleta (12)
- ¡Al toro! (9)
- Salir al ruedo (9)
- Pillar el toro [a alguien] (5)
- Hasta el rabo todo es toro (4)
- Ver los toros desde la barrera (3)
- Tomar la alternativa (2)
- Estar cabreado como un toro de lidia (1)

Aquí tenemos **18 expresiones**, un número muy cercano al listado del fútbol, por lo que la aplicación de los criterios va a ser similar a lo que se vio en 4.2.4.1 y en 4.2.4.2:

**PRIMER CRITERIO:** es el de la **rentabilidad comunicativa** con los tres símbolos que se propusieron en los dos subapartados anteriores (++/+/-); sin embargo, en este listado no aparecen expresiones vulgares, igual que pasó con el fútbol, por lo que solo

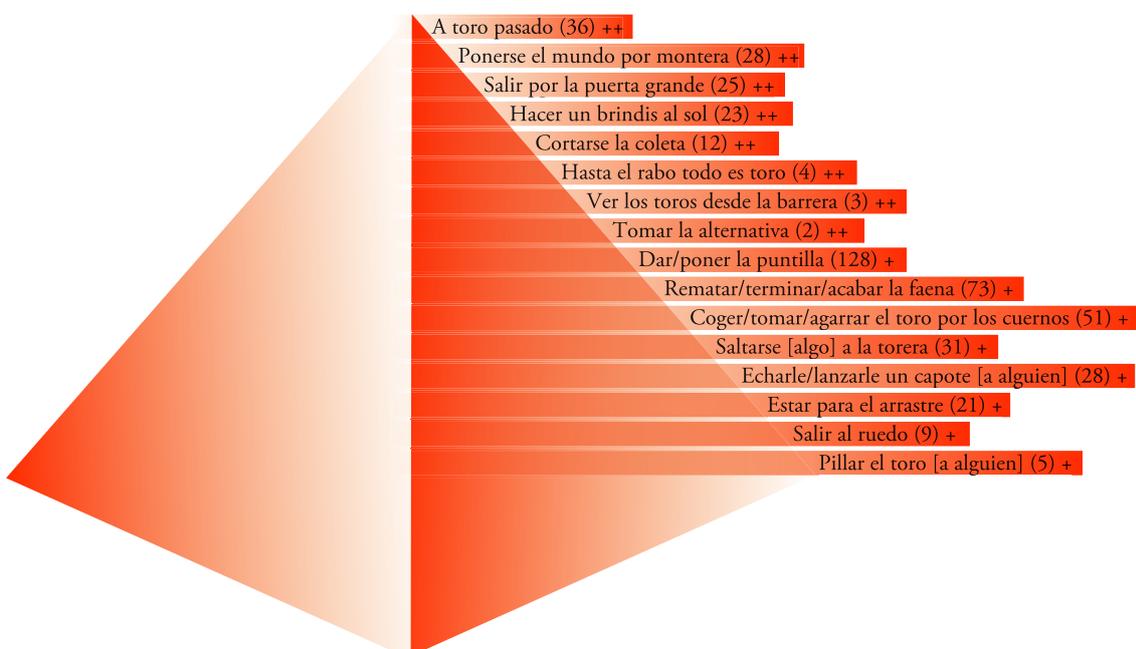
120. Se ampliará esta idea en las conclusiones al analizar las líneas abiertas de investigación en 5.3.

se utilizarán las marcas ++ y +<sup>121</sup>.

**SEGUNDO CRITERIO:** en el caso de las **interjecciones**, se eliminará ¡Al toro! para ser coherentes con el segundo criterio aplicado en las otras dos pirámides.

**TERCER CRITERIO:** tal y como se comentó en 4.2.4.1, se incluirán en la pirámide y se marcarán todos los refranes con ++ por tratarse de expresiones que aportan preciada información etnolingüística, como es el caso de *hasta el rabo todo es toro*.

Por último, se prescindirá de *estar cabreado como un toro de lidia*, la única expresión con un hápax y que por lo tanto se considera poco relevante. Teniendo en cuenta estos criterios, así es como queda la pirámide:



**Gráfico 5. Pirámide 1/1 de los toros en español**

El resultado final es una pirámide con **16 expresiones**, un número suficientemente reducido que ofrece, igual que ocurrió con la pirámide del fútbol, una visualización clara y concisa de cómo influye el ámbito de los toros en la lengua real, y con el que se completa el segundo objetivo (B) de este proyecto de investigación.

Curiosamente, vemos que aquí las expresiones más frecuentes son las más coloquiales; sin embargo, ocupan la parte inferior en la pirámide por ser consideradas menos rentables. Es difícil prever lo que pueda pasar dentro de unos años y queda por ver si los futuros usuarios de la lengua seguirán utilizando estas expresiones pese a la constante decadencia de la tau-

121. Si el DRAE marca la expresión como coloq.

romaquia y al escaso interés que despierta en la sociedad actual. Algunas de ellas están muy fosilizadas en la lengua y es posible que sigan existiendo, aunque su permanencia dependerá en muchas ocasiones de factores impredecibles como los memes, las modas y las tendencias de los hablantes.

A partir de aquí, será interesante cotejar estas pirámides del español con las pirámides correspondientes del italiano para demostrar, a partir de la representación visual de datos, que la cultura influye de manera diferente en cada lengua, incluso con ámbitos culturales afines (objetivo C). En el caso concreto de los toros, un ámbito cultural prácticamente inexistente en Italia, se verá claramente la gran divergencia que existe entre las dos lenguas<sup>122</sup>, pero incluso los ámbitos culturales afines depararán más de una sorpresa, como se verá en 4.4.

---

122. Esta es una de nuestras hipótesis de partida (1.2) que se aspira a demostrar mediante el presente trabajo.

### 4.3. EL PROCESO DE CREACIÓN DE LAS PIRÁMIDES ETNOLINGÜÍSTICAS: TRES EJEMPLOS PARA EL ITALIANO

#### 4.3.1. FASE 1: EL *BRAINSTORMING*

El *brainstorming* en italiano se realizó el miércoles 14 de diciembre de 2016 a las 16.00h en un despacho cercano a la Universitat de Barcelona con tres profesoras nativas que se dedican a la enseñanza de la lengua y cultura italiana desde hace muchos años: Ilaria Bada, Antonietta Vinciguerra y Ornella Bernardi, a quienes agradezco su paciencia y disponibilidad<sup>123</sup>. Al día siguiente, se realizó la prueba con dos informantes nativos más residentes en Italia y sin conocimientos de la lengua española. Esto se hizo principalmente por dos razones: en primer lugar, lenguas como el inglés y el español presentan una extensión geográfica muy amplia y, con la llegada de internet y de las redes sociales, su interferencia con lenguas menos habladas es aún mayor; en segundo lugar, las tres profesoras de italiano residen desde hace tiempo en Barcelona, y es posible que su disponibilidad léxica tenga alguna influencia del catalán y/o del castellano, por eso se quiso equilibrar el *brainstorming* con dos informantes nativos de nivel sociocultural alto que solo hablan italiano y que además residen en Italia. El *brainstorming* duró alrededor de veinte minutos y las directrices fueron las mismas que para la prueba en español (4.2.1). En este caso, cinco personas también fueron suficientes para obtener un buen número de palabras que se analizarán a continuación:

##### 4.3.1.1. Listado obtenido tras la Fase 1 para el ámbito de la religión

## A

Abito  
Aborto  
Abside  
Acquasanta  
Allah  
Apostoli  
Agnostici  
Albero  
Anello  
Anima  
Assisi  
Atei  
Avemaria

## B

Bambini  
Battesimo  
Bestemmia  
Bianco  
Bibbia  
Bigotti  
Birre  
Bisogno  
Blasfemia  
Bomboniere  
Buddha

## C

Calvi  
Candela  
Canto  
Cappella Sistina  
Catechismo  
Cattedrale  
Cera  
Chiesa  
Chitarra  
Cielo  
Cimitero  
Clausura  
Colpa  
Colombi  
Compassione  
Comunione  
Conchiglia  
Conventi  
Corale  
Corpo  
Cottolengo  
Credenti  
Credenza  
Cresima  
Cristo  
Croce  
Crociate  
Crocifisso  
Curia

123. Se indican sus nombres porque contamos con su consentimiento para ello.

**D**

Diavolo  
Dieci comandamenti  
Dio  
Dolci  
Domenica

**E**

Ebraismo  
Elezioni  
Eresia

**F**

Famiglia  
Fazzoletto  
Fede  
Fedeli  
Feverbachs  
Fonte  
Freddo

**G**

Gerusalemme  
Gesù  
Gesuiti  
Giallo  
Ginzburg  
Giuseppe  
Gotico  
Guerra

**I**

Ideología  
Impressionante  
Impuro  
Incenso  
Indulgenza  
Inginocchiarsi  
Inginocchiatoio  
Inquisizione  
Irrazionalità  
Islam  
Israele

**L**

La Mecca  
Lenzuolo  
Lino  
Liquori  
Lutero

**M**

Madonna  
Marmo  
Martirio  
Marx  
Matrimonio  
Menzogna  
Messa  
Mitologia  
Monaca  
Monasteri  
Monoteismo  
Moschea

**N**

Natale  
Navata

**O**

Oferta  
Omosessualità  
Oppio  
Oppressione  
Oratorio  
Órgano  
Ostia

**P**

Pace  
Padrenostro  
Paganesimo  
Palestina  
Pane  
Papa  
Papalina  
Pasqua  
Pedofilia  
Pellegrini  
Pentimento  
Persecuzione  
Pianto  
Pietà  
Portone  
Poveri

**R**

Raccoglimento  
Razze  
Reclusione  
Regali  
Ripetere  
Ritiro

**S**

Sandali  
Santino  
Santo  
Santiago  
Scalinata  
Scultura  
Scouts  
Sindone  
Soldi  
Suore  
Superstizione

**T**

Trascendenza  
Trinità

**V**

Vangelo  
Vaticano  
Villaggi  
Vino  
Viola  
Voto

**4.3.1.2. Listado obtenido tras la Fase 1 para el ámbito del fútbol****A**

Abbonamento  
Adidas  
Adolescenza  
Aereo  
Africa  
Ala  
Allenamento  
Arbitro  
Attaccante

**B**

Bandiera  
Balotelli  
Barça  
Berlusconi  
Bianconero  
Birra  
Bomba  
Bomber  
Buffon

**C**

Calciatori  
Calciopoli  
Calzetti  
Calzettoni  
Campetto  
Campionato  
Campo  
Canalis  
Canne  
Canzone  
Capitano  
Cartellino giallo  
Cartellino rosso

Classifica  
Clooney  
Cocaina  
Colori  
Cori  
Corner  
Corruzione  
Cronista  
Cronometro  
Curva

**D**

Dallara  
D'Amico  
De Gregori  
Denaro  
Derby  
Diadora  
Distrazione  
Divisa  
Dodici

**E**

Erba  
Espulsi  
Europei

**F**

Falli  
FIFA  
Finte  
Fischietto  
Fischio  
Fuorigioco

**G**

Giocatori  
Gioco  
Gol  
Gradinata  
Gridare  
Gullit

Domenica

Donne

Doping

Dribblare

## H

Hooligans

## I

Inghilterra

Intervista

## J

Juve

## L

Lacci

Libro

Litigare

Lusso

## M

Maglie

Mai dire Gol

Maschilismo

Messi

Match

Miliardi

Mister

Moggi

Mondiali

Moviola

Muscoli

## N

Nazionale

Nike

## O

Oppio

## P

Palla

Pallone

Palo

Pantaloncini

Partita

Passione

Pattine

Perone

Piqué

Porta

Portiere

Prato

Pubblico

Pulcini

Puzza

## Q

Quadruplicite

## R

Radiolina

Ragazzi

Razzismo

Resistenza

Rete

Rigore

## S

Scarpe

Schedina

Sciarpa

Scudetto

Scuola Calcio

Sedili

Sfida

Sgambetto

Shakira

Sky

Soldi

Sport

Sputare

## T

Tacchetti

Tecnica

Televisione

Tifare

Tifo

Toro

Trasmettere

Traversa

Trofeo

**U**

UEFA  
Uomini

**V**

Veline  
Verde  
Vieri  
Violenza

**W**

Waka-waka  
Weekend

Squadra  
Stadio  
Stampa  
Stopper  
Stress  
Striscioni  
Stupidità  
Sudore  
Supplementari  
Supporters

**4.3.1.3. Listado obtenido tras la Fase 1 para el ámbito de los toros****A**

Allevamento  
Almodovar  
Alcool  
Andalusia  
Anello  
Animalismo  
Animalisti  
Anziani  
Arena  
Argento

**B**

Barcelona

**C**

Calamari fritti  
Cappellino  
Carica  
Cattiveria  
Cibo  
Conservatore  
Coraggio  
Corna  
Cornata  
Correre  
Corrida  
Costume  
Crudeltà

**D**

Destra  
Dolore

**E**

El Cordobés

**F**

Fama  
Fazzoletto  
Ferire  
Ferite  
Fieno  
Forza

**G**

García Lorca  
Giallo  
Giornale  
Gradinata  
Granata  
Grida

**H**

Hemingway

**I**

Incornata  
Incornare

**O**

Olé  
Omicidio  
Orecchie  
Osborne

**S**

Sabbia  
Sangue  
Sbuffare  
Seviglia  
Soldi  
Spade  
Spagna  
Spavaldo  
Spettacolo  
Staccionata  
Stalle  
Sterco  
Stupidità  
Stupro  
Sudore  
Superga

Francia  
Freccia  
Fuoco

**L**

Labirinto

**P**

Pamplona  
Pezzato  
Picadores  
Puzza

**T**

Terra  
Testosteroni  
Torero  
Tori  
Torino  
Tortura  
Tradizione

**Z**

Zoccolo

**M**

Malvagità  
Mandria  
Manifesto  
Mansioni  
Mantella  
Miguel Bosé  
Minosse  
Minotauro  
Molestia sessuale  
Momento della verità  
Morte  
Mucca  
Muleta

**R**

Rischio  
Rosa  
Rosso

**U**

Ubriachi  
Uccidere

**N**

Nero

**V**

Violenza

Al cotejar este *brainstorming* con el equivalente que se realizó en español en 4.2.1, se puede apreciar que el número de palabras que aportaron los informantes es bastante parejo, y también en este caso hay muchos vocablos que están relacionados de manera indirecta con el tema en cuestión y que se deben a la cosmovisión personal de cada uno y al sistema de creencias que hemos desarrollado en nuestro subconsciente. En el caso de los toros, los informantes italianos han proporcionado muchas más palabras indirectas al tratarse de un ámbito cultural poco desarrollado en Italia; lo cual influirá, según nuestra hipótesis de partida (1.2), en el número de expresiones que se encontrarán con la búsqueda en los corpus (fase 3). En 4.3.2 se someterán estos listados a la **criba** para ver qué palabras seleccionamos y por qué.

### 4.3.2. FASE 2: LA OBTENCIÓN DEL LISTADO DE TRABAJO: CRIBA, METODOLOGÍA Y CRITERIOS

Tal como se hizo con el español en 4.2.2, es menester aplicar una serie de criterios para seleccionar aquellas palabras del *brainstorming* que más se ajusten a nuestro propósito y que permitan alcanzar el primer objetivo (A) a partir de la búsqueda de expresiones en los corpus (fase 3 del proceso de creación). Para que al análisis contrastivo final sea lo más objetivo posible, y como el resultado del *brainstorming* presenta características similares al español (nombres propios de personas, topónimos, nombres que no tienen una relación directa con el tema propuesto, etc.), se ha decidido mantener los mismos criterios propuestos en 4.2.2, de modo que:

- 1) **Se excluirán aquellos nombres propios de personas que no estén directamente relacionados con la temática**, como *Feverbachs*, *Ginzburg*, *Giuseppe*, *Lutero*, *Maria* y *Marx* (religión); *Berlusconi*, *Canalis*, *Clooney*, *D'Amico*, *De Gregori* y *Shakira* (fútbol); *Almodovar*, *García Lorca*, *Hemingway* y *Miguel Bosé* (toros).
- 2) **Se excluirán los topónimos y los nombres de continentes, países y regiones por su falta de extensión**, como *Assisi*, *Betlemme*, *Gerusalemme*, *Israele*, *La Mecca*, *Palestina*, *Santiago* y *Vaticano* (religión); *Africa*, *Andalusia*, *Barcellona*, *Francia*, *Inghilterra*, *Pamplona*, *Seviglia*, *Spagna* y *Torino*<sup>124</sup> (fútbol y toros).
- 3) **Se excluirán los nombres de las iglesias o instituciones, de los estadios y de las marcas (siempre por su falta de extensión)**, como *Cappella Sistina* y *Cottolengo* (religión); *Adidas*, *Dallara*, *Diadora* y *Nike* (fútbol); *Osborne* y *Superga* (toros).
- 4) **En el caso concreto del fútbol, se prescindirá de cualquier nombre de programa o servicio televisivo, de los nombres de clubes y de las federaciones (falta de**

124. Podría referirse también al nombre del equipo de fútbol, pero se descartaría de todas formas con el tercer criterio.

**extensión**), como *Barça, Calciopoli*<sup>125</sup>, *FIFA, Mai dire gol, Sky y UEFA*.

Por lo que se refiere a los **sustantivos, verbos y adjetivos**, se seguirán aplicando los mismos criterios del proceso de creación de las pirámides que se vieron en 4.2.2:

1) Con los sustantivos o nombre comunes, **se excluirán aquellos que no guarden una relación directa con el ámbito cultural y cuya aparición en el brainstorming se deba a factores emotivos, personales, etc.** Así pues, se prescindió de *abito, aborto, albero, anello, bambini, birre, bisogno, bomboniere, calvi, chitarra, colombi, conchiglia, dolci, elezioni, familia, fazzoletto, fonte, freddo, giallo, gotico, guerra, irrazionalità, lenzuolo, lino, liquori, marmo, menzogna, offerta, omosessualità, oppio, oppressione, organo, papalina, pedofilia, persecuzione, portone, poveri, razzie, regali, ripetere, sandali, scalinata, scultura, scouts, soldi, trascendenza, villaggi y viola (religión); abbonamento, adolescenza, aereo, birra, bomba, canne, canzone, cocaina, colori, corruzione, cronometro, denaro, distrazione, divisa, donne, erba, gioco, lusso, maschilismo, miliardi, muscoli, oppio, passione, pattine, perone, puzza, quadricipite, ragazzi, sciarpa, soldi, sport, sputare, stampa, stress, stupidità, sudore, televisione, trasmettere, uomini, veline, violenza, wakawaka y weekend (fútbol); allevamento, alcool, anello, animalismo, animalisti, anziani, argento, calamari fritti, cappellino, catteriveria, cibo, coraggio, costume, crudeltà, destra, dolore, fama, fazzoletto, ferite, fieno, forza, freccia, fuoco, giornale, grida, labirinto, machismo, macho, malvagità, mandria, manifesto, mansioni, mantella, olé, omicidio, puzza, rischio, sabbia, soldi, spade, spettacolo, stalle, sterco, stupidità, stupro, sudore, terra, testosteroni, tortura, tradizione, ubriachi y violenza (toros).*

2) Respecto a los verbos, **se aplicará un criterio muy similar al anterior, aunque su presencia en el brainstorming es mínima.** Se han quitado, por su falta de relación directa, *inginocchiarsi* (religión), *gridare* (fútbol), *correre, ferire, sbuffare y uccidere* (toros).

3) Con los adjetivos pasa algo muy similar y su presencia es mínima. Se prescindirá de *bianco, corale e impressionante* (religión), *verde* (fútbol), *conservatore, giallo, granata, pezzato, spavaldo, rosa e rosso* (toros).

4) Respecto a los compuestos pluriverbales, se han incluido *dieci comandamenti* (religión), *cartellino giallo, cartellino rosso y scuola calcio* (fútbol). Se ha decidido excluir *momento della verità* (toros) por considerar que su relación con el ámbito cultural propuesto es indirecta y poco relevante.

Tras aplicar estos criterios, así es como quedarían los tres listados.

125. Es el nombre de un proceso que se llevó a cabo en 2006 contra la corrupción de algunos clubes de la Serie A y que afectó principalmente a la Juventus F. C. No nos parece relevante y por eso decidimos excluirlo del listado definitivo.

### 4.3.2.1. Listado definitivo tras la Fase 2 para el ámbito de la religión

## A

Abside  
Acquasanta  
Allah  
Apostoli  
Agnostici  
Anima  
Avemaria

## B

Battesimo  
Bestemmia  
Bibbia  
Bigotti  
Blasfemia  
Buddha

## C

Candela  
Canto  
Catechismo  
Cattedrale  
Cera  
Chiesa  
Cielo  
Cimitero  
Clausura  
Colpa  
Compassione  
Comunione

Conventi  
Corpo  
Credenti  
Credenza  
Cresima  
Cristo  
Croce  
Crociate  
Crocifisso  
Curia

## D

Diavolo  
Dieci comandamenti  
Dio  
Domenica

## E

Ebraismo  
Eresia

## F

Famiglia  
Fede  
Fedeli

## G

Gesù  
Gesuiti

## I

Ideologia  
Impuro  
Incenso  
Indulgenza  
Inginocchiatoio  
Inquisizione  
Islam

## L

Lutero

## M

Madonna  
Martirio  
Matrimonio  
Messa  
Mitologia  
Monaca  
Monasteri  
Monoteismo  
Moschea

## N

Natale  
Navata

## O

Oratorio  
Ostia

## P

Pace  
Padrenostro  
Paganesimo  
Pane  
Papa  
Pasqua  
Pellegrini

## R

Raccoglimento  
Reclusione  
Ritiro

## S

Santino  
Santo  
Sindone  
Suore  
Superstizione

	Pentimento
	Pianto
	Pietà
	Pregare
	Preghiera
	Presepe
	Preti
	Profeta
	Pulpito
<b>T</b>	<b>V</b>
Trinità	Vangelo
	Vaticano
	Vino
	Voto

#### 4.3.2.2. Listado definitivo tras la Fase 2 para el ámbito del fútbol

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
Ala	Bandiera	Calciatori	Derby
Allenamento	Balotelli	Calzetti	Divisa
Arbitro	Bianconero	Calzettoni	Domenica
Attaccante	Bomber	Campetto	Dribblare
	Buffon	Campionato	
		Campo	
		Capitano	
		Cartellino giallo	
		Cartellino rosso	
		Classifica	
		Cori	
		Corner	
		Cronista	
		Curva	
<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>
Espulsi	Falli	Giocatori	Hooligans
Europei	Finte	Gioco	
	Fischietto	Gol	
	Fischio	Gradinata	
	Fuorigioco	Gullit	

**J**

Juve

**L**

Lacci

**M**

Maglie

Messi

Mondiali

Moviola

**N**

Nazionale

**P**

Palla

Pallone

Palo

Pantaloncini

Partita

Piqué

Porta

Portiere

Prato

Pulcini

**R**

Radiolina

Rete

Rigore

**S**

Scarpe

Schedina

Scudetto

Scuola calcio

Sfida

Sgambetto

Squadra

Stadio

Stopper

Striscioni

Supplementari

Supporters

**T**

Tacchetti

Tecnica

Tifare

Tiro

Toro

Traversa

Trofeo

**V**

Vieri

**4.3.2.3. Listado definitivo tras la Fase 2 para el ámbito de los toros****A**

Arena

**C**

Carica

Corna

Cornata

Corrida

**E**

El Cordobés

**I**

Incornata

Incornare

**M**

Minotauro

Morte

Muleta

**O**

Orecchie

**P**

Picadores

**S**

Sangue

Staccionata

**T**

Torero

Tori

**Z**

Zoccolo

Si se cotejan estos resultados con los listados del español tras la criba (4.2.2), se puede apreciar que el número de palabras de las que disponemos para buscar en el corpus (fase 3 del proceso de creación) es bastante parejo: **125 del español frente a 93 del italiano para la religión, 79 frente a 84 para el fútbol y 53 frente a 19 para los toros**. Donde se aprecia la mayor diferencia es en el apartado relacionado con el mundo de los toros, un ámbito cultural mucho menos desarrollado en Italia y por consiguiente con una disponibilidad léxica muy inferior<sup>126</sup>.

Como se vio en el apartado 4.2.2, cada *brainstorming* es diferente, pero confiamos en que estos tres listados definitivos nos resulten de utilidad para encontrar el mayor número posible de expresiones y alcanzar el primer objetivo (A) con el italiano.

A continuación, se procederá a introducir todas las palabras de los listados definitivos en los corpus.

### 4.3.3. FASE 3: LA BÚSQUEDA EN LOS CORPUS: DE LAS OCURRENCIAS A LAS EXPRESIONES

Para poder realizar las pirámides etnolingüísticas en italiano, se seguirá el proceso de creación que en 4.2 se ha ejemplificado con el español y se buscarán las palabras obtenidas con la criba (fase 2) en el CORIS/CODIS, apoyándose también en el LIP y en la red como corpus. La diferencia principal con el español estriba en que el CORPES también incluye transcripciones orales, mientras que el CORIS/CODIS solo está centrado en el italiano escrito, como se vio en 3.3.1, por lo que se ha decidido realizar todas las búsquedas también en el LIP para subsanar esta pequeña diferencia. Cabe señalar que para el italiano también se ha optado por seguir el orden alfabético de cada listado, ya que resulta más sencillo. Así pues, hay que acceder al CORIS/CODIS, el corpus de referencia que se ha seleccionado para el italiano (3.3.1), e introducir las palabras en el motor de búsqueda marcándolas entre comillas. En las páginas siguientes se recogen una por una las palabras del listado definitivo para el ámbito cultural de la religión, el fútbol y los toros en busca del mayor número posible de expresiones pluriverbales (objetivo A).

#### 4.3.3.1. Fase 3 con el listado de la religión en italiano

Al igual que en España, la religión es un ámbito cultural muy arraigado en Italia desde hace más de dos mil años, y eso tiene unas repercusiones lingüísticas en la lengua que deberían verse reflejadas en un gran número de expresiones, en principio mucho mayor respecto al fútbol y sobre todo a los toros<sup>127</sup>. Hay que señalar que el CORIS/CODIS no ofrece la posibilidad de escoger entre *lema y forma* (el corpus no está lematizado), por lo que hemos tenido que buscar

126. Véase hipótesis de partida en 1.2.

127. Vale la pena subrayar, una vez más, que tanto el CORPES como el CORIS/CODIS son corpus que se van actualizando constantemente, y estas búsquedas se realizaron a finales de 2016, por lo que es posible que en la actualidad el número de ocurrencias haya variado.

las palabras en todas sus flexiones<sup>128</sup>. Se ha buscado siempre por *forma* y este es el resultado de la búsqueda:

## A

**Abside:** 112 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Acquasanta:** 9 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *essere come il diavolo e l'acquasanta* (3 ocurrencias), para referirse a dos cosas incompatibles entre ellas.

**Allah:** 694 ocurrencias en el CORIS<sup>129</sup> y 0 en el LIP. Ninguna expresión en los corpus.

**Apostolo:** 816 ocurrencias en el CORIS y 9 en el LIP. Ninguna expresión en los corpus.

**Agnostico:** 84 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Anima:** 10.028 ocurrencias en el CORIS y 6 en el LIP. Destacan en el CORIS las expresiones *non esserci anima viva* (185 ocurrencias, con significado de “nadie”), *essere/trovare l'anima gemella* (126 ocurrencias, para referirse a la persona que más se parece a nosotros mismos y que nos complementa), *essere come un'anima in pena* (40 ocurrencias, para referirse a alguien inquieto/atormentado), *vendere/cedere l'anima al diavolo* (27 ocurrencias, para referirse a alguien que está dispuesto a todo con tal de conseguir sus objetivos) y *raccomandare/rendere l'anima a Dio* (14 ocurrencias en el CORIS, con significado de “morir”).

**Ateo:** 325 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Avemaria:** 10 ocurrencias en el CORIS y 0 en el LIP. Aparece en el corpus la expresión *essere innocente come un'avemaria* (1 ocurrencia).

## B

**Battesimo:** 1.116 ocurrencias en el CORIS y 5 en el LIP. Destaca en los corpus la expresión *tenere a battesimo* (128 ocurrencias en el CORIS y 1 en el LIP) con significado de “hacer de padrino o de inaugurar algo”.

**Bestemmia:** 551 ocurrencias en el CORIS y 0 en el LIP. Destaca en el corpus la expresión *essere una bestemmia* (90 ocurrencias), para referirse a algo disparatado y fuera de lugar.

**Bibbia:** 135 ocurrencias en el CORIS y 1 en el LIP. Destaca en el corpus la expresión *essere/diventare la Bibbia* (40 ocurrencias en el CORIS), con el mismo significado que la expresión española (4.2.3.1).

**Bigotto:** 217 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en el corpus.

**Blasfemia:** 29 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en el corpus.

**Buddha:** 465 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en el corpus.

## C

**Candela:** 1.811 ocurrencias en el CORIS y 7 en el LIP. Destaca la expresión *consumarsi/spengersi come una candela* (10 ocurrencias en el CORIS, con significado de “consumirse

128. Volveremos sobre este tema en 5.1, dedicado a las aportaciones y limitaciones de esta tesis.

129. El sistema de búsqueda distingue entre mayúsculas y minúsculas, por lo que se buscarán los nombres propios tanto con minúscula como con mayúscula inicial.

rápidamente”, en alusión a las velas que se encienden en las iglesias)<sup>130</sup>.

**Canto:** 6.132 ocurrencias en el CORIS y 12 en el LIP. Ninguna expresión en los corpus relacionada con la religión.

**Catechismo:** 335 ocurrencias en el CORIS y 1 en el LIP. Ninguna expresión en los corpus.

**Cattedrale:** 1.361 ocurrencias en el CORIS y 0 en el LIP. Destaca la expresión *come una cattedrale* (7 ocurrencias en el CORIS), con el mismo significado que en español (4.2.3.1).

**Cera:** 1.094 ocurrencias en el CORIS y 3 en el LIP. Se encontraron en el CORIS 28 ocurrencias de la expresión *avere una brutta cera*, pero se descartó por no tener relación con la religión.

**Chiesa:** 8.780 ocurrencias en el CORIS y 31 en el LIP<sup>131</sup>. Se encontró en el corpus la expresión *essere tutto casa e chiesa* (3 ocurrencias en el CORIS) para referirse a una persona muy religiosa. A veces se utiliza de manera irónica para referirse a un falso devoto.

**Cielo:** 14.534 ocurrencias en el CORIS y 21 en el LIP. Se encontraron en los corpus algunas interjecciones muy frecuentes en la lengua como *santo cielo!* (210 ocurrencias en el CORIS y 1 en el LIP), *per l'amor del cielo!* (80 ocurrencias en el CORIS y 2 en el LIP) y *grazie al cielo!* (19 ocurrencias en el CORIS y 2 en el LIP). Como expresión, destaca *essere al settimo cielo* (82 ocurrencias en el CORIS) y *non stare né in cielo né in terra* (1 ocurrencia en el CORIS, para referirse a algo absurdo)<sup>132</sup>.

**Cimitero:** 2.501 ocurrencias en el CORIS y 9 en el LIP. Destaca en los corpus la expresión *essere/esserci un cimitero [di qualcosa]* (22 ocurrencias en el CORIS y 1 en el LIP) para indicar abundancia de algo en algún lugar, a menudo con una connotación despectiva.

**Clausura:** 231 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Colpa:** 10.279 ocurrencias en el CORIS y 22 en el LIP. Ninguna expresión en los corpus relacionada con la religión.

**Compassione:** 1.046 ocurrencias en el CORIS y 0 en el LIP. Destacan en el corpus las expresiones *avere compassione* (82 ocurrencias, con significado de “apiadarse de alguien/algo”) y *fare compassione* (15 ocurrencias, con significado de “dar pena”).

**Comunione:** 1.541 ocurrencias en el CORIS y 14 en el LIP. Ninguna expresión en los corpus.

**Convento:** 1.764 ocurrencias en el CORIS y 1 en el LIP. Destacan las expresiones *accontentarsi con quello che passa il convento* (17 ocurrencias en el CORIS) e *il convento non passa altro* (1 ocurrencia en el CORIS), ambas para indicar que hay que contentarse con lo que se tiene.

**Corpo:** 34.887 ocurrencias en el CORIS y 42 en el LIP. Destaca la expresión *avere il*

130. La búsqueda en el CORIS proporcionó otras expresiones como *reggere la candela* o *il gioco non vale la candela*, pero no se tendrán en consideración porque no guardan ninguna relación con la religión.

131. Cabe subrayar que el CORIS, al no estar lematizado, no distingue las categorías gramaticales, por lo que confunde el sustantivo plural *chiese* con el homógrafo del pretérito perfecto simple de tercera persona singular del verbo *chiedere*, lo que entorpece bastante la búsqueda. Se volverá sobre ello en 5.1.

132. Esta última no se tendrá en cuenta por su escasa frecuencia y por no guardar una relación clara con el ámbito religioso.

*diavolo in corpo* (12 ocurrencias en el CORIS) para referirse a alguien muy inquieto.

**Credente:** 1.268 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Credenza:** 1.408 ocurrencias en el CORIS y 4 en el LIP. Ninguna expresión en los corpus.

**Cresima:** 65 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Cristo:** 8.383 ocurrencias en el CORIS y 64 en el LIP (un número muy inferior respecto a la búsqueda en español)<sup>133</sup>. Destacan en los corpus las expresiones *essere un povero cristo* (96 ocurrencias en el CORIS, para referirse a una persona desamparada), *essere un bravo cristo* (5 ocurrencias en el CORIS, con significado de “buena persona”) y *non esserci cristoli* (3 ocurrencias en el CORIS, con significado de “no haber manera de hacer algo”). Las interjecciones *Cristo!* y *Cristo santo!* aparecen respectivamente en 140 y en 60 ocasiones.

**Croce:** 2.786 ocurrencias en el CORIS y 12 en el LIP. Destacan las expresiones *mettere in croce* (44 ocurrencias en el CORIS, con significado de “atormentar/molestar a alguien”), *dare/buttare/gettare la croce addosso [a qualcuno]* (8 ocurrencias en el CORIS y 1 en el LIP, con significado de “atribuir toda la culpa a alguien”) y *ciascuno ha/porta la sua croce* (5 ocurrencias en el CORIS, para indicar que cada uno tiene algún tipo de problema).

**Crociata:** 817 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Crocifisso:** 511 ocurrencias en el CORIS y 1 en el LIP<sup>134</sup>. Ninguna expresión en los corpus.

**Curia:** 238 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## D

**Diavolo:** 3.789 ocurrencias en el CORIS y 8 en el LIP<sup>135</sup>. Además de las interjecciones *chel/cosa diavolo!* (respectivamente 168 y 166 ocurrencias en el CORIS), destacan las expresiones *mandare al diavolo* (45 ocurrencias en el CORIS y 1 en el LIP), *fare l'avvocato del diavolo* (35 ocurrencias en el CORIS, utilizada cuando se desmonta una opinión ajena tratando de demostrar su inconsistencia), *avere un diavolo per capello* (23 ocurrencias en el CORIS, con significado de “estar de muy mal humor”), *saperne una più del diavolo* (17 ocurrencias en el CORIS, para indicar a alguien muy astuto) y *essere come il diavolo e l'acquasanta*, que ya se vio en la entrada *acquasanta*.

**Dieci comandamenti:** 64 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en el corpus.

**Dio:** 27.208 ocurrencias en el CORIS y 146 en el LIP. Destacan en los corpus las interjecciones *Dio mio!/mio Dio!* (777 ocurrencias en el CORIS y 0 en el LIP), *per l'amor di Dio!* (182 ocurrencias en el CORIS y 2 en el LIP), *santo Dio!/Dio santo!* (151 ocurrencias en el CORIS y 0 en el LIP), *ah/oh Dio!* (89 ocurrencias en el CORIS y 18 en el LIP), *se Dio vuole!* (47 ocurrencias en el CORIS y 0 en el LIP), *grazie a Dio!* (28 ocurrencias en el CORIS y 2 en el LIP), *come /quanto è vero Dio!* (8 ocurrencias en el CORIS y 0 en el LIP,

133. Véase 4.2.3.1.

134. Incluyendo el participio pasado del verbo *crocifiggere*.

135. Como curiosidad, vale la pena comentar que esta palabra aparece casi siempre en minúscula (más de 3.300 casos), a diferencia de la palabra *Dio*, con la que ocurre exactamente lo contrario.

para enfatizar una amenaza o una afirmación), *che Dio ce la mandi buona!* (6 ocurrencias en el CORIS y 1 en el LIP), *Dio canel!* (6 ocurrencias en el CORIS y 0 en el LIP) y *per carità di Dio!* (0 ocurrencias en el CORIS y 2 en el LIP). En cuanto a las expresiones, encontramos *ogni/tutto questo ben di Dio* (114 ocurrencias en el CORIS y 0 en el LIP, para referirse a cosas importantes y/o de valor), *Dio solo sa* (74 ocurrencias en el CORIS y 0 en el LIP, para indicar desconocimiento acerca de algo), *come Dio comanda* (40 ocurrencias en el CORIS y 0 en el LIP, para referirse a algo bien hecho), *essere figlio di un Dio minore* (25 ocurrencias en el CORIS, para referirse a una persona más desafortunada), *Dio sa come* (22 ocurrencias en el CORIS y 0 en el LIP, utilizada cuando algo ocurre de manera inexplicable), *raccomandare/rendere l'anima a Dio* (14 ocurrencias en el CORIS y 0 en el LIP, como se vio en la entrada *anima*), *Dio li fa e poi li accoppia*<sup>136</sup> (12 ocurrencias en el CORIS y 0 en el LIP), *mandato da Dio* (9 ocurrencias en el CORIS y 0 en el LIP, para referirse a algo providencial), *che Dio la manda* (8 ocurrencias en el CORIS y 0 en el LIP, para referirse normalmente a la intensidad de la lluvia o nieve), *a Dio piacendo* (8 ocurrencias en el CORIS y 0 en el LIP, con significado de “si Dios quiere”), *come Dio vuole* (5 ocurrencias en el CORIS y 0 en el LIP, para referirse a que algo ocurre según la voluntad de Dios), *non si muove foglia che Dio non voglia* (4 ocurrencias en el CORIS y 0 en el LIP, para indicar que todo ocurre según la voluntad de Dios) y *Dio non paga il sabato* (3 ocurrencias en el CORIS y 0 en el LIP, para indicar que el castigo divino tarde o temprano siempre llega).

**Domenica:** 12.046 ocurrencias en el CORIS y 88 en el LIP. Ninguna expresión en los corpus.

## E

**Ebraismo:** 407 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Eresia:** 616 ocurrencias en el CORIS y 2 en el LIP. Destaca en los corpus la expresión *essere un'eresia* (28 en el CORIS y 1 en el LIP) para referirse a algo absurdo y disparatado.

## F

**Famiglia:** 41.985 ocurrencias en el CORIS y 113 en el LIP. Ninguna expresión en los corpus relacionada con la religión.

**Fede:** 8.956 ocurrencias en el CORIS y 13 en el LIP. Destacan en los corpus las expresiones *in buona fede* (1.034 ocurrencias en el CORIS y 4 en el LIP)<sup>137</sup>, *in cattiva fede* (17 ocurrencias en el CORIS y 0 en el LIP), *avere fede [in qualcuno/qualcosa]* (124 ocurrencias en el CORIS y 0 en el LIP) y *far fede* (6 ocurrencias en el CORIS y 0 en el LIP, con significado de “garantizar algo”).

**Fedele:** 5.342 ocurrencias en el CORIS y 4 en el LIP. Ninguna expresión en los corpus.

136. Es equivalente al refrán español “Dios los cría y ellos se juntan”, aunque con un matiz diferente, como se verá en 4.4.

137. Se ha optado por ponerlos separados debido a la gran diferencia de ocurrencias. Es evidente que el italiano tiene una clara predisposición a la forma con *buona*.

## G

**Gesù:** 7.047 ocurrencias en el CORIS y 0 en el LIP. Solo se encontraron en los corpus las interjecciones *Gesù!* (52 ocurrencias), *Gesù mio!* (17 ocurrencias) y *Gesù e Maria!* (2 ocurrencias).

**Gesuita:** 806 ocurrencias en el CORIS y 1 en el LIP. Ninguna expresión en los corpus.

## I

**Ideologia:** 2.434 ocurrencias en el CORIS y 8 en el LIP. Ninguna expresión en los corpus.

**Impuro:** 342 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Incenso:** 392 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Indulgencia:** 567 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *essere un venditore di indulgenze* (2 ocurrencias) para referirse a un charlatán.

**Inginocchiatoio:** 40 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Inquisizione:** 143 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Islam:** 1.965 ocurrencias en el CORIS y 3 en el LIP. Ninguna expresión en los corpus.

## L

**Lutero:** 410 ocurrencias en el CORIS y 3 en el LIP. Ninguna expresión en los corpus.

## M

**Madonna:** 2.986 ocurrencias en el CORIS y 55 en el LIP. Destacan en los corpus las expresiones *della madonna* (60 ocurrencias en el CORIS y 5 en el LIP, para referirse a algo grande e/o impactante) y *tirar giù le madonne* (2 ocurrencias en el CORIS y 0 en el LIP)<sup>138</sup>. Además, se encontraron las interjecciones *Madonna mia!* (20 ocurrencias en el CORIS y 14 en el LIP), *Madonna!* (42 ocurrencias en el CORIS y 50 en el LIP) y *Madonna santa!* (10 ocurrencias en el CORIS).

**Martirio:** 466 ocurrencias en el CORIS y 0 en el LIP. Destaca en el corpus la expresión *essere un martirio* (14 ocurrencias) para referirse a una carga pesada que nos aflige.

**Matrimonio:** 9.998 ocurrencias en el CORIS y 20 en el LIP. Ninguna expresión en los corpus.

**Messa:** 9.249 ocurrencias en el CORIS (de las cuales solo unas 1.500 del sustantivo *mesa*) y 0 en el LIP<sup>139</sup>. Ninguna expresión en los corpus.

**Mitologia:** 576 ocurrencias en el CORIS y 3 en el LIP. Ninguna expresión en los corpus.

**Monaco:** 2.529 ocurrencias en el CORIS y 2 en el LIP. Destaca en el CORIS el refrán popular *l'abito non fa il monaco* (26 ocurrencias), para indicar que no hay que confiar en las

138. En italiano existe también la expresión *avere le madonne* (estar de mal humor) pero no apareció en ninguna búsqueda, por lo que, lógicamente, no se tendrá en cuenta (Data-driven).

139. En la búsqueda aparece también el *homógrafo* *mesa*, participio pasado femenino del verbo *mettere*, porque no es posible filtrar por categoría gramatical.

apariencias.

**Monastero:** 1.354 ocurrencias en el CORIS y 6 en el LIP. Ninguna expresión en los corpus.

**Monoteísmo:** 59 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Moschea:** 986 ocurrencias en el CORIS y 3 en el LIP. Ninguna expresión en los corpus.

## N

**Natale:** 707 ocurrencias en el CORIS y 23 en el LIP. Destaca en el CORIS la expresión *durare da Natale a Santo Stefano* (1 ocurrencia), para referirse a algo que dura poco tiempo.

**Navata:** 440 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## O

**Oratorio:** 382 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *da oratorio* (8 ocurrencias) para referirse a algo vulgar y de bajo nivel.

**Ostia:** 202 ocurrencias en el CORIS y 0 en el LIP. Solo aparece una ocurrencia en el CORIS de la expresión *non capire un'ostia*, con significado de “no entender nada”.

## P

**Pace:** 12.168 ocurrencias en el CORIS y 48 en el LIP. Solo se encontró en el CORIS la expresión *andare in pace* (11 ocurrencias) pronunciada normalmente por el cura al final de la misa y utilizada a menudo de manera figurada con el significado de “ir cada uno por su cuenta”.

**Padre nostro:** 4 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Paganesimo:** 102 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Pane:** 5.983 ocurrencias en el CORIS y 26 en el LIP. Solo hay una expresión relacionada con la religión en el CORIS y es *dire pane al pane e vino al vino* (27 ocurrencias), para referirse, igual que en español, a la idea de llamar las cosas por su nombre<sup>140</sup>.

**Papa:** 7.508 ocurrencias en el CORIS y 9 en el LIP. Se encontraron las expresiones *a ogni morte di papa* (9 ocurrencias en el CORIS, con significado de “muy raramente”) y *essere il papa di un grupo di persone* (3 ocurrencias en el CORIS, para referirse, a menudo de forma irónica, a la persona más importante de un determinado grupo social).

**Pasqua:** 1.720 ocurrencias en el CORIS y 18 en el LIP. Destaca en el CORIS la expresión *essere felice/contento come una pasqua*<sup>141</sup> (19 ocurrencias), para indicar que una persona está alegre y contenta.

**Pellegrino:** 1.344 ocurrencias en el CORIS y 1 en el LIP. Solo se encontró en el CORIS la expresión *idee pellegrine* (1 ocurrencia) para referirse a ideas extravagantes.

**Pentimento:** 568 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Pianto:** 2.634 ocurrencias en el CORIS y 1 en el LIP<sup>142</sup>. Ninguna expresión en los corpus

140. Véase 2.3.3.

141. *Pasqua* aparece tanto con mayúscula como con minúscula

142. Incluyendo el participio pasado del verbo *piangere*.

relacionada con el ámbito religioso.

**Pietà:** 2.141 ocurrencias en el CORIS y 0 en el LIP. Se encontraron en el corpus las expresiones *senza pietà* (289 ocurrencias, con el mismo significado que en español que se vio en 4.2.3.1), *avere pietà* (171 ocurrencias en el CORIS, con el mismo significado que en español) y *far pietà* (22 ocurrencias, para referirse a algo que está mal o ha salido mal).

**Pregare:** 5.732 ocurrencias en el CORIS y 116 en el LIP buscando una por una las formas conjugadas. Destaca en el CORIS la expresión *farsi pregare* (45 ocurrencias) con significado de “hacerse rogar”.

**Preghiera:** 4.155 ocurrencias en el CORIS y 8 en el LIP. Se encontró en el CORIS la expresión *esaudire la preghiera [di qualcuno]* (4 ocurrencias), con significado de “cumplir los deseos de alguien”.

**Presepe:** 328 ocurrencias en el CORIS y 1 en el LIP. Ninguna expresión en los corpus.

**Prete:** 3.740 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Profeta:** 1.803 ocurrencias en el CORIS y 4 en el LIP. Se encontró en el CORIS la expresión *nessuno è profeta in patria* (20 ocurrencias) con clara referencia a Jesús, a quien no se le prestaba ningún tipo de atención en su tierra natal.

**Pulpito:** 305 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *da che pulpito viene la predica* (4 ocurrencias), utilizada de manera irónica cuando alguien no predica con el ejemplo.

## R

**Raccoglimento:** 188 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *osservare un minuto de raccoglimento* (17 ocurrencias), con significado de “guardar silencio durante un minuto en señal de respeto y devoción a una tragedia ocurrida”.

**Reclusione:** 1.691 ocurrencias en el CORIS y 1 en el LIP. Ninguna expresión en los corpus.

**Ritiro:** 2.490 ocurrencias en el CORIS y 11 en el LIP. Ninguna expresión en los corpus.

## S

**Santino:** 142 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Santo:** 16.459 ocurrencias en el CORIS y 42 en el LIP. Se encontraron en los corpus las interjecciones *santo cielo!* (ya se vio en la entrada cielo), *Cristo santo!* (entrada Cristo) *santo Dio!* (entrada Dio), Madonna santa! (entrada Madonna), *per tutti i santi!* (3 ocurrencias en el CORIS y 0 en el LIP), y las expresiones *essere un santola* (140 ocurrencias en el CORIS y 1 en el LIP), *dare/picchiare/sonare di santa ragione* (75 ocurrencias en el CORIS, con significado de “dar una paliza”), *tutto il santo giorno* (28 ocurrencias en el CORIS y 0 en el LIP, para delatar impaciencia), *non essere uno stinco di santo* (26 ocurrencias en el CORIS, para referirse a una persona poco honesta), *parole sante* (20 ocurrencias en el CORIS, dicho para demostrar nuestro acuerdo respecto a alguna afirmación), *avere qualche santo in paradiso* (9 ocurrencias en el CORIS y 0 en el LIP, para indicar que alguien tiene un ángel de la guarda) y *non supere a che santo votarsi* (2 ocurrencias en el CORIS y 0 en el LIP, utilizada en situaciones desesperadas, cuando ya no se sabe a quién acudir).

**Sindone:** 19 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Suora:** 1.577 ocurrencias en el CORIS y 4 en el LIP. Ninguna expresión en los corpus.

**Superstizione:** 454 ocurrencias en el CORIS y 4 en el LIP. Ninguna expresión en los corpus.

## T

**Trinità:** 70 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## V

**Vangelo:** 630 ocurrencias en el CORIS y 11 en el LIP. Se encontró en el CORIS la expresión *essere vangelo* (5 ocurrencias) para referirse a algo que es considerado verdad absoluta.

**Vaticano:** 2.762 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en el corpus.

**Vino:** 9.830 ocurrencias en el CORIS y 21 en el LIP. Destaca en el CORIS la expresión *dire pane al pane e vino al vino* (27 ocurrencias), que ya se vio en la entrada pane<sup>143</sup>.

**Voto:** 15.168 ocurrencias en el CORIS y 59 en el LIP. Destaca en el CORIS la expresión *fare voto [di qualcosa]* (38 ocurrencias), con significado de “perseguir algo siguiendo los principios religiosos”.

Como se puede observar, pese a tener un tamaño más reducido respecto al CORPES y a no estar lematizados, tanto el CORIS como el LIP han proporcionado varias expresiones italianas relacionadas con la religión, lo que corresponde al primer objetivo (A) de esta tesis de investigación.

### 4.3.3.2. Fase 3 con el listado del fútbol en italiano

En este subapartado se analizarán una por una las palabras del listado del fútbol obtenidas tras la criba para encontrar el mayor número posible de expresiones relacionadas con este ámbito (objetivo A). Es evidente que el fútbol tiene mucha trascendencia también en Italia, pero la interacción con la lengua es mucho más reciente respecto a la religión, por lo que el caudal de expresiones en principio será inferior. Vamos a ver ahora qué ocurre en italiano buscando las palabras del listado por *forma* en el CORIS y en el LIP<sup>144</sup>.

## A

**Ala:** 2.012 ocurrencias en el CORIS y 1 en el LIP. No se encontró en los corpus ninguna expresión relacionada con el fútbol.

**Allenamento:** 1.624 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Arbitro:** 2.637 ocurrencias en el CORIS y 2 en el LIP. Se encontró en el CORIS la expresión

143. Véase también 2.3.3.

144. Las búsquedas se realizaron a finales de 2016, por lo que es posible que en la actualidad los resultados obtenidos no se ajusten por completo debido a los cambios y a las actualizaciones que se van introduciendo en el CORIS.

sión *fare/fungere da arbitro* (14 ocurrencias), con significado metafórico de “mediar en un asunto”.

**Attaccante:** 1.578 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## B

**Bandiera:** 3.586 ocurrencias en el CORIS y 7 en el LIP. Destaca en el CORIS la expresión *fare/segnare il gol della bandiera*<sup>145</sup> (17 ocurrencias).

**Balotelli:** 114 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Bianconero:** 555 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Bomber:** 762 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *essere un bomber*<sup>146</sup> (7 ocurrencias).

**Buffon:** 457 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## C

**Calciatore:** 1.663 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Calzetto:** 7 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Calzettoni:** 112 ocurrencias en el CORIS y 1 en el LIP. Ninguna expresión en los corpus.

**Campetto:** 95 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Campionato:** 4.595 ocurrencias en el CORIS y 8 en el LIP. Destaca en el CORIS la expresión *da campionato* (5 ocurrencias) para referirse a algo de gran importancia.

**Campo:** 35.617 ocurrencias en el CORIS y 72 en el LIP<sup>147</sup>. Se encontraron en los corpus las expresiones *scendere in campo* (610 ocurrencias en el CORIS y 3 en el LIP, con significado de “participar en algo”) y *a tutto campo* (386 ocurrencias en el CORIS y 1 en el LIP)<sup>148</sup>.

**Capitano:** 4.703 ocurrencias en el CORIS y 5 en el LIP. Ninguna expresión en los corpus.

**Cartellino giallo:** 68 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Cartellino rosso:** 457 ocurrencias en el CORIS y 55 en el LIP. Se encontró en el CORIS la expresión *essere da cartellino rosso* (2 ocurrencias), equivalente a la española (4.2.3.2).

**Classifica:** 4.623 ocurrencias en el CORIS y 8 en el LIP. Destaca en el CORIS la expresión *essere/balzare in testa alla classifica* (161 ocurrencias), con significado de “llegar a la primera posición”<sup>149</sup>.

**Coro:** 2.762 ocurrencias en el CORIS y 7 en el LIP. Se encontró en el CORIS la expresión

145. Se utiliza, incluso en ámbitos no *futbolísticos*, cuando se consigue marcar al menos un tanto frente a una clara derrota.

146. Se utiliza con un significado metafórico similar a la expresión española ser un crack (4.2.3.2).

147. Se trata de una palabra polisémica muy frecuente en la lengua, lo que justifica el gran número de ocurrencias.

148. Esta última procede del ámbito futbolístico y se refiere a hacer algo de manera completa y exhaustiva, abarcando, literalmente, todas las zonas del campo.

149. Es una expresión que en italiano se utiliza de forma muy generalizada para cualquier contexto.

*accogliere/ricevere [qualcuno] con cori da stadio* (5 ocurrencias), utilizada cuando alguien es recibido con abucheos y cánticos típicos de los campos de fútbol.

**Corner:** 234 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *salvarsì/rifugiarsi in corner* (12 ocurrencias) cuando alguien se salva de un peligro *in extremis*.

**Cronista:** 1.708 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Curva:** 4.234 ocurrencias en el CORIS y 19 en el LIP. Ninguna expresión en los corpus relacionada con el fútbol.

## D

**Derby:** 1.028 ocurrencias en el CORIS y 1 en el LIP. Ninguna expresión en los corpus.

**Divisa:** 4.226 ocurrencias en el CORIS y 10 en el LIP<sup>150</sup>. Ninguna expresión en los corpus relacionada con el fútbol.

**Domenica:** 12.046 ocurrencias en el CORIS y 88 en el LIP. Ninguna expresión en los corpus.

**Dribblare:** 142 ocurrencias en el CORIS y 0 en el LIP buscando una por una todas las formas flexionadas. Se encontró en el corpus la expresión *dribblare ogni domanda/ogni polemica* (12 ocurrencias, con significado de “esquivar”)<sup>151</sup>.

## E

**Espulso:** 457 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Europei:** 508 ocurrencias en el CORIS y 0 en el LIP buscando solo la forma en plural y con mayúscula inicial para no confundir con el adjetivo. Ninguna expresión en los corpus.

## F

**Fallo:** 1.296 ocurrencias en el CORIS y 11 en el LIP<sup>152</sup>. Ninguna expresión en los corpus.

**Finta:** 3.176 ocurrencias en el CORIS y 20 en el LIP. Se encontró en los corpus la expresión *far finta di niente*, muy frecuente en la lengua, pero no se tendrá en cuenta por no proceder del ámbito futbolístico.

**Fischietto:** 259 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Fuorigioco:** 386 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *andare/mettere in fuorigioco* (3 ocurrencias, con significado de “encontrarse o de poner a alguien en una situación desfavorable y no correcta”).

## G

**Giocatore:** 7.254 ocurrencias en el CORIS y 36 en el LIP. Ninguna expresión en los corpus.

150. Incluyendo también el participio pasado del verbo *dividere*, que es homógrafo.

151. Aparece solo con esos dos sustantivos, pero es plausible pensar que en un futuro y en otra búsqueda posterior pueda aparecer acompañada de nuevos sustantivos.

152. Incluyendo el imperativo del verbo *fare* con el pronombre lo/li, que es homógrafo.

**Gioco:** 24.174 ocurrencias en el CORIS y 62 en el LIP<sup>153</sup>. No se ha encontrado en los corpus ninguna expresión relacionada directamente con el fútbol.

**Gol:** 6.060 ocurrencias en el CORIS y 4 en el LIP<sup>154</sup>. Ninguna expresión a excepción de *fare/segnare il gol della bandiera*, que ya se vio en la entrada *bandiera*.

**Gradinata:** 290 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Gullit:** 79 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## H

**Hooligans:** 53 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## J

**Juve:** 3.471 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## L

**Laccio:** 553 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## M

**Maglia:** 2.433 ocurrencias en el CORIS y 7 en el LIP. Ninguna expresión en los corpus relacionada con el fútbol.

**Messi:** 170 ocurrencias en el CORIS y 0 en el LIP<sup>155</sup>. Ninguna expresión en los corpus.

**Match:** 1.850 ocurrencias en el CORIS y 1 en el LIP.<sup>156</sup> No se encontró ninguna expresión en los corpus relacionada con el fútbol.

**Mister:** 569 ocurrencias en el CORIS y 3 en el LIP. Ninguna expresión en los corpus.

**Moggi:** 321 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Mondiali:** 2.099 ocurrencias en el CORIS y 4 en el LIP<sup>157</sup>. Ninguna expresión en los corpus.

**Moviola:** 115 ocurrencias en el CORIS y 0 en el LIP. Solo se encontró en el corpus la expresión *lavorare alla moviola* (2 ocurrencias) para indicar lentitud en el trabajo.

153. Aparecen muchas expresiones porque se incluye la primera y la segunda persona del verbo *giocare*, que es homógrafa. Una vez más, el problema de la lematización entorpece la búsqueda y esperamos que en el futuro se encuentre alguna solución a este problema (5.1).

154. Se ha buscado también la forma inglesa *goal* y han aparecido solo 287 ocurrencias en el CORIS, lo que indica que el italiano, en este caso, se ha decantado por el calco semántico en lugar del préstamo bruto.

155. Se buscó solo la forma con mayúscula inicial, ya que en italiano existe un homógrafo con el participio pasado plural del verbo *mettere*.

156. Se trata de un préstamo bruto del inglés con bastante frecuencia en la lengua.

157. Se buscó solo la forma en plural, que es la que se refiere al acontecimiento futbolístico. Lógicamente, en las ocurrencias aparece también al adjetivo plural de *mondiale*, que es homógrafo.

**N**

**Nazionale:** 28.685 ocurrencias en el CORIS y 81 en el LIP<sup>158</sup>. No se encontró ninguna expresión en los corpus relacionada con el fútbol.

**P**

**Palla:** 4.977 ocurrencias en el CORIS y 45 en el LIP. Se trata de una palabra polisémica con muchas expresiones, pero la única que procede del ámbito futbolístico y que se encontró en el CORIS es *palla al centro* (19 ocurrencias) con significado de “punto y aparte” o “borrón y cuenta nueva”.

**Pallone:** 2.790 ocurrencias en el CORIS y 10 en el LIP. Destaca en los corpus la expresión *andare nel pallone* (50 ocurrencias en el CORIS y 1 en el LIP), usada con significado de “ir en confusión”<sup>159</sup>.

**Pantaloncino:** 201 ocurrencias en el CORIS y 5 en el LIP. Ninguna expresión en los corpus.

**Partita:** 14.764 ocurrencias en el CORIS y 53 en el LIP<sup>160</sup>. Solo se encontró en el CORIS la expresión *non esserci partita* (3 ocurrencias), utilizada cuando, en una lucha entre dos bandos, uno gana fácilmente al otro.

**Piqué:** 4 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Porta:** 41.212 ocurrencias en el CORIS y 105 en el LIP<sup>161</sup>. Ninguna expresión en los corpus relacionada con el fútbol.

**Portiere:** 2.170 ocurrencias en el CORIS y 6 en el LIP<sup>162</sup>. Ninguna expresión en los corpus.

**Prato:** 2.904 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Pulcini:** 236 ocurrencias en el CORIS y 2 en el LIP<sup>163</sup>. Ninguna expresión en los corpus.

**R**

**Radiolina:** 88 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Rete:** 29.947 ocurrencias en el CORIS y 27 en el LIP<sup>164</sup>. No se encontró ninguna expresión en los corpus relacionada con el fútbol.

**Rigore:** 3.870 ocurrencias en el CORIS y 7 en el LIP. Ninguna expresión en los corpus.

158. Aparecen muchísimas ocurrencias debido a la presencia del adjetivo *nazionale*, palabra polisémica que se suele utilizar en Italia para referirse a la Selección nacional de fútbol.

159. Su etimología no está clara y entendemos que su relación directa con el mundo del fútbol es cuestionable. Sin embargo, más de la mitad de las ocurrencias proceden de artículos sobre el fútbol, por lo que se va a tener en cuenta para la pirámide.

160. La mayoría de las ocurrencias son del participio pasado femenino del verbo *partire*.

161. Aparecen muchísimas ocurrencias del verbo *portare* en tercera persona singular.

162. Se trata de una palabra polisémica con diferentes significados.

163. Se buscó solo la forma en plural porque es la que se utiliza para referirse, en los equipos de fútbol, a los alevines.

164. Se trata de otra palabra polisémica muy frecuente en la lengua, lo que justifica el gran número de casos.

## S

**Scarpa:** 5.903 ocurrencias en el CORIS y 17 en el LIP. Se encontró en el CORIS la expresión *appendere le scarpe al chiodo* (4 ocurrencias) con significado de “abandonar una actividad”.

**Schedina:** 189 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Scudetto:** 1.527 ocurrencias en el CORIS y 4 en el LIP. Ninguna expresión en los corpus.

**Scuola calcio:** 9 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Sfida:** 7.900 ocurrencias en el CORIS y 10 en el LIP. Ninguna expresión en los corpus relacionada con el ámbito futbolístico.

**Sgambetto:** 124 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *fare lo sgambetto a qualcuno*, pero no se tendrá en cuenta por no proceder del ámbito futbolístico.

**Squadra:** 14.978 ocurrencias en el CORIS y 28 en el LIP. Destacan en los corpus las expresiones *fare gioco di squadra* (61 ocurrencias en el CORIS, con significado de “trabajar en equipo”) y *essere un uomo squadra* (4 ocurrencias en el CORIS y 1 en el LIP, para referirse al líder de un grupo).

**Stadio:** 4.921 ocurrencias en el CORIS y 17 en el LIP. Aparece en el CORIS la expresión *accogliere/ricevere [qualcuno] con cori da stadio* (5 ocurrencias), que ya se comentó en coro.

**Stopper:** 54 ocurrencias en el CORIS y 0 en el LIP<sup>165</sup>. Ninguna expresión en los corpus.

**Striscione:** 722 ocurrencias en el CORIS y 9 en el LIP. Ninguna expresión en los corpus.

**Supplementari:** 527 ocurrencias en el CORIS y 2 en el LIP. Ninguna expresión en los corpus.

**Supporters:** 45 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## T

**Tacchetto:** 29 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *avere il morale sotto i tacchetti* (1 ocurrencia) para indicar que la moral de un grupo es muy baja.

**Tecnica:** 18.995 ocurrencias en el CORIS y 23 en el LIP<sup>166</sup>. Ninguna expresión en los corpus.

**Tifare:** 272 ocurrencias en el CORIS y 0 en el LIP<sup>167</sup>. Ninguna expresión en los corpus.

**Tifo:** 617 ocurrencias en el CORIS y 1 en el LIP<sup>168</sup>. Con la segunda acepción, se acuñó en la lengua la expresión *fare il tifo [per qualcuno/qualcosa]* (221 ocurrencias en el CORIS y 1 en el LIP), que puede ser sustituida por el verbo *tifare*. De cara a la pirámide, se va a tener

165. Se trata de un préstamo bruto del inglés utilizado en italiano para designar al defensa central de un equipo de fútbol.

166. Se buscaron las formas tanto del sustantivo como del adjetivo.

167. Se trata de una verbalización del sustantivo tifo, cuyo significado está estrechamente vinculado con el mundo del deporte (seguir/animar a un equipo).

168. Se trata de una palabra polisémica con dos acepciones principales: la primera se refiere a la enfermedad, mientras que la segunda, mucho más utilizada, designa la animosidad y la pasión hacia un equipo o un grupo en concreto.

en cuenta solo *fare el tifo*, ya que el primer objetivo (A) consiste en buscar solo expresiones pluriverbales en los corpus a partir del listado definitivo obtenido tras la fase 2 (4.3.2.2).

**Toro:** 1.246 ocurrencias en el CORIS y 0 en el LIP<sup>169</sup>. Ninguna expresión en los corpus.

**Traversa:** 771 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Trofeo:** 688 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## V

**Vieri:** 1.473 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

A primera vista, se tiene la sensación de que el listado del fútbol en italiano también ha proporcionado un buen número de expresiones, y ha servido para alcanzar, gracias a los dos primeros pasos de la triangulación metodológica mixta, el primer objetivo (A) de este proyecto de investigación. Queda por analizar el listado de los toros y luego se pasará a la fase 4 del proceso de creación (4.3.4).

### 4.3.3.3. Fase 3 con el listado de los toros en italiano

El mundo taurino es un ámbito cultural que tiene muy poca presencia en la cultura y en la tradición italiana, por eso en 1.2 se formuló la hipótesis de que el número de expresiones encontradas en los corpus sería muy inferior respecto al español. Veamos ahora qué aporta esta última búsqueda en los corpus, con la que se dará por finalizada la fase 3 de creación de las pirámides etnolingüísticas en italiano. Se buscarán las palabras del listado por *forma* en el CORIS/CODIS y en el LIP<sup>170</sup>.

## A

**Arena:** 614 ocurrencias en el CORIS y 0 en el LIP. Vale la pena señalar, aunque su relación no esté directamente vinculada con la arena de toros, la expresión *essere/scendere nell'arena* (178 ocurrencias en el CORIS) con significado metafórico de “enfrentarse a un reto o una competición”, y también de “estar en medio de algo importante”. No se tendrá en cuenta para la pirámide.

## C

**Carica:** 8.391 ocurrencias en el CORIS y 11 en el LIP<sup>171</sup>. No se encontró en los corpus ninguna expresión relacionada con el mundo de los toros.

**Corno:** 1.188 ocurrencias en el CORIS y 4 en el LIP. Se encontraron en los corpus las expresiones *fare le corna e mettere le corna*, pero no se tendrán en cuenta por no proceder del

169. Se buscó solo la forma con mayúscula inicial porque se trata de un diminutivo del nombre del equipo Torino.

170. Las búsquedas se realizaron a finales de 2016, por lo que es posible que en la actualidad los resultados obtenidos no se ajusten por completo debido a los cambios y a las actualizaciones que se van introduciendo en el CORIS.

171. El gran número de ocurrencias se debe a la presencia de la tercera persona singular del presente de indicativo del verbo *caricare* y al hecho de que se trata de una palabra polisémica.

mundo taurino.

**Cornata:** 15 ocurrencias en el CORIS y 0 en el LIP. Se encontró en el corpus la expresión *fare a cornate [con qualcuno/qualcosa]* (2 ocurrencias) con significado de “luchar a brazo partido con alguien o algo”.

**Corrida:** 110 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## E

**El Cordobés:** 0 ocurrencias en el CORIS y 0 en el LIP.

## I

**Incornata:** 45 ocurrencias en el CORIS y 0 en el LIP<sup>172</sup>.

**Incornare:** 15 ocurrencias en el CORIS y 0 en el LIP buscando todas las formas flexionadas del verbo. Ninguna expresión en los corpus.

## M

**Minotauro:** 23 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Morte:** 30.113 ocurrencias en el CORIS y 41 en el LIP. Ninguna expresión en los corpus relacionada con el mundo de los toros.

**Muleta:** 12 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## O

**Orecchia/o:** 7.932 ocurrencias en el CORIS y 19 en el LIP<sup>173</sup>. Ninguna expresión en los corpus relacionada con los toros.

## P

**Picadores:** 2 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## S

**Sangue:** 16.085 ocurrencias en el CORIS y 10 en el LIP. Ninguna expresión en los corpus relacionada directamente con el mundo de los toros.

**Staccionata:** 163 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

## T

**Torero:** 106 ocurrencias en el CORIS y 0 en el LIP. Ninguna expresión en los corpus.

**Toro:** 938 ocurrencias en el CORIS y 2 en el LIP. Destaca en el CORIS la expresión *prendere il toro per le corna* (24 ocurrencias), usada con el mismo significado de “coger al toro

172. No se encontró ninguna expresión relacionada con el mundo de los toros, pero sí con el mundo del fútbol, ya que nos encontramos frente a un caso de ampliación de significado que ha trascendido de un ámbito cultural a otro. En italiano, una *incornata* es un cabezazo de un futbolista hacia la portería rival, y la metáfora procede del movimiento que hace el toro con los cuernos. No se tendrá en cuenta para la pirámide de los toros, pero es una curiosidad digna de mención en este trabajo de investigación y se retomará en 4.4.2.

173. Se buscó tanto la forma masculina como la femenina (también para el plural), ya que existe en los dos géneros.

por los cuernos”, una expresión española que se vio en 4.2.3.3.

## Z

**Zoccolo:** 777 ocurrencias en el CORIS y 1 en el LIP. Ninguna expresión en los corpus.

Tras la búsqueda en los corpus, se puede comprobar que la influencia del mundo taurino en la lengua italiana es muy reducida y apenas hay expresiones; de hecho, es bastante llamativo el listado que quedó después de la criba (fase 2, como se vio en 4.3.2.3), que solo presenta 18 palabras, de las cuales muchas son generales y su relación con los toros es bastante indirecta. Con este hecho quedaría demostrada la hipótesis de partida (1.2) según la cual si un determinado ámbito cultural tiene una presencia escasa o nula en una comunidad de hablantes, su impacto en esa lengua será a su vez escaso o nulo, y la presencia de algunas expresiones se debe más bien a la interculturalidad y a la interferencia con otras lenguas afines<sup>174</sup>.

Finalmente, tras realizar la búsqueda en los corpus y conseguir las expresiones, se han consultado el *Dizionario Sabatini Coletti*, disponible en la red para consultas, y el *Devoto-Oli, Vocabolario della lingua italiana*, y se ha comprobado que la gran mayoría de las expresiones que aparecen en el CORIS/CODIS y en el LIP coinciden con las que presentan los diccionarios, por lo que se puede colegir que la triangulación metodológica mixta a partir del *brainstorming* y de la búsqueda en los corpus, junto con el método inductivo Data-driven, ha servido para alcanzar el primer objetivo (A) de este proyecto tanto para el español como para el italiano y en los tres ámbitos culturales propuestos.

La última fase del proceso de creación de las pirámides etnolingüísticas consistirá en ordenar estas expresiones por frecuencia y rentabilidad comunicativa, siguiendo los mismos criterios generales que se ejemplificaron con el español en 4.2.4.

### 4.3.4. FASE 4: SELECCIÓN Y ORDENACIÓN DE LAS EXPRESIONES DENTRO DE LAS PIRÁMIDES

La fase 4 representa el último paso en del proceso de creación de las pirámides etnolingüísticas. Al igual que se vio con la lengua española en 4.2.4, se procederá a listar todas las expresiones encontradas con la búsqueda para luego aplicar una serie de criterios que se detallarán en los subapartados siguientes.

En primer lugar, se ordenarán las expresiones **según el número de ocurrencias**; es decir, ateniéndonos únicamente a su **frecuencia**. En segundo lugar, se aplicará el criterio de la **rentabilidad lingüística** para destacar aquellas expresiones que se pueden utilizar en la mayoría de los contextos comunicativos sin que resulten vulgares o demasiado coloquiales. En cuanto a los refranes, se tratarán de la misma manera que en las pirámides del español y se les otorgará una rentabilidad alta por su carga etnolingüística. También se realizará, en algunos casos, una segunda búsqueda en la web como corpus para corroborar algunos aspectos lingüísticos de uso, como se verá a continuación.

174. Estas ideas se retomarán en 4.4, ya que constituyen el foco del tercer objetivo (C) de esta tesis.

### 4.3.4.1. La pirámide etnolingüística de la religión en italiano

La religión ejerce una gran influencia en la lengua italiana desde hace más de dos mil años, y era inevitable que fuera el ámbito cultural con el mayor número de expresiones, como se demostró con la búsqueda en los corpus de la fase 3 (4.3.3.1). A continuación, se listarán primero todas las expresiones por número de ocurrencias y luego se decidirá si crear más de una pirámide (como se hizo con el español en 4.2.4.1) para que la visualización de datos (tercer paso de la triangulación metodológica) sea lo más clara y efectiva posible.

#### > 80 ocurrencias

In buona fede (1.038)<sup>175</sup>  
 Dio mio/mio Dio! (777)  
 Che/cosa diavolo! (334)  
 Senza pietà (289)  
 Santo cielo! (211)  
 Non esserci anima viva (185)  
 Per l'amor di Dio! (184)  
 Avere pietà (171)  
 Santo Dio/Dio santo! (151)  
 Essere un santo (141)  
 Cristo! (140)  
 Tenere a battesimo (128)  
 Essere/trovare l'anima gemella (126)  
 Avere fede [in qualcuno/qualcosa] (124)  
 Ogni/tutto questo ben di Dio (114)  
 Ah/oh Dio! (107)  
 Essere un povero cristo (96)  
 Madonna! (92)  
 Essere una bestemmia (90)  
 Essere al settimo cielo (82)  
 Avere compassione (82)  
 Per l'amor del cielo! (82)

#### Entre 20 y 79 ocurrencias

Dio solo sa (74)  
 Della madonna (65)  
 Cristo santo! (60)  
 Gesù! (52)  
 Se Dio vuole! (47)  
 Mandare al diavolo (46)

<sup>175</sup>. Esta cifra es la suma de las ocurrencias encontradas en el CORIS y en el LIP.

Farsi pregare (45)  
 Mettere in croce (44)  
 Come Dio comanda (40)  
 Essere/diventare la Bibbia (40)  
 Essere come un'anima in pena (40)  
 Fare voto [di qualcosa] (38)  
 Fare l'avvocato del diavolo (35)  
 Madonna mia! (30)  
 Grazie a Dio! (30)  
 Essere un'eresia (29)  
 Tutto il santo giorno (28)  
 Dire pane al pane e vino al vino (27)  
 Vendere/cedere l'anima al diavolo (27)  
 L'abito non fa il monaco (26)  
 Avere un diavolo per capello (23)  
 Esserci un cimitero [di qualcosa] (23)  
 Dio sa come (22)  
 Far pietà (22)  
 Grazie al cielo! (21)  
 Nessuno è profeta in patria (20)

### Entre 7 y 19 ocurrencias

Essere felice /contento come una pasqua (19)  
 Gesù mio! (17)  
 In cattiva fede (17)  
 Saperne una più del diavolo (17)  
 Osservare un minuto di raccoglimento (17)  
 Accontentarsi con quello che passa il convento (17)  
 Fare compassione (15)  
 Dio li fa e poi li accoppia (15)  
 Essere un martirio (14)  
 Raccomandare/ rendere l'anima a Dio (14)  
 Avere il diavolo in corpo (12)  
 Andare in pace (11)  
 Madonna santa! (10)  
 Consumarsi/spegnersi come una candela (10)<sup>176</sup>  
 Mandato da Dio (9)  
 A ogni morte di papa (9)  
 Che Dio la manda (8)  
 A Dio piacendo (8)

176. Para la pirámide se ha escogido la variante más frecuente con *consumarsi*.

Gettare la croce addosso [a qualcuno] (8)<sup>177</sup>  
 Da oratorio (8)  
 Come/quanto è vero Dio! (8)  
 Come una cattedrale (7)

### ◀ 7 ocurrencias

Far fede (6)  
 Che Dio ce la mandi buona! (6)  
 Dio cane! (6)  
 Come Dio vuole (5)  
 Ciascuno ha/porta la sua croce (5)  
 Essere un bravo cristo (5)  
 Essere vangelo (5)  
 Esaudire la preghiera [di qualcuno] (4)  
 Non si muove foglia che Dio non voglia (4)  
 Per tutti i santi! (3)  
 Dio non paga il sabato (3)  
 Non esserci cristo/i (3)  
 Essere tutto casa e Chiesa (3)  
 Essere il papa di un grupo di persone (3)  
 Essere come il diavolo e l'acguasanta (3)  
 Essere un venditore di indulgenze (2)  
 Tirar giù le madonne (2)  
 Non sapere a che santo votarsi (2)  
 Per carita di Dio! (2)  
 Gesù e Maria! (2)  
 Non stare né in cielo né in terra (1)  
 Durare da Natale a Santo Stefano (1)  
 Essere inocente come un'avemaria (1)  
 Non capire un'ostia (1)  
 Il convento non passa altro (1)  
 Idee pellegrine (1)

Como se puede apreciar, en este caso también se dispone de mucha información cuantitativa, razón por la que se establecerán los mismos criterios que ya se aplicaron con el español en 4.2.4.1 y que servirán para ofrecer una mejor visualización de los datos en las pirámides (tercer paso de la triangulación metodológica):

**1) PRIMER CRITERIO:** es el de la **rentabilidad comunicativa** con el que se usarán los símbolos ++ en caso de que la expresión sea rentable en la mayoría de los

177. Con sus variantes *dare/buttare la croce addosso* [a qualcuno]. Para la pirámide se ha escogido la variante más frecuente con el verbo *gettare*.

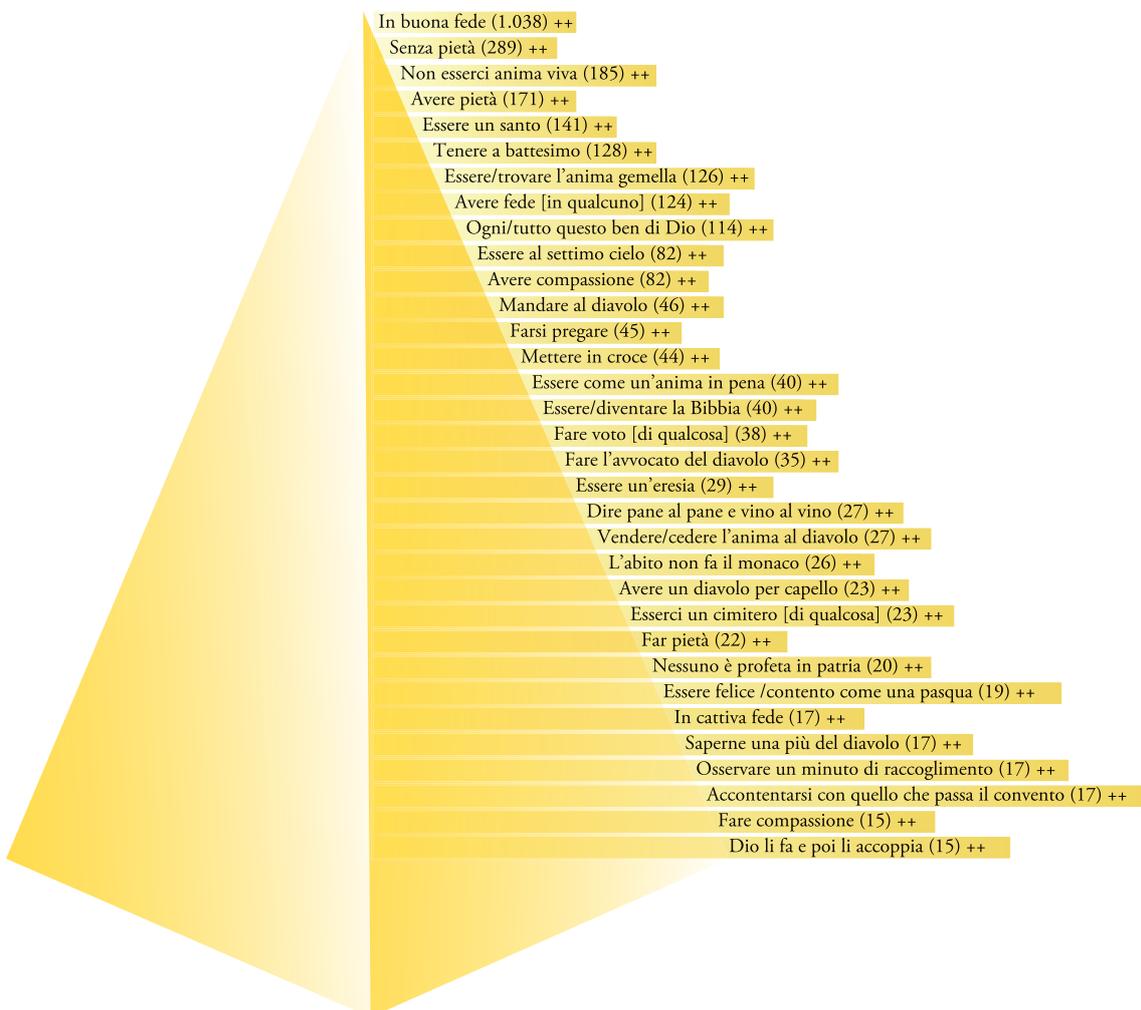
contextos, + en caso de que se considere coloquial y – con expresiones vulgares. Para asignar los valores nos hemos apoyado principalmente en el *Dizionario italiano Devoto-Oli*, que suele utilizar las marcas *pop.* o *fam.* (símbolo + en la pirámide) para designar las expresiones coloquiales, y también se ha recurrido en algunos casos a la web como corpus, aunque somos conscientes de que Google cambia de forma muy rápida y es poco probable que la búsqueda que se llevó a cabo a finales de 2016 sea la misma en el momento en el que se lea esta tesis. Si nos centramos en el listado obtenido tras la fase 3, veremos que solo hay una expresión vulgar (*¡dio cane!*), que se excluirá de la pirámide por ser una blasfemia de pésimo gusto y sin valor etnolingüístico. En cuanto a los refranes, se incluirán todos con marca ++, igual que en la pirámide del español (4.2.4.1).

**2) SEGUNDO CRITERIO:** se prescindirá de todas las interjecciones por considerar que no son demasiado relevantes para extraer información etnolingüística.

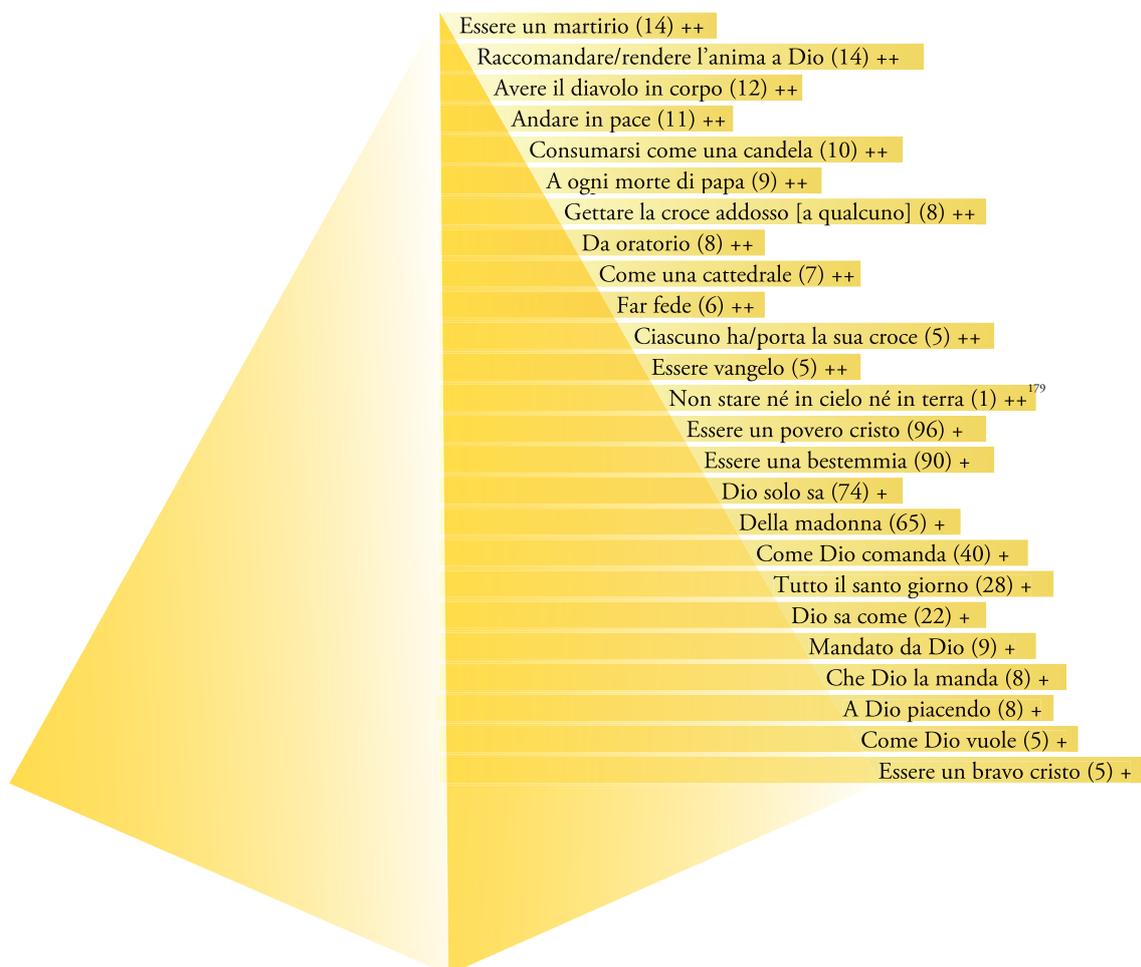
**3) TERCER CRITERIO:** se excluirán aquellas expresiones que presenten menos de 5 ocurrencias en el CORIS, que no aparezcan en el LIP y que den pocos resultados en la web como corpus, siempre por su escasa relevancia.

Aplicando estos criterios, se ha decidido repartir las expresiones en dos pirámides para que la visualización de datos sea más clara y ordenada de cara a un análisis contrastivo entre el español y el italiano (objetivo B). En la primera pirámide aparecen las expresiones más rentables y frecuentes, mientras que en la segunda las palabras menos rentables y/o con una frecuencia inferior a 15. Así es como quedarían<sup>178</sup>:

178. Debido al auge de internet, las redes sociales, los memes y el alud informativo al que estamos sometidos a diario, es posible que una expresión varíe su frecuencia y su rentabilidad comunicativa en tan solo unos años, razón por la que se ha decidido enmarcar las “pirámides etnolingüísticas” en un período limitado de tiempo (2001-2015).



**Gráfico 6. Pirámide 1/2 de la religión en italiano. Expresiones rentables (++) y frecuentes**



**Gráfico 7. Pirámide 2/2 de la religión en italiano. Expresiones rentables (++) , pero menos frecuentes y expresiones coloquiales (+)**

En el caso del italiano, no tenemos expresiones vulgares, pero llama la atención el hecho de que la mayoría de las expresiones que contienen la palabra *dio*, *madonna* y *cristo* (sin tener en cuenta los refranes) aparezcan en la segunda pirámide y sean consideradas familiares o coloquiales por el *Dizionario Devoto-Oli*, por lo que su rentabilidad comunicativa es inferior. Es probable que a estas dos pirámides se les puedan aplicar más criterios para simplificarlas y reducir el tamaño, pero consideramos que así se perdería algo de información cualitativa, punto central del segundo objetivo (B) de este proyecto de investigación, alcanzado mediante

179. Se ha decidido incluir esta expresión pese a tener menos de cinco ocurrencias porque, a raíz de la búsqueda en la web como corpus, nos salieron más de 23 millones de resultados, lo que avala su fuerte presencia en la lengua actual.

la visualización de datos en las pirámides. De todos modos, se retomará este concepto en el apartado final dedicado a las aportaciones y limitaciones de la investigación (5.1), así como en las líneas abiertas de investigación (5.3).

#### 4.3.4.2. La pirámide etnolingüística del fútbol en italiano

Con el fútbol se espera un resultado similar al que se obtuvo en español (4.2.4.2), pues se trata de un ámbito cultural bastante afín que lleva poco más de cien años en contacto con las dos lenguas. El número de expresiones es claramente inferior respecto a la religión, por lo que solo tendremos una pirámide y la tarea de clasificación resulta mucho más rápida y sencilla. A continuación, vamos a listar todas las expresiones que se obtuvieron tras la búsqueda en el CORIS/CODIS y en el LIP (fase 3, como se vio en 4.3.3.2):

Scendere in campo (613)  
 A tutto campo (387)  
 Fare il tifo (221)  
 Essere/balzare in testa alla classifica (161)  
 Fare gioco di squadra (61)  
 Andare nel pallone (51)  
 Palla al centro (19)  
 Fare/segnare il gol della bandiera (17)  
 Fare da arbitro (14)  
 Salvarsi/rifugiarsi in corner (12)  
 Dribblare ogni domanda/polemica (12)  
 Essere un bomber (7)  
 Da campionato (5)  
 Ricevere [qualcuno] con cori da stadio (5)  
 Appendere le scarpe al chiodo (4)  
 Essere un uomo squadra (4)  
 Non esserci partita (3)  
 Andare/mettere in fuorigioco (3)  
 Essere da cartellino rosso (2)  
 Lavorare alla moviola (2)  
 Avere il morale sotto i tacchetti (1)

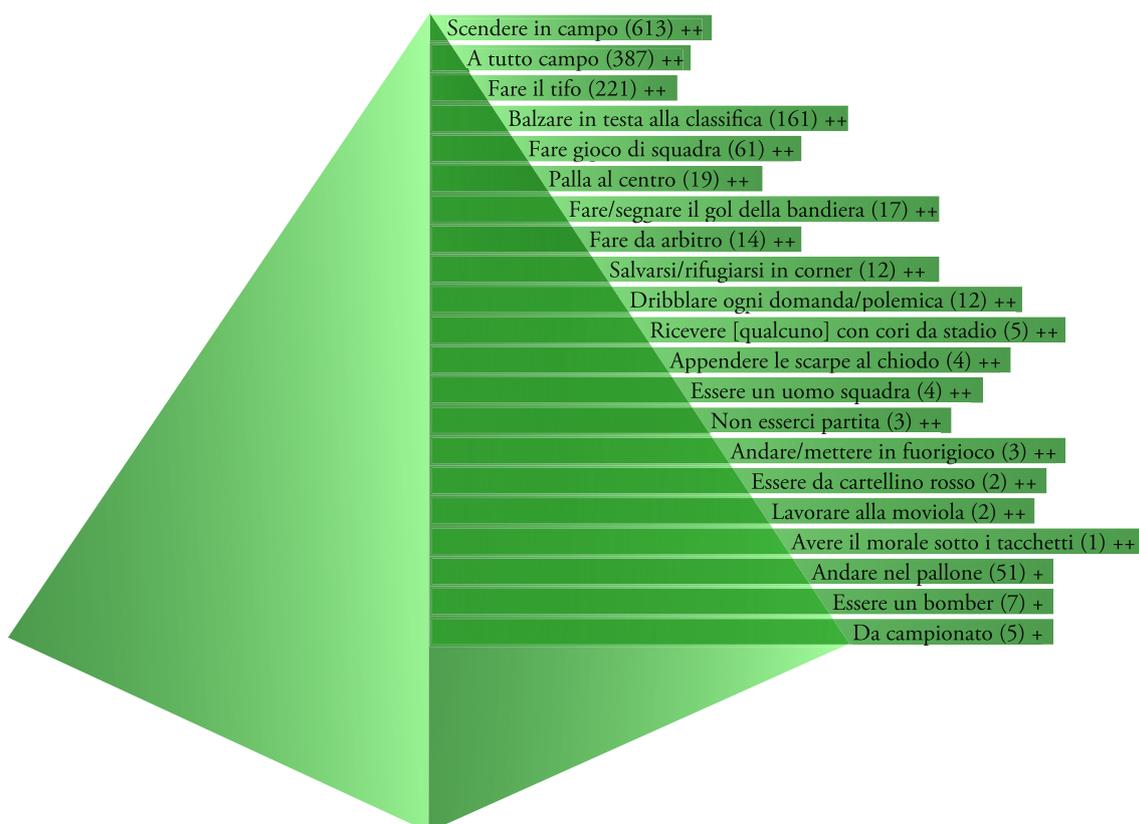
Con el italiano, igual que ocurrió con el español (4.2.4.2), el listado del fútbol es mucho más reducido respecto al de la religión, por eso será relativamente sencillo aplicar estos tres criterios, que ya se aplicaron en el subapartado anterior (4.3.4.1):

**1) PRIMER CRITERIO:** es el de la **rentabilidad comunicativa** apoyándose en las marcas *fam.* y *pop.* del *Dizionario Devoto-Oli*.

2) **SEGUNDO CRITERIO**: se prescindirá de las interjecciones por considerar que no aportan información etnolingüística sustancial.

3) **TERCER CRITERIO**: debido al número reducido de expresiones, se incluirán incluso aquellas con una frecuencia inferior a las 5 ocurrencias, igual que en la pirámide española del fútbol (4.2.4.2).

Aplicando estos tres criterios al listado, así es como quedaría la pirámide etnolingüística del fútbol para el italiano:



### Gráfico 8. Pirámide 1/1 del fútbol en italiano

El resultado final es una pirámide con **21 expresiones**, un número reducido que ofrece una visualización clara (tercer paso de la triangulación metodológica) de cómo influye el fútbol en la lengua real, mostrando, a partir de la recopilación de datos (Data-driven), qué expresiones se utilizan en Italia en el período seleccionado (2001-2015), lo que corresponde al segundo objetivo (B) de este proyecto de investigación.

Como ya se comentó al final de 4.2.4.2, es probable que la pirámide del fútbol sufra cambios en un futuro no muy lejano, ya que este deporte tiene un impacto social cada vez más fuerte en la sociedad y es normal que se vayan incorporando a la lengua nuevas expresiones. Creemos que sería interesante realizar otra pirámide dentro de unos años para cotejarla con

esta en busca de nuevas expresiones y posibles cambios etnolingüísticos<sup>180</sup>.

#### 4.3.4.3. La pirámide etnolingüística de los toros en italiano

Las corridas de toros nunca han sido populares en Italia y la tauromaquia ha sido vista siempre como una marca típicamente española que apenas ha tenido repercusiones en la cosmovisión de los hablantes italianos, ni siquiera como **cultura compartida** (2.2) Volviendo al concepto de intensidad y extensión que se introdujo en 2.4.1, es evidente que si no hay un input cultural no hay necesidad de forjar nuevas expresiones, y es justamente lo que ocurre en italiano con el ámbito de los toros. En la fase 3 (4.3.3.3) se encontraron únicamente dos expresiones: *prendere il toro per le corna* (24 ocurrencias) y *fare a cornate [con qualcuno/qualcosa]* (2 ocurrencias). Ambas presentan una rentabilidad ++, por lo que tendríamos la siguiente pirámide:



#### Gráfico 9. Pirámide 1/1 de los toros en italiano

El resultado final es una micropirámide con tan solo **2 expresiones**, un número muy reducido que, de todos modos, ofrece una visualización clara de lo poco que influye el ámbito de los toros en italiano, y con el que se completa el segundo objetivo (B) de este proyecto de investigación. En el siguiente apartado (4.4) se cotejarán todas las pirámides obtenidas en español y en italiano para demostrar, a partir de la representación visual de datos (tercer paso de la triangulación metodológica), que la cultura influye de manera diferente en cada lengua, incluso con ámbitos culturales afines (objetivo C). En el caso concreto de los toros, un ámbito cultural prácticamente inexistente en Italia, se verá claramente la gran divergencia que existe entre las dos lenguas.

180. Se ampliará esta idea en las consideraciones finales al analizar las líneas abiertas de la investigación (5.3).

## 4.4. EL ANÁLISIS CONTRASTIVO ENTRE EL ESPAÑOL Y EL ITALIANO A PARTIR DE LA VISUALIZACIÓN DE DATOS EN LAS PIRÁMIDES ETNOLINGÜÍSTICAS

Este apartado está dedicado al análisis contrastivo entre las pirámides etnolingüísticas del español y del italiano que se obtuvieron mediante las cuatro fases analizadas a lo largo de 4.2 y 4.3. Tras obtener la visualización de la información cuantitativa y cualitativa en las pirámides (tercer paso de la triangulación metodológica mixta de esta tesis) y alcanzar los primeros dos objetivos (A y B) que se propusieron en 1.2, ha llegado el momento de cotejar los resultados visuales en los tres ámbitos culturales escogidos (la religión, el fútbol y los toros) para analizar el grado de afinidad que existe entre el español y el italiano. Para ello, se tendrán muy en cuenta aspectos generales de cosmovisión, semantividad, creatividad, frecuencia y rentabilidad comunicativa que se expusieron a lo largo del capítulo 2 y que constituyen el marco teórico del que bebe este trabajo de investigación. A partir de aquí, se pretende demostrar (de acuerdo con las hipótesis de partida expuestas en 1.2) que la cultura influye de manera diferente en cada lengua, incluso con ámbitos culturales afines, y que la cuña de expresiones se debe a factores diversos, *in primis* la creatividad de los hablantes, la imitación producida por las neuronas espejo y la historicidad.

Una de las grandes preocupaciones a la hora de realizar un análisis contrastivo entre dos lenguas, especialmente si está más enfocado al ámbito fraseológico, es el de las equivalencias, que han sido analizadas y tratadas en numerosas tesis doctorales y en varios artículos académicos<sup>181</sup>. Nosotros no pretendemos entrar a fondo en materia fraseológica porque el presente trabajo se centra principalmente en aspectos etnolingüísticos a partir de la visualización de datos, pero sí es cierto que a lo largo del análisis contrastivo **nos apoyaremos en los conceptos de equivalencia total, parcial y nula** que propone Mellado Blanco como punto de apoyo para alcanzar el tercer objetivo (C) de este trabajo. Esta idea nos vino a partir de la teoría de Gloria Corpas Pastor, según la cual muchas equivalencias que a primera vista parecen totales en realidad son parciales, ya que varían en las dos lenguas en términos de frecuencia, contenido semántico, connotaciones, rentabilidad y uso (“Acerca de la [in]traducibilidad de la fraseología”, 483-522).

A continuación, se compararán todas las pirámides etnolingüísticas obtenidas en español y en italiano tras la búsqueda en los corpus (fase 3) para los tres ámbitos culturales seleccionados: la religión, el fútbol y los toros.

### 4.4.1. LAS PIRÁMIDES ETNOLINGÜÍSTICAS DE LA RELIGIÓN EN CONTRASTE

Por lo que se ha podido comprobar en 4.2.4.1 y 4.3.4.1 (fase 4), la religión es el ámbito cul-

181. Por ejemplo, el de la Dra. Carmen Mellado Blanco, de la Universidad de Santiago de Compostela, quien vuelve a centrarse en la equivalencia en el nivel sistémico, lexicográfico y textual, caracterizándose la primera por su gradualidad (equivalencia total, parcial o nula) y las otras dos por su valor funcional-comunicativo (“Parámetros específicos de equivalencia en las unidades fraseológicas, con ejemplos del español y el alemán”, 153-174).

tural que más impacto ha tenido tanto en la lengua española como en la italiana, como era de esperar dada la influencia constante a lo largo de los últimos dos mil años. El resultado de la búsqueda en los corpus (fase 3) proporcionó un gran número de expresiones que se ordenaron en tres pirámides para el español (4.2.4.1) y en dos para el italiano (4.3.4.1). Para alcanzar el tercer objetivo (C), se analizarán las expresiones que aparecen dentro de las pirámides en español y en italiano, confiando en que la visualización de datos (tercer paso que cierra la triangulación metodológica) permita detectar de forma inductiva el grado de afinidad que existe entre estas dos lenguas y ver de qué manera el ámbito cultural escogido ha influido en cada una (Data-driven).

Somos conscientes de que la religión, al llevar tantos años en contacto con las dos lenguas, tiene un impacto mucho más fuerte y variado respecto al fútbol o a los toros, lo cual se traduce en mucha información cuantitativa que no siempre resulta fácil manejar; por eso mismo, se establecieron diferentes criterios a lo largo del proceso de creación, como la criba del *brainstorming* (fase 2) y, sobre todo, la selección de las expresiones que se iban a incluir en las pirámides etnolingüísticas (paso 4) con el fin de poder aprovechar mejor el aspecto cualitativo del material obtenido. A continuación, se pondrán las cinco pirámides de la religión juntas para que la representación visual de los datos de investigación resulte más clara:

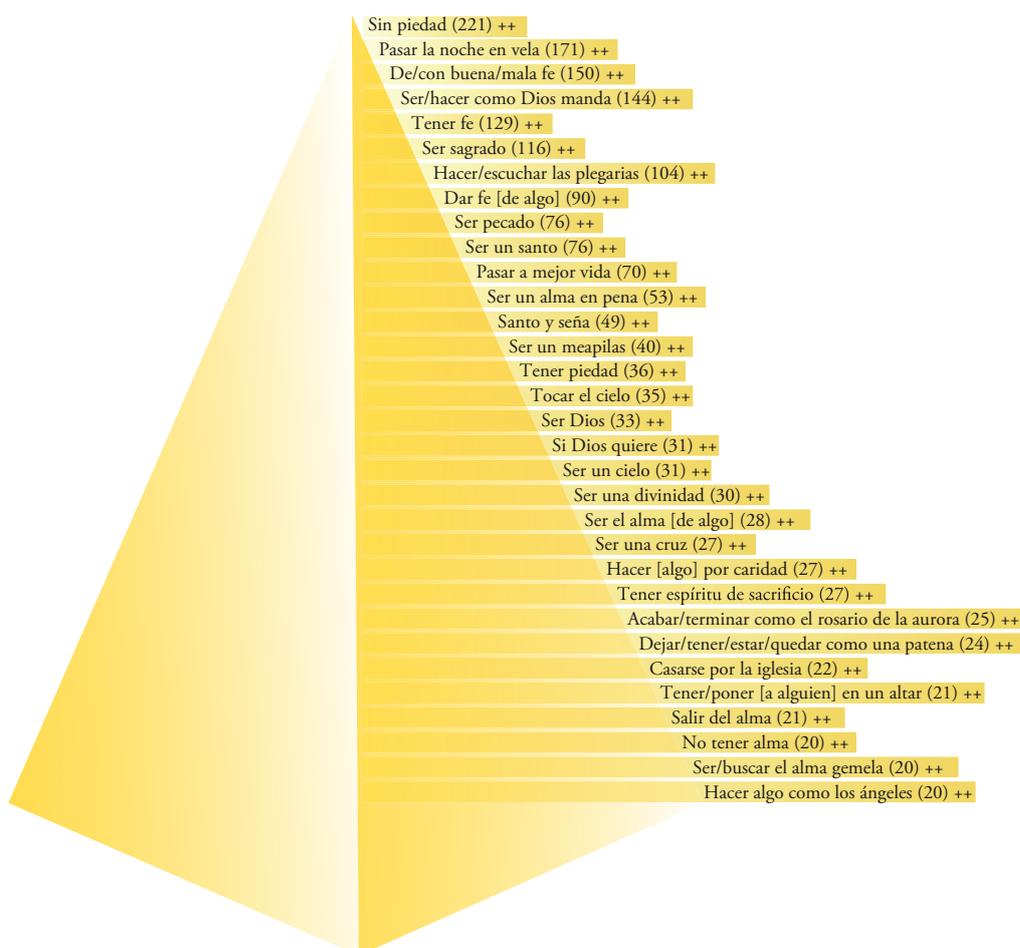


Gráfico 10. Pirámide 1/3 de la religión en español. Expresiones rentables (++) y frecuentes

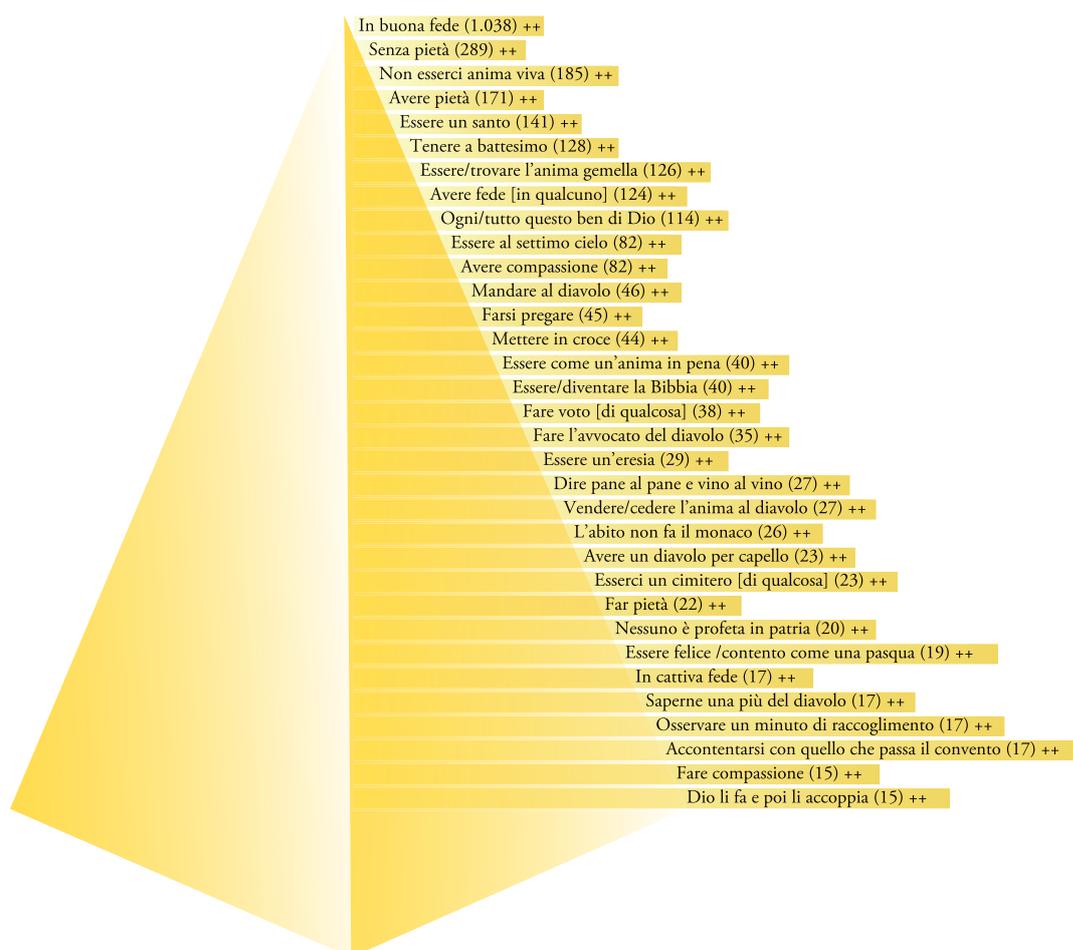
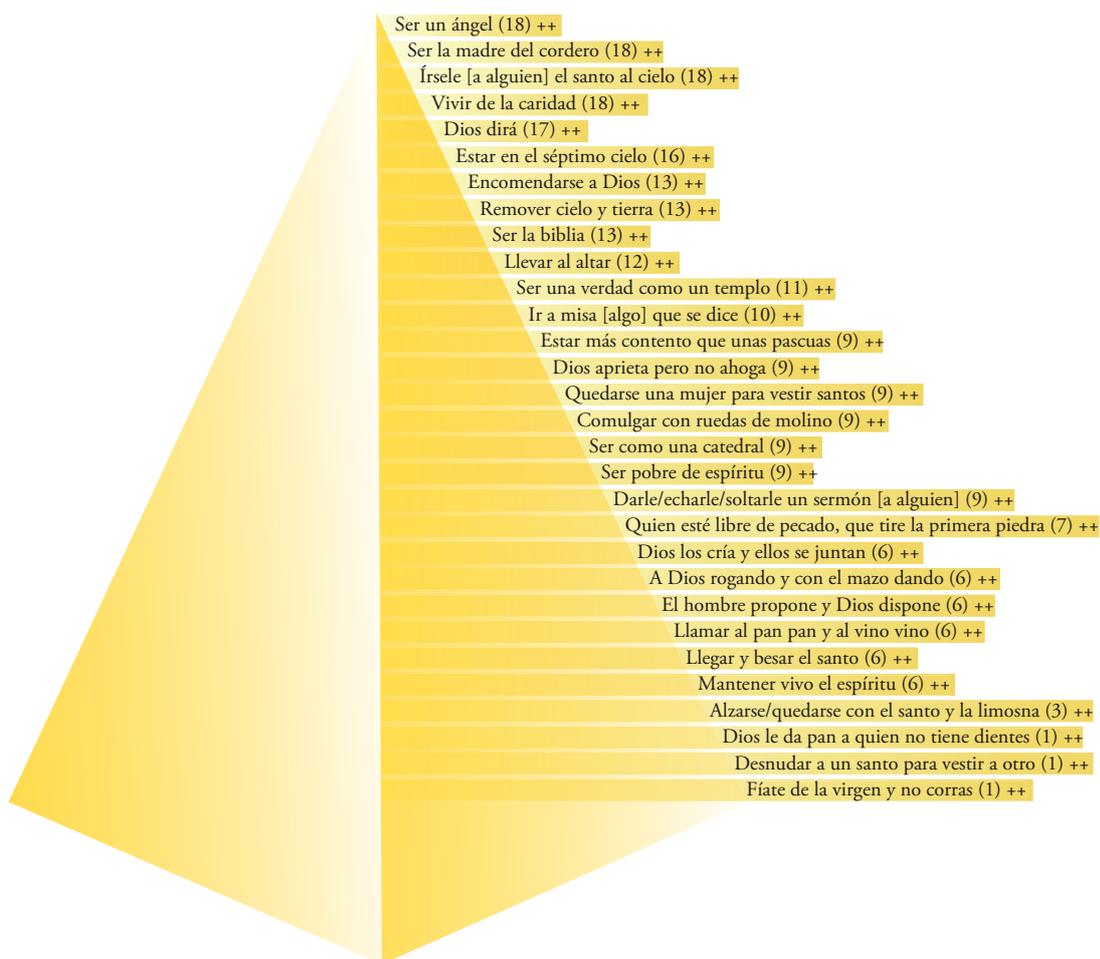
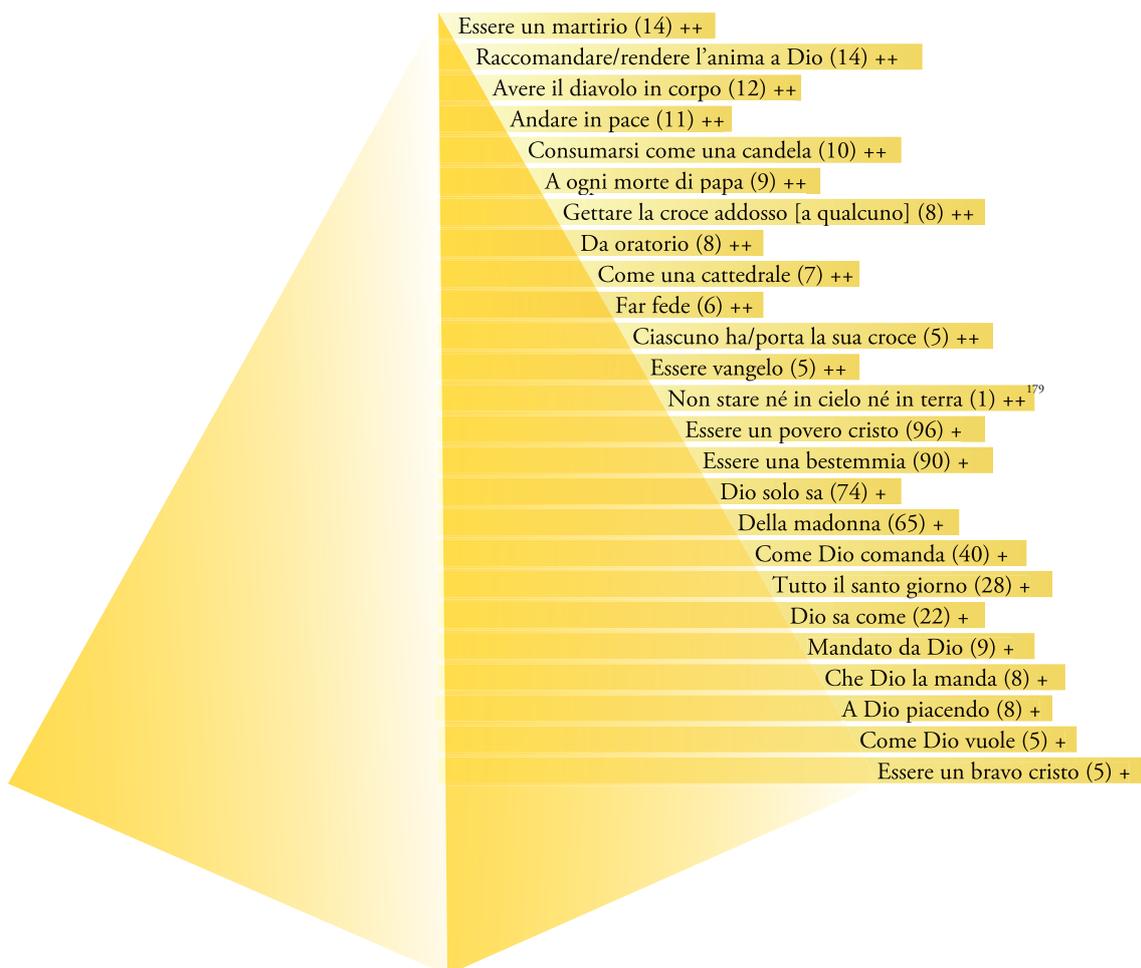


Gráfico 11. Pirámide 1/2 de la religión en italiano. Expresiones rentables (++) y frecuentes



**Gráfico 12.** Pirámide 2/3 de la religión en español. Expresiones rentables (++), pero menos frecuentes



**Gráfico 13.** Pirámide 2/2 de la religión en italiano. Expresiones rentables (++), pero menos frecuentes y expresiones coloquiales (+)

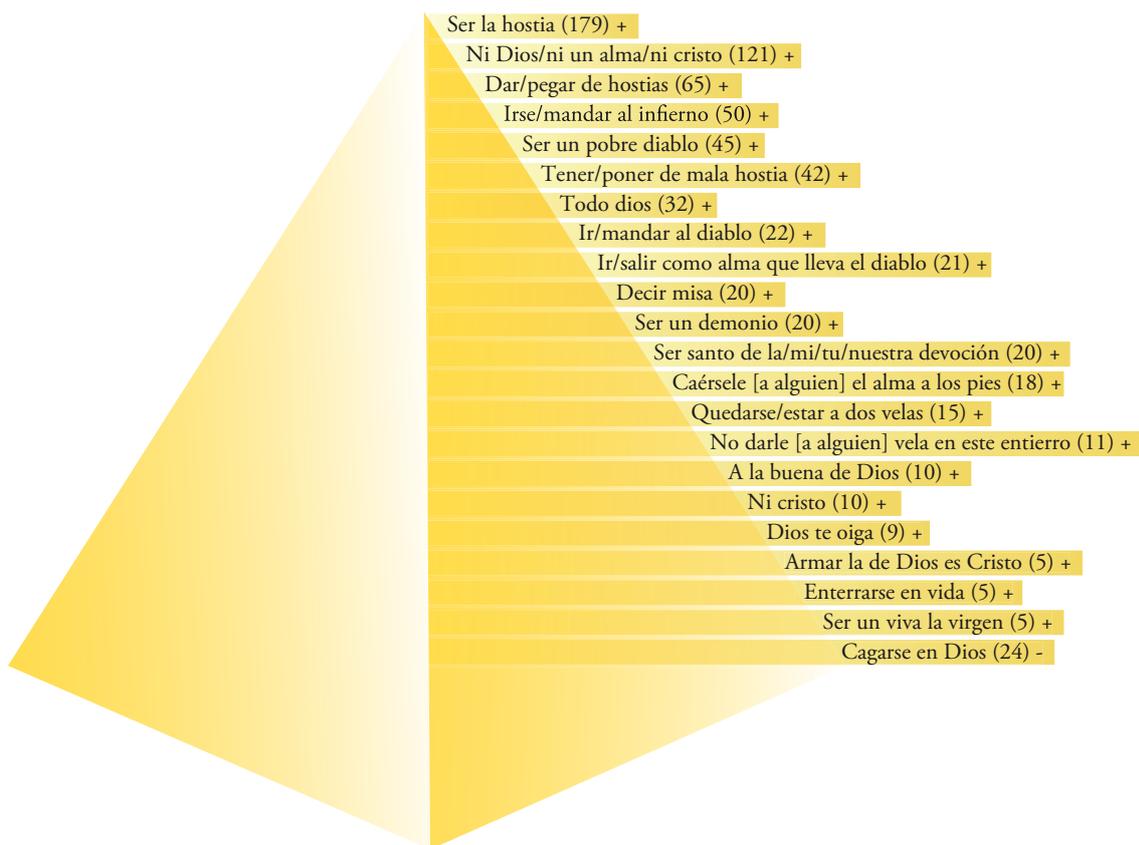


Gráfico 14. Pirámide 3/3 de la religión en español. Expresiones coloquiales (+) y vulgares (-)

Comparando las cinco pirámides, lo primero que salta a la vista es que la **lengua española presenta un mayor número de expresiones (84 expresiones frente a 58)**; sin embargo, de este dato no se puede inferir ni que la religión tenga un impacto más fuerte en español que en italiano, ni que el *brainstorming* haya sido más rico en español. La diferencia de expresiones sin duda se debe a los corpus utilizados, ya que el CORPES es más grande que el CORIS/CODIS, incluye transcripciones de la lengua oral y está lematizado, hecho que sin duda ha agilizado y facilitado la búsqueda<sup>182</sup>. El *brainstorming*, como se ha visto en 4.2.1 y en 4.3.1, dio como resultado un número de palabras bastante parejo en las dos lenguas.

Si se someten las pirámides a un análisis más detallado, se detectan algunos aspectos que vale la pena comentar. En primer lugar, **el número de expresiones menos rentables (marca +) es bastante superior en español (21 frente a 12)**, y la mayoría contienen las palabras *Dios/Dio*, *Cristo*, *Madonna* y *Virgen*, lo que indicaría que **en las dos lenguas se considera coloquial e incluso vulgar pronunciar esos nombres en vano**, especialmente en italiano, donde la pirámide muestra que 10 expresiones sobre 12 con marca + contienen *Dio/Cristo*. Respecto a ello, y retomando los conceptos de intensión y extensión mencionados en 2.4.1, se puede ver que en italiano la palabra *Cristo* presenta una mayor extensión con la expresión

182. Esto demostraría la importancia de la representatividad, de la que se habló en 3.1.1, y evidenciaría el riesgo que conlleva extraer conclusiones cuantitativas sin tener en cuenta el volumen del corpus.

*essere un povero cristo* (96, +)<sup>183</sup>, y que el equivalente español<sup>184</sup> prefiere utilizar la palabra *diablo* en *ser un pobre diablo* (45, +), presentando de todos modos una frecuencia muy inferior. La expresión *essere un bravo cristo* (5, +) también tiene una mayor extensión en *cristo*, y en este caso ni siquiera hay un equivalente en español con la misma carga cultural<sup>185</sup>.

Desde un punto de vista semántico, la diferencia entre las dos lenguas es mucho más palpable en las expresiones frecuentes y rentables, que ocupan gran parte de las pirámides. Hay cierta similitud en la parte más alta, especialmente con equivalencias que se podrían considerar casi totales, donde aparece *de buena/mala fe* (150, ++)<sup>186</sup> frente a *in buona fede* (1038, ++)<sup>186</sup>, aunque la diferencia de frecuencia es muy llamativa y llevaría a suponer que el impacto de la buena fe y la importancia que se le da es mucho más fuerte en italiano que en español. También hay similitudes entre *tener fe* (129, ++)<sup>187</sup> frente a *avere fede* (124, ++), *ser un alma en pena* (53, ++)<sup>187</sup> frente a *essere come un'anima in pena* (40, ++)<sup>187</sup> y *estar contento como unas pascuas* (9, ++)<sup>187</sup> frente a *essere felice come una pasqua* (19, ++), mientras que con otras expresiones hay una discreta diferencia de frecuencia, como *dar fe* (90, ++)<sup>187</sup> frente a *far fede* (6, ++), *ser un santo* (76, ++)<sup>187</sup> frente a *essere un santo* (141, ++), *tener piedad* (36, ++)<sup>187</sup> frente a *avere pietà* (171, ++), *estar en el séptimo cielo* (35, ++)<sup>187</sup> frente a *essere al settimo cielo* (82, ++)<sup>187</sup>, *ser/buscar el alma gemela* (20, ++)<sup>187</sup> frente a *essere/trovare l'anima gemella* (126, ++), e incluso como *Dios manda* (144, ++)<sup>187</sup> frente a *come Dio comanda* (40, +) donde no solo hay una diferencia de frecuencia, sino también de rentabilidad comunicativa. Estas diferencias demostrarían, una vez más, la tendencia más marcada del italiano a considerar *coloquiales* las expresiones que contienen la palabra *Dio*. Respecto a ello, y como se adelantó en 2.2.1 al exponer los conceptos de semanticidad y creatividad, vemos que en español tenemos 3 expresiones con significado de “nadie”: *ni Dios/ni un alma/ni cristo* (121, +), mientras que en italiano solo se encontró la expresión *non esserci anima viva* (185, ++). Esto señalaría una sustancial diferencia en términos de extensión/intensión (2.4.1), ya que, así como en español se alternan diferentes palabras procedentes del ámbito religioso, en italiano se opta por utilizar únicamente el término *anima*, evitando mencionar tanto a *Cristo* como a *Dio*.

También llaman la atención dos refranes populares que presentan una equivalencia casi total en las dos lenguas, aunque con frecuencias inferiores en español: *Dios los cría y ellos se juntan* (6, ++)<sup>187</sup> frente a *Dio li fa e poi li accoppia* (15, ++)<sup>187</sup> y *llamar al pan pan y al*

183. *Cristo* se refiere a cualquier persona, por eso también se suele escribir con minúscula inicial.

184. En términos semánticos y de uso.

185. En español existe ser buena gente, pero es mucho más neutro y no tiene la misma carga cultural que la expresión italiana.

186. En italiano se ha separado *in cattiva fede* (17, ++)<sup>187</sup> porque había demasiada diferencia de frecuencia con respecto a *in buona fede*.

187. En la pirámide 1 del español aparece también tocar el cielo (35, ++), que tiene un significado similar y que de alguna manera compensaría la diferencia de frecuencia. Así pues, la expresión italiana tendría más intensidad respecto a las dos españolas y, por consiguiente, menos extensión.

*vino vino* (6, ++) frente a *dire pane al pane e vino al vino* (27, ++) <sup>188</sup>. En la primera, que a primera vista parece una equivalencia total, llama la atención el cambio de sujeto en la segunda parte de la oración: en español *Dios los cría*, pero luego *ellos se juntan*, mientras que en italiano es el mismo Creador quien *li fa* y luego *li accoppia*. Lo que se podría colegir, al menos de este refrán, es que Dios para un español solo se encarga de crear, y luego declina toda responsabilidad hacia el hombre; para un italiano, en cambio, *Dio* es amo y señor del destino del hombre, ya que primero lo crea y después decide con quién se junta. Estos matices, detectados a partir de la representación visual de datos, aportan preciada información sobre la verdadera cosmovisión de un pueblo, lo que constituye el foco del tercer objetivo (C).

Por otro lado, si nos fijamos en las cinco pirámides, **nos encontramos con varias expresiones que no tienen un equivalente ni parcial en la otra lengua** <sup>189</sup>, bien por ser refranes que proceden del ámbito popular <sup>190</sup>, bien por ser expresiones acuñadas a partir de una determinada cosmovisión o acontecimiento histórico. Veamos algunos ejemplos:

#### ESPAÑOL

Pasar la noche en vela (171, ++)  
 Escuchar las plegarias (104, ++)  
 Pasar a mejor vida (70, ++)  
 Santo y seña (49, ++)  
 Ser un meapilas (40, ++)  
 Terminar como el rosario de la aurora (25, ++)  
 Quedar como una patena (24, ++)  
 Ser la madre del cordero (18, ++)  
 Irse el santo al cielo (18, ++)  
 Dios aprieta pero no ahoga (9, ++)  
 Comulgar con ruedas de molino (9, ++)  
 A Dios rogando y con el mazo dando (6, ++)  
 Llegar y besar el santo (6, ++)  
 Dios le da pan a quien no tiene dientes (1, ++)  
 Desnudar a un santo para vestir a otro (1, ++)  
 Fíjate de la virgen y no corras (1, ++)  
 Poner de mala hostia (42, +)  
 Ir/salir como alma que lleva el diablo (21, +)  
 Ser santo de la devoción (20, +)  
 Caérsele [a alguien] el alma a los pies (18, +)  
 Quedarse a dos velas (15, +)  
 No dar vela en este entierro (11, +)  
 Armar la de Dios es Cristo (5, +)  
 Ser un viva la virgen (5, +)

#### ITALIANO

Tenere a battesimo (128, ++)  
 Tutto questo ben di Dio (114, ++)  
 Avere un diavolo per capello (23, ++)  
 Nessuno è profeta in patria (20, ++)  
 Accontentarsi di quello che passa il convento (17, ++)  
 Saperne una più del diavolo (17, ++)  
 Avere il diavolo in corpo (12, ++)  
 Da oratorio (8, ++)  
 Non stare né in cielo né in terra (1, ++)  
 Dio sa come (22, +)  
 Che Dio la manda (8, +)  
 A Dio piacendo (8, +)

188. Como se ejemplificó en 2.3.3.

189. Según los datos obtenidos con el método Data-driven.

190. El proyecto BADARE, dirigido por el Dr. José Antonio Gargallo en la Universitat de Barcelona, ofrece ejemplos muy esclarecedores al respecto.

En estos dos listados tenemos **24 expresiones en español** frente a **12 en italiano**, lo cual representa casi un cuarto de las pirámides en ambas lenguas. Se considera que esta diversidad se debe no solo a eventos históricos que hayan tenido lugar en un país y no en el otro (historicidad), sino sobre todo a la **creatividad** de los hablantes, quienes desempeñan un papel de **artista creador**, como se vio en 2.2.4 con la teoría del idealismo lingüístico de Karl Vossler. A veces, se acuña una determinada expresión a partir de un hecho histórico (historicidad), como por ejemplo *terminar como el rosario de la aurora*, y luego la propia comunidad lingüística decide si esa expresión va a calar en la lengua y se va a utilizar en el futuro. En el caso de *comulgar con ruedas de molino* se presenta el contraste entre la hostia y la rueda del molino, objeto absolutamente imposible de tragar, como metáfora ante una situación en la que intentan hacernos creer algo imposible. Estas dos expresiones son muy usadas en la lengua española, pero no tienen ningún equivalente en italiano: la primera porque, según parece, se acuñó a partir de un hecho histórico concreto que ocurrió en España<sup>191</sup>; la segunda porque procede de una metáfora producida por semejanza (2.4.3) que se presta a la creatividad de una determinada comunidad lingüística y que trasciende la mera influencia de la cultura en la lengua.

También cabe señalar que en italiano es bastante frecuente la expresión *l'abito non fa il monaco* (26, ++), cuyo equivalente español *el hábito no hace al monje* apenas presentó 4 ocurrencias tras la búsqueda en los corpus (fase 3, como se vio en 4.2.3.1) y ni siquiera se tuvo en cuenta para las pirámides de la religión al no cumplir el tercer criterio de selección (4.3.4.1). Se trata de una equivalencia total que aparece en la mayoría de los diccionarios fraseológicos, pero siguiendo el método Data-driven y dejándose guiar únicamente por los datos, se hace evidente la gran diferencia de uso que existe entre el español y el italiano con esta expresión, lo que parece avalar la teoría de Dell Hymes (2.2.3), según la cual las diferencias culturales entre lenguas no se dan tanto en su estructura como en su uso.

Finalmente, **el hecho de que una expresión se asiente y se trasmite entre los hablantes es posible solo gracias a la imitación y a la aceptación de la misma por parte de nuestro cerebro; es decir, gracias a la función de las neuronas espejo**, de ahí que se les dedicara el apartado 2.3. La cultura compartida y el intercambio de información entre dos comunidades lingüísticas no siempre se ven plasmados en las expresiones, y a menudo resulta difícil predecir la evolución de una lengua, especialmente por lo que se refiere a las expresiones relacionadas con un ámbito cultural concreto. Gracias al *brainstorming* y a los corpus (primeros dos pasos de la triangulación metodológica) se pueden extrapolar estas expresiones y visualizarlas en las pirámides (tercer paso) para analizar puntos afines y divergentes en un determinado período histórico, que para esta tesis han sido los quince años que van desde 2001 hasta 2015.

En el caso del italiano y del español, la representación visual de datos, repartida en cinco pirámides de la religión, **demuestra que la afinidad entre las dos lenguas es solo parcial (objetivo C), y que existen un buen número de expresiones que no tienen equivalente en la otra lengua, ni semántico ni sintáctico**, bien por la influencia de hechos históricos, bien

191. Existen varias teorías acerca del origen de esta expresión, y la más difundida afirma que en un pueblo gaditano, durante el rosario que se rezaba justo antes de la salida del sol (aurora), dos cofradías enemistadas coincidieron por un paso estrecho, y la tensión por ver quién pasaría primero desencadenó en una gran trifulca donde muchas personas resultaron heridas.

por la creación de expresiones a partir de metáforas y metonimias que se prestan al libre albedrío de una determinada comunidad lingüística. Como ya se señaló en 4.2.4.1 y en 4.3.4.1, somos conscientes de que con la religión se ha obtenido una gran cantidad de información cuantitativa que a menudo resulta difícil de gestionar y que tal vez se podría someter a ulteriores criterios de selección para que el número de pirámides sea inferior y el análisis contrastivo resulte más claro<sup>192</sup>. Este problema no subsiste con el fútbol y con los toros, donde la visualización de datos está mucho más acotada en ambas lenguas, como se verá a continuación.

#### 4.4.2. LAS PIRÁMIDES ETNOLINGÜÍSTICAS DEL FÚTBOL EN CONTRASTE

El análisis contrastivo entre las pirámides del fútbol va a resultar, en principio, más sencillo, y eso se debe básicamente a la menor cantidad de datos que se manejan. Mediante la representación visual de las pirámides, se pretende comparar las expresiones obtenidas en la fase 4 (4.2.4.2 y 4.3.4.2) para detectar el grado de afinidad que existe entre el español y el italiano en el período seleccionado<sup>193</sup> y con un ámbito cultural compartido en las dos lenguas. Veamos las dos pirámides juntas en las siguientes páginas:

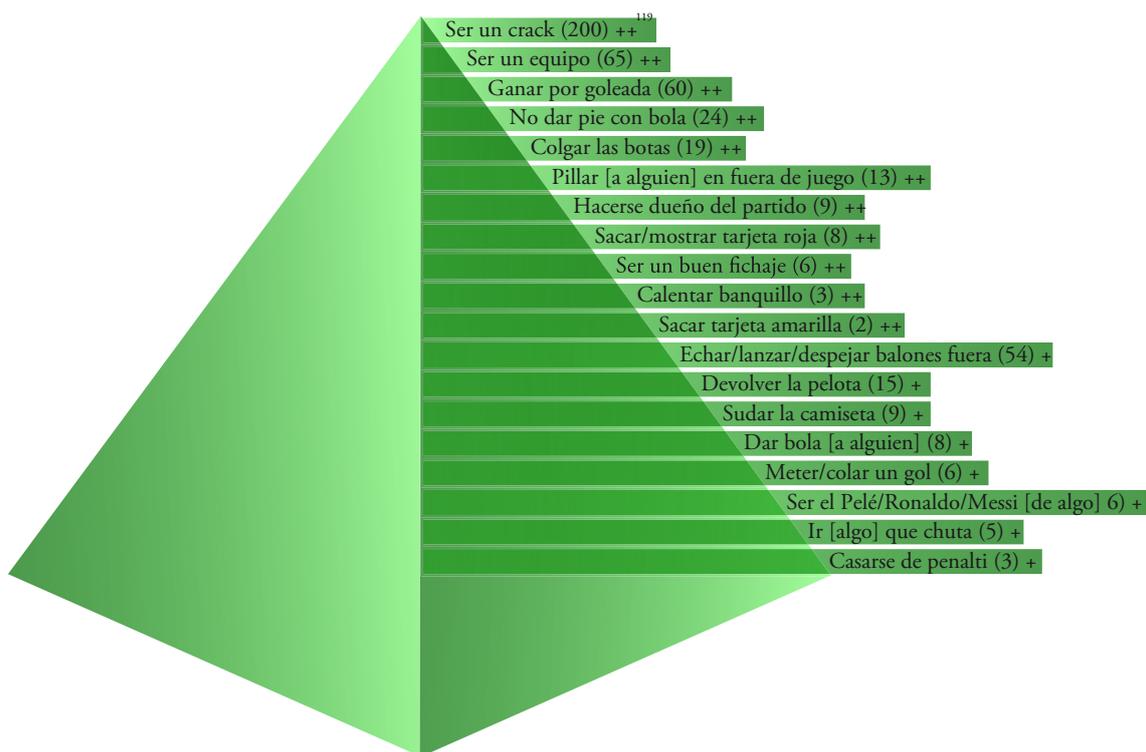


Gráfico 15. Pirámide 1/1 del fútbol en español

192. Se volverá sobre ello en 5.1 y 5.3.

193. 2001-2015.

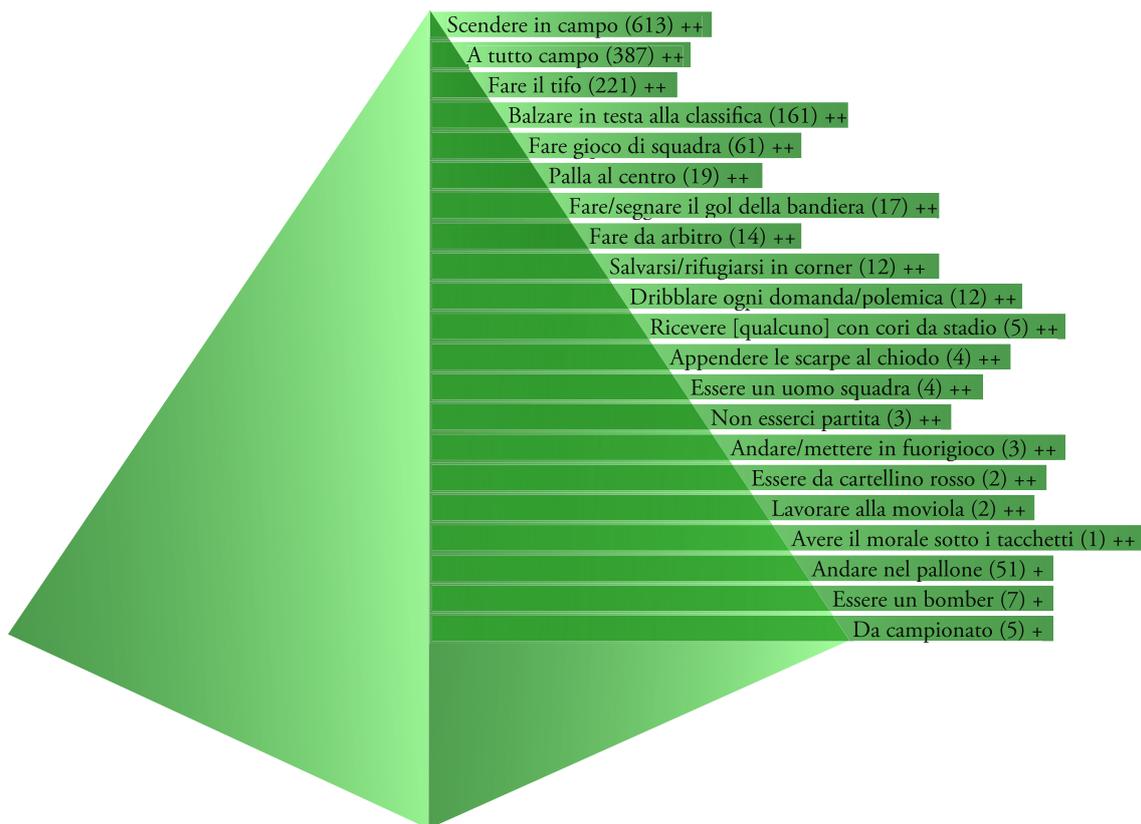


Gráfico 16. Pirámide 1/1 del fútbol en italiano

Como se puede apreciar, el número de expresiones es mucho más parejo en comparación con la pirámide de la religión, y tenemos **19 expresiones en español frente a 21 en italiano**. La primera reflexión que se podría hacer llegados a este punto es la siguiente: el fútbol lleva poco más de cien años en contacto con ambas lenguas, y se podría deducir, a primera vista, que la relación bidireccional lengua-cultura de la que se habló en 2.1 es casi idéntica, al menos por los que se refiere al aspecto cuantitativo. Sin embargo, al observar las dos pirámides, se pueden detectar algunos aspectos que vale la pena analizar. En primer lugar, **en español solo tenemos una expresión con una frecuencia muy alta (*ser un crack*, 200, ++), mientras que en italiano hay cuatro que superan las 150 ocurrencias y que encabezan la pirámide: *scendere in campo* (613, ++), *a tutto campo* (387, ++), *fare il tifo* (221, ++) y *balzare in testa alla classifica* (161, ++)**. Comparando la primera expresión de la pirámide española con la primera de la italiana, se ve además que hay una diferencia considerable (200 ocurrencias frente a 613). Esto podría indicar, de forma somera, que en italiano la influencia del fútbol en la lengua es particularmente marcada y hay algunas expresiones muy frecuentes y de uso diario entre los hablantes, mientras que en español parece que la influencia es inferior.

En segundo lugar, **tenemos pocas expresiones afines o con equivalencia total/parcial**<sup>194</sup>, como es el caso de *ser un equipo* (65, ++) frente a *fare gioco di squadra* (61 ++), *colgar las botas* (19, ++) frente a *appendere le scarpe al chiodo* (4, ++), *sacar/mostrar tarjeta roja*

194. Entendiéndose principalmente desde el punto de vista semántico, y no morfosintáctico.

(8, ++) frente a *essere da cartellino rosso* (2, ++) y *ser un crack* (200, ++) frente a *ser un bomber* (7, ++), aunque en este último caso la equivalencia semántica es solo parcial, ya que en italiano esta expresión se utiliza solo en contextos informales y no ha tenido tanto impacto como en español, quedando todavía muy vinculada solo al entorno futbolístico y siendo la pirámide prueba fehaciente de ello.

En tercer lugar, tenemos algunas expresiones aparentemente similares, pero con un significado semántico diferente, como *pillar [a alguien] en fuera de juego* (13, ++) frente a *mettere [qualcuno] in fuorigioco* (3, ++), donde la acción del sujeto es diferente, y *despejar balones fuera* (54, +) frente a *salvarsi in corner* (12, ++), ya que la expresión española remite a alguien que intenta esquivar una situación, mientras que la italiana hace referencia a alguien que se ha salvado por los pelos en una situación comprometida.

Por último, nos encontramos, igual que con la religión, con varias expresiones que no tienen un equivalente en la otra lengua<sup>195</sup>. Veamos algunos ejemplos:

#### ESPAÑOL

Ganar por goleada (60, ++)  
 No dar pie con bola (24, ++)  
 Hacerse dueño del partido (9, ++)  
 Ser un buen fichaje (6, ++)  
 Calentar banquillo (3, ++)  
 Sacar tarjeta amarilla (2, ++)  
 Devolver la pelota (15, +)  
 Sudar la camiseta (9, +)  
 Dar bola [a alguien] (8, +)  
 Colar un gol (6, +)  
 Ser el Pelé/Ronaldo/Messi [de algo] (6, +)  
 Ir [algo] que chuta (5, +)  
 Casarse de penalti (3, +)

#### ITALIANO

Scendere in campo (613, ++)  
 A tutto campo (387, ++)  
 Fare il tifo (221, ++)  
 Balzare in testa alla classifica (161, ++)  
 Palla al centro (19, ++)  
 Segnare il gol della bandiera (17, ++)  
 Fare da arbitro (14, ++)  
 Salvarsi in corner (12, ++)  
 Dribblare ogni polemica (12, ++)  
 Ricevere [qualcuno] con cori da stadio (5, ++)  
 Essere un uomo squadra (4, ++)  
 Non esserci partita (3, ++)  
 Lavorare alla moviola (2, +)  
 Avere il morale sotto i tacchetti (1, ++)  
 Andare nel pallone (51, +)  
 Da campionato (5, +)

En total, tenemos **13 expresiones en español** y **16 en italiano** sin un equivalente en la otra lengua, lo que representa más de dos tercios de cada pirámide. Este resultado parece avalar la hipótesis de partida (1.2) de que la afinidad etnolingüística entre estas dos lenguas partiendo de **un brainstorming (primer paso de la triangulación metodológica)** y realizando luego una **búsqueda en los corpus (segundo paso)** y una **visualización de los datos (tercer paso)** existe solo desde un punto de vista cuantitativo y numérico, puesto que hay muchas diferencias por lo que se refiere al aspecto cualitativo, que se refleja en el tipo de expresiones. Esto enlazaría con el tercer objetivo de esta tesis (C), y apuntaría, una vez más, al hecho de que la cosmovisión de una determinada comunidad lingüística (2.2), junto con el papel de artista creador que tienen los hablantes (2.2.4), influyen de manera determinante en la forja de nuevas expresiones, y la aceptación de las mismas depende principalmente de las neuronas espejo

195. Según los datos obtenidos con el método Data-driven.

(2.3). Por lo tanto, aunque dos lenguas compartan un mismo ámbito cultural, no siempre se tienen las mismas expresiones y, como se acaba de ver con el fútbol, que forma parte tanto de la cultura adquirida como de la compartida en español y en italiano (2.2), son más las diferencias que los aspectos en común. Paolo Balboni habla de palabras en común y de culturas diferentes para referirse a la competencia comunicativa en un contexto de interculturalidad (*Parole comuni, culture diverse. Guida alla comunicazione interculturale*, cap. 1); aquí, en cambio, nos encontramos justamente con lo contrario, y la visualización de datos en las pirámides ofrece una imagen clara de que **hay culturas en común, pero palabras diferentes, lo que corresponde al tercer objetivo (C) de este trabajo de investigación.**

Finalmente, está claro que el fútbol tiene en la actualidad una presencia muy fuerte en la sociedad, y todo apunta a que dicha presencia seguirá aumentando. Por eso mismo, no se descarta que dentro de unos años, con el auge cada vez mayor de la tecnología y de los medios de comunicación, las dos pirámides del fútbol se amplíen y se enriquezcan con nuevas expresiones. También confiamos en una evolución de las herramientas lingüísticas, especialmente en italiano, ya que con un corpus lematizado se podrán realizar unas búsquedas mucho más afinadas y certeras que sin duda agilizarán y mejorarán todo el proceso de visualización<sup>196</sup>.

En tercer y último lugar, se cotejarán las pirámides de los toros, donde se verá, mediante la representación visual, lo que ocurre con un ámbito cultural que tiene una influencia muy marcada en español y casi nula en italiano.

#### **4.4.3. LAS PIRÁMIDES ETNOLINGÜÍSTICAS DE LOS TOROS EN CONTRASTE**

Con la religión y el fútbol se detectaron bastantes diversidades entre la lengua española y la italiana, especialmente por los que se refiere al aspecto cualitativo y al grado de equivalencia de las expresiones obtenidas tras la búsqueda en los corpus (fase 3); pese a ello, el aspecto cuantitativo ha sido bastante equilibrado, debido a que se trata de dos ámbitos culturales que están compartidos por ambas comunidades lingüísticas. El mundo taurino, en cambio, es característico solo de la lengua española, como se puede apreciar a partir de la visualización de estas dos pirámides:

---

196. Según los datos obtenidos con el método Data-driven.

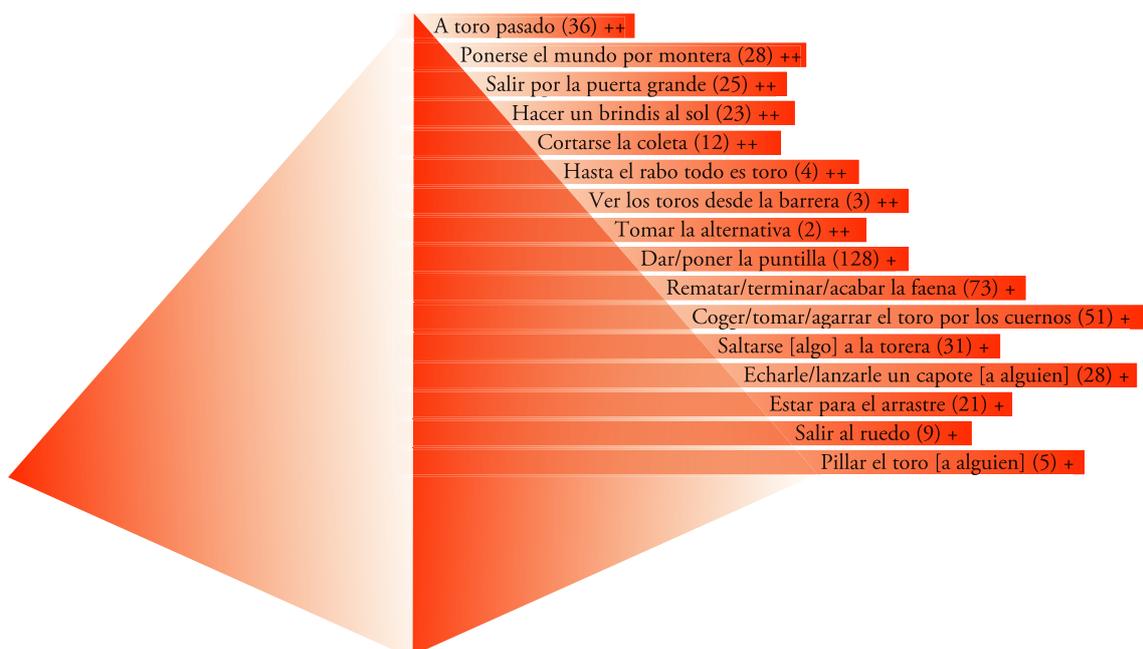


Gráfico 17. Pirámide 1/1 de los toros en español

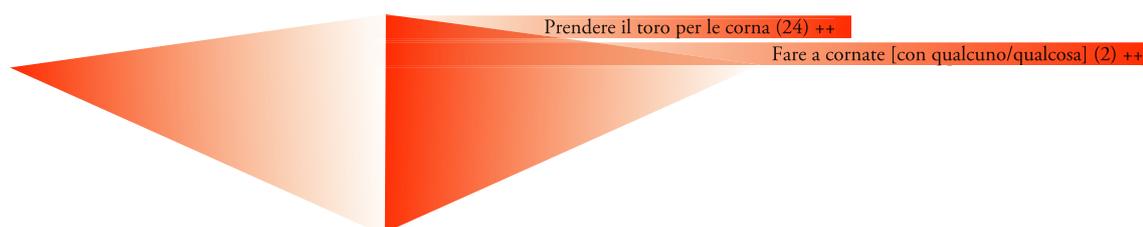


Gráfico 18. Pirámide 1/1 de los toros en italiano

La representación visual de las pirámides permite detectar al instante una diferencia evidente entre las dos lenguas, con **15 expresiones en español** frente a **2 en italiano**, de las cuales *prendere el toro per le corna* (24, ++) es una equivalencia total de *coger al toro por los cuernos* (51, ++), mientras que *fare a cornate* presenta tan solo 2 ocurrencias y su aparición se puede deber a interferencias culturales y lingüísticas; es decir, a la influencia de la cultura compartida (2.2). Este análisis contrastivo refleja visualmente el hecho de que el alcance del artista creador (2.2.4) se ve muy reducido si un ámbito cultural concreto no forma parte de su cosmovisión. En el caso de la pirámide en español, hay varias expresiones que reflejan claramente las huellas de la cultura en la lengua y que no tienen ningún tipo de equivalencia en italiano. Por ejemplo, con *ponerse el mundo por montera* (28, ++) se asocia el gesto del torero que se pone la montera antes del combate con la idea de tomar una decisión sin importar lo que digan los demás. Se podría considerar como un caso de metonimia por contigüidad (2.4.3), donde el gesto del torero representaría la valentía en la vida real. Otros casos destacables son *echarle/lanzarle un capote [a alguien]* (28, +), que ha ampliado su

significado pasando a designar cualquier tipo de ayuda, incluso fuera del ruedo, y *estar para el arrastre* (21, +), una hipérbole muy frecuente en la lengua que se utiliza para expresar un gran estado de cansancio, y en la que el hablante se compara con el toro muerto al final de la lidia, estando listos los dos para ser arrastrados fuera de la arena. Por último, vale la pena mencionar *salir al ruedo* (9, +), utilizada cuando alguien se dispone a pasar a la acción. Es cierto que en italiano no existe un equivalente relacionado con el mundo taurino, pero sí con el mundo del fútbol, ya que se encontró en el corpus *scendere in campo* (613, ++), una expresión que ocupa la parte más alta de la pirámide del fútbol en italiano, como se vio en 4.3.4.2 y en 4.4.2. A este respecto, también se mencionó en 4.3.3.3., en una nota al pie, que en italiano se utiliza la palabra *incornata* para referirse a un cabezazo hacia la portería rival durante un partido. No se tuvo en cuenta para la pirámide porque procede de otro ámbito cultural, pero estos casos son una prueba clara de lo flexible y elástica que puede llegar a ser la malla lingüística de una lengua viva. Con la representación visual de las pirámides se pretende ofrecer una visión en conjunto de estos aspectos, lo que corresponde al tercer objetivo (C) de esta tesis de investigación.

## 4.5. RESUMEN DE LA PARTE II

Hemos llegado al final de la segunda parte de la tesis, en la que se ha tratado de alcanzar los tres objetivos principales (A, B y C) que se presentaron en 1.2 poniendo en práctica la metodología mixta que sigue la idea de triangulación propuesta por Zoltan Dornyei (*Research Methods in Applied Linguistic: Quantitative, Qualitative, and Mixed Methodologies*, cap. 2). Dicha metodología se concreta en tres pasos: el *brainstorming*, la búsqueda en los corpus y la visualización de datos en las pirámides. A partir de aquí, se dedicó gran parte del capítulo 4 a explicar las cuatro fases para crear las pirámides etnolingüísticas en español (4.2) y en italiano (4.3) con el fin de encontrar el mayor número posible de expresiones relacionadas con el ámbito cultural en cuestión (objetivo A), visualizar toda esa información en las pirámides (objetivo B) y, por último, demostrar que la cultura influye de manera diferente en cada lengua a partir de la representación visual y el análisis contrastivo de las expresiones obtenidas tras la búsqueda en los corpus (objetivo C).

A lo largo de esta *Parte II* se ha manejado mucha información cuantitativa, especialmente con el ámbito de la religión en 4.2.3.1 y en 4.3.3.1, donde la visualización de datos se ha tenido que repartir en varias pirámides debido al elevado número de expresiones que se encontraron en los corpus. Así pues, con el fin de mejorar el aspecto visual de los datos dentro de las pirámides, se establecieron una serie de criterios en 4.2.4 y en 4.3.4 para seleccionar la información cualitativa que más interesaba en vista de los tres objetivos de esta tesis (A, B y C).

Finalmente, se dedicó todo el apartado 4.4 al análisis contrastivo entre el español y el italiano partiendo de la comparación visual de las pirámides. Aquí se han retomado los conceptos teóricos expuestos en la primera parte para analizar el grado de influencia que ejercen los tres ámbitos culturales escogidos (religión, fútbol y toros) en español y en italiano, así como el grado de afinidad que existe entre las expresiones obtenidas tras la búsqueda en los corpus. En 4.4.1 y en 4.4.2 se observó que con ámbitos culturales afines como la religión y el fútbol la información cuantitativa es bastante similar, pero hay diferencias considerables por lo que se refiere a la información cualitativa (equivalencia semántica entre las expresiones). En cambio, el mundo taurino nunca ha tenido particular relevancia en la comunidad lingüística italiana, y en 4.4.3 se pudo apreciar la diferencia que existe entre el español y el italiano en cuanto al número de expresiones, demostrando la hipótesis de partida que se planeó en 1.2.

Esta tesis se cerrará con una *Parte III*, que incluirá las consideraciones finales donde se sopesarán las aportaciones y limitaciones de la investigación, además de señalar las líneas que hayan podido quedar abiertas y que puedan servir de inspiración para futuros proyectos en el campo de la etnolingüística.





# **PARTE III**

## **5. CONSIDERACIONES FINALES**



# **5. CONSIDERACIONES FINALES**

## 5.1. APORTACIONES Y LIMITACIONES DE LA INVESTIGACIÓN

Con la investigación llevada a cabo en esta tesis se ha querido proponer un nuevo sistema de presentación de expresiones culturalmente marcadas a partir de unos conceptos teóricos sobre etnolingüística, semántica, neurociencia y comunicación que se han expuesto en la primera parte. El presente trabajo de investigación se gestó a raíz de **dos preguntas iniciales que se formularon** en 1.2: por un lado, **cómo se podría utilizar la etnolingüística para comparar dos lenguas romances como el español y el italiano**; por otro, **cómo se manejaría la información cuantitativa y cualitativa**. Para contextualizar la primera pregunta, se optó primero por acotar el ámbito cultural, lo que llevó a seleccionar tres áreas concretas: **la religión, el fútbol y los toros**. Dicha elección se basaba en la idea de cotejar dos ámbitos culturales compartidos y uno que solo fuera característico de una lengua, como es el caso del mundo taurino para el español. Respecto a la segunda pregunta, era menester acceder a una gran cantidad de datos que reflejaran el uso real de la lengua, por eso se barajó la posibilidad de apoyarse en los corpus lingüísticos para realizar las búsquedas y acceder tanto a la información cuantitativa como cualitativa.

Estas dos preguntas de partida nos han llevado a establecer **tres grandes objetivos** en 1.2 que servirían para emprender el proceso investigador y ofrecer una aportación en el ámbito etnolingüístico. El **primer objetivo (A)**, partiendo de la **hipótesis** de que es en las expresiones donde más se puede apreciar la influencia de la cultura en la lengua, **consistía en encontrar el mayor número posible de expresiones relacionadas con los tres ámbitos culturales seleccionados** y, como se vio en 4.2.3 y en 4.3.3, se pudo alcanzar tras un paso previo: la realización de un *brainstorming* (fase 1) con informantes nativos<sup>197</sup>. Este fue el **primer paso** de la metodología mixta basada en la triangulación que se ha utilizado en este trabajo de investigación y que ha permitido solventar aspectos relacionados con la recaudación y visualización de datos (Data-driven). Mediante el *brainstorming* se obtuvo un gran número de palabras que se sometieron a una criba (fase 2) con el fin de conseguir información cualitativa para la búsqueda en los corpus (fase 3). Dicha búsqueda representó el **segundo paso** de la triangulación metodológica y ha sido un puente fundamental que nos ha permitido ir de las ocurrencias a las expresiones en cada uno de los tres ámbitos escogidos. Esto demostraría, como **primera aportación**, la **hipótesis** inicial de que **la combinación del brainstorming con la búsqueda en los corpus permite acceder a información cuantitativa y cualitativa**. Es cierto que para realizar las búsquedas se necesita al menos un corpus general que sea lo más representativo posible de la lengua en cuestión y que, a ser posible, esté lematizado. En el caso del español, se escogió el CORPES, el corpus de la Real Academia Española que, pese a encontrarse todavía en versión beta en el período en que se realizaron las búsquedas (2016), está lematizado y presenta una interfaz muy cómoda y sencilla. Para el italiano, en cambio, se recurrió al CO-RIS/CODIS, un corpus que recoge solo el canal escrito y que no está lematizado, dos factores que dificultaron bastante la tarea de búsqueda. El primer problema se solventó realizando

197. 5 para el español y 5 para el italiano, como se detalla en 4.2.1. y en 4.3.1.

búsquedas complementarias en el LIP<sup>198</sup>, un corpus de italiano oral, mientras que el segundo obstáculo nos obligó a buscar todas las formas de las palabras del listado, complicando la obtención de las expresiones. Esta es quizá **una primera limitación de la investigación**, y demuestra lo difícil que es comparar dos corpus generales con características diferentes para realizar un mismo tipo de búsqueda.

Estos resultados enlazaron con el **segundo objetivo (B)** de esta tesis, con el que se pretendía **ordenar y visualizar la información cuantitativa y cualitativa para poder realizar un análisis contrastivo entre el español y el italiano**. Es evidente que hoy en día vivimos un auge tecnológico sin precedentes, y el aspecto visual ha ido ganando mucho terreno en todos los ámbitos, por eso se decidió enmarcar las expresiones seleccionadas dentro de unas pirámides y ordenarlas de acuerdo con dos criterios principales: el de **frecuencia** y el de **rentabilidad comunicativa**. Esta visualización de datos representa el **tercer paso** y el cierre de la triangulación metodológica que empezó con el *brainstorming* y que siguió con la búsqueda en los corpus. Asimismo, la presentación visual de las pirámides etnolingüísticas constituye la cuarta y última fase del proceso de creación que se vio en 4.2.4 y en 4.3.4. Por lo tanto, la **segunda aportación** de esta tesis se ha materializado **en la visualización de las pirámides, donde ha sido posible reflejar tanto la información cuantitativa como cualitativa**, eje central del segundo objetivo (B). Pese a ello, se vio con la religión en 4.2.4.1 y en 4.3.4.1 que se disponía de demasiada información cuantitativa, por lo que nos vimos obligados a aplicar estrictos criterios de selección en vista de la visualización final. No obstante, y a diferencia del fútbol y de los toros, el ámbito de la religión dio lugar a tres pirámides para el español y dos para el italiano, lo que sobrepasa el número de pirámides previstas, **limitando en parte una visualización que se considera ideal cuando solo hay una pirámide por lengua**.

Por último, se dedicó todo el apartado 4.4 al **tercer objetivo (C)**, que consistía en **demostrar, mediante un análisis contrastivo a partir de la representación visual de las pirámides, la hipótesis de que la cultura influye de manera diferente en cada lengua y que, incluso con ámbitos culturales afines como la religión y el fútbol, son más las expresiones diferentes que las similares, puesto que influyen muchos factores, entre otros la creatividad de los hablantes, el proceso de imitación y la historicidad**. Para alcanzar este tercer objetivo se recurrió en varias ocasiones a los conceptos que se expusieron en la primera parte y que perfilan el marco teórico de toda la investigación. Se pudieron observar diferencias considerables entre el italiano y el español por lo que se refiere al grado de equivalencia semántica de las expresiones y se vio claramente que con la religión y con el fútbol el número de expresiones era bastante parejo, pero más de un tercio no presentaban ningún equivalente en la otra lengua. Así pues, la **representación visual de datos** ha hecho posible que, a partir de conceptos semánticos y de neurociencia, se pudiera llevar a cabo un análisis contrastivo entre el español y el italiano, que se traduce en la consecución del tercer objetivo (C) del presente trabajo.

Por último, cabe mencionar que estos tres objetivos han compartido **la voluntad de ofrecer un nuevo enfoque en el campo de la etnolingüística a partir de una herramienta**

198. Se decidió utilizar este corpus como apoyo al CORIS pese a que sus muestras son de mediados de los años noventa, como se vio en 3.3.3. Somos conscientes de ello y lo asumimos como limitación de esta investigación.

visual que facilite tanto el aprendizaje de la lengua como la inmersión cultural. La aportación general de dicho enfoque, pese a las limitaciones mencionadas arriba, consiste en haber demostrado la hipótesis de que es posible servirse de un *brainstorming* y de los corpus con fines etnolingüísticos para detectar la influencia real de un ámbito cultural concreto en la lengua, y que esa influencia se puede apreciar gracias a la visualización de las pirámides.

## 5.2. CONCLUSIONES

Tras analizar las aportaciones y limitaciones de la investigación, se ha llegado a la parte conclusiva de esta tesis, en la que se ha querido proponer, a raíz de nuestra pasión por la etnolingüística, una nueva herramienta de visualización de datos que permita ahondar más en la relación lengua-cultura, detectando patrones que pueden ser de ayuda tanto en el ámbito contrastivo como en la enseñanza de una L2. Como se vio en 1.2, el presente trabajo se ha articulado a partir de tres hipótesis que constituyen el eje vertebrador de la investigación y que han llevado a las siguientes conclusiones:

- 1) Que es en las expresiones donde se puede apreciar la influencia de la cultura en la lengua;
- 2) Que es posible utilizar los corpus para demostrar que la afinidad etnolingüísticas entre dos lenguas aparentemente similares como el español y el italiano es solo parcial, y que dicha influencia se puede visualizar en las pirámides etnolingüísticas;
- 3) Que, con ámbitos culturales no compartidos, como en el caso de los toros entre el español y el italiano, la visualización de las expresiones presenta grandes diferencias.

Para llegar a estas conclusiones, ha desempeñado un papel fundamental la metodología escogida (1.4); por un lado, se decidió utilizar una metodología mixta basada en la triangulación de datos (*brainstorming*, búsqueda en los corpus y visualización en las pirámides); por otro, se optó por seguir el método inductivo Data-driven, con el que se pretendía demostrar las hipótesis iniciales a partir de la recopilación y análisis de datos lingüísticos, y sin recurrir a diccionarios fraseológicos en busca de las expresiones. Así, pues, y tras realizar el proceso investigador a lo largo de tres años, se ha podido corroborar, pese a algunas limitaciones que se analizaron en 5.1, que las expresiones contienen mucha información etnolingüística (conclusión 1), y que gracias al *brainstorming* y a la búsqueda en los corpus es posible apreciar que el grado de afinidad entre dos lenguas romances como el español y el italiano es parcial por los que se refiere a la religión y al fútbol (conclusión 2), y casi nulo con el ámbito cultural de los toros (conclusión 3). Cabe señalar que para llegar a estas conclusiones y completar el proceso investigador también han resultado de ayuda algunos seminarios a los que asistí en el período de redacción de esta tesis, y de los que cabe destacar *Primeras jornadas de formación para profesores de idiomas: aprender del pasado y enseñar hacia el futuro*, organizado por la Universitat de Barcelona (28 de enero de 2017), *Jornada didáctica de formación para profesores de ELE*, organizado por Sociedad General Española de Librería (SGEL) y la Universitat de Barcelona (10 de marzo de 2017), con conferencias de neurociencia y de gramática cognitiva, y *4º Incontro pratico per insegnanti di italiano*, organizado por la Escola d'Idiomes Moderns de la Universitat de Barcelona (29 de abril de 2017).

A modo de cierre, con esta tesis se ha querido demostrar también que la cultura y la lengua no solamente ejercen entre ellas una influencia bidireccional que se ve plasmada en el uso real de los hablantes, sino que además existe un factor impredecible relacionado con el papel del artista creador y con la aceptación de una determinada expresión que depende muchas

veces de factores sociales, y por supuesto de la empatía, dirigida en todo ser humano por las neuronas espejos. Pensamos que la mejor manera de aprender y, sobre todo, de comprender una lengua es con una gran inmersión lingüística y cultural que permita entender más a fondo la cosmovisión de sus hablantes. La etnolingüística presenta nuevos recursos con los que se podría ahondar mucho más en la relación lengua-cultura, ofreciendo puntos de vista novedosos y esclarecedores sobre el impacto de la cultura en una determinada comunidad lingüística. También queremos recalcar, una vez más, el papel fundamental de la neurociencia en este trabajo de investigación. El cerebro humano sigue siendo, en ciertos aspectos, un gran desconocido para el ser humano, y los avances de las últimas décadas subrayan la importancia de algunos mecanismos cognitivos como la imitación o la actividad del sistema límbico, verdaderos motores del aprendizaje y de la integración sociocultural.

### 5.3. LÍNEAS ABIERTAS DE INVESTIGACIÓN

Los resultados de la investigación sugieren una serie de líneas abiertas de investigación, especialmente por lo que se refiere a la obtención y visualización de datos en las pirámides etnolingüísticas. En primer lugar, y como se apuntó al final de 4.2.4.1 y 4.3.4.1, la información cuantitativa obtenida a partir de la búsqueda en los corpus (fase 3) fue considerable, y se tuvieron que aplicar una serie de criterios para que el resultado visual resultara más esclarecedor. Pese a ello, se tuvieron que repartir los datos en tres pirámides para el español y dos para el italiano, lo que sigue siendo un número algo abultado que dificulta una visualización óptima. Tal vez, de cara a nuevas líneas de investigación, se podrían ampliar aún más los criterios de selección para conseguir una sola pirámide.

En segundo lugar, se escogieron dos corpus generales para realizar la búsqueda de las ocurrencias (fase 3): el CORPES para el español y el CORIS/CODIS para el italiano. El primero está lematizado, pero el segundo no, lo que dificultó bastante el proceso de búsqueda. También nos apoyamos en el Val.Es.Co. para el español y en el LIP para el italiano, además de la web como corpus, pero no se descarta la posibilidad de haberse dejado alguna expresión por el camino, así que se podría realizar el mismo trabajo con otros corpus generales, como por ejemplo el Corpus del español o el Corpus Paisà, y tal vez el resultado final de la pirámide sea algo distinto.

También animamos a futuros investigadores a seguir esta línea para ampliar y mejorar este camino de contraste etnolingüístico, incluso seleccionando y comparando otras lenguas, ya que el proceso seguiría los mismos pasos, incluso entre lenguas diferentes que no procedan del latín. La única condición es poder realizar el *brainstorming* con informantes nativos y tener acceso al menos a un corpus general de la lengua en cuestión.

Creemos que, por tratarse de un proyecto totalmente nuevo que propone un modelo algo atípico, como es la visualización de datos en las pirámides etnolingüísticas a partir de un *brainstorming* y de la búsqueda en los corpus, presenta un amplio margen de mejora y, en este sentido, deseáramos que nuestra aportación inspirara futuras investigaciones que partieran de esta propuesta etnolingüística. Así pues, confiamos en que estas ideas despierten la curiosidad de nuevos investigadores. Como comentó el Dr. Francisco Mora en la conferencia sobre Neuroeducación que impartió en la Universitat de Barcelona el 10 de marzo de 2017<sup>199</sup>, “no hay pensamiento sin el fuego emocional que lo alimente, ni creatividad sin equivocación, y la verdadera curiosidad estará siempre unos pasos más allá de las limitaciones impuestas en los mapas”.

199. Seminario Educación, comunicación y gramática II, en ocasión del II Foro Universidad de Barcelona-SGEL.



# BIBLIOGRAFÍA

- ABELLA, Carlos. *¡Derecho al toro! El lenguaje taurino y su influencia en lo cotidiano*. Madrid: Anaya y Mario Muchnik, 1996.
- ALINEI, Mario. *La struttura del lessico*. Bologna: Il Mulino, 1974.
- ALMELA, Ramón; WOTJIAK, Gerd y RAMÓN TRIVES, Estanislao. *Fraseología Contrastiva: Con Ejemplos tomados del Alemán, Español, Frances e Italiano*. Murcia: Edicum, 2005.
- ALONSO, Amado. *Estudios lingüísticos. Temas españoles*. Madrid: Gredos, 1982.
- ALONSO JIMÉNEZ, Elisa. "Linguee y las nuevas formas de traducir." *Skopos: revista internacional de traducción e interpretación*, 2: 5–28, 2013. [http://rio.upo.es/xmlui/bitstream/handle/10433/851/2013\\_Alonso\\_Linguee\\_Skopos.pdf?sequence=1](http://rio.upo.es/xmlui/bitstream/handle/10433/851/2013_Alonso_Linguee_Skopos.pdf?sequence=1)
- ALONSO PÉREZ-ÁVILA, Elena. "El Corpus lingüístico en la didáctica del léxico del español como LE." *Boletín de ASELE (Asociación para la Enseñanza del Español como Lengua Extranjera)*, 37: 11–25, 2007. <http://formespa.rediris.es/pdfs/asele37.pdf>
- AMBADY, Nalini y ROSENTHAL, Robert. "Half minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness". *Journal of Personality and Social Psychology*, 64: 431-441, 1993.
- ANGLADA ARBOIX, Emília. *Lexicografía española*. Barcelona: Universitat de Barcelona, 2005.
- ARBIB, Michael. "From monkeylike action recognition to human language: An evolutionary framework for neurolinguistics", *Behavioural and Brain Science*, 28: 105-124, 2005.
- ARIÑO, Antonio. *Sociología de la cultura. La constitución simbólica de la sociedad*. Barcelona: Ariel, 2000.
- AUSTIN, John Langshaw. *Come fare cose con le parole*. Torino: Marietti, 1987. Ed. or. 1962.
- BALBONI, Paolo. *Parole comuni, culture diverse. Guida alla comunicazione interculturale*. Venezia: Marsilio, 1999.
- BALCONI, Michela. *Neuropsicologia delle emozioni*. Roma: Carocci, 2004.
- BARCROFT, Joe. *Input-Based incremental vocabulary instruction*. Alexandria, VA: TESOL International Association, 2012.
- BARQUERO CABRERO, José Daniel y ARBÓS BERTRÁN, Albert. *Relaciones públicas multiculturales: cómo evitar el choque de culturas y civilizaciones*. Barcelona: Furtwangen Editores, 2008.
- BARROS GARCÍA, Benamí y KHARNÁSOVA, Galina. "La interculturalidad como macrocompetencia en la enseñanza de lenguas extranjeras: Revisión Bibliográfica y Conceptual." *Porta Linguarum*, Revista: 97-114, 2012. <https://dialnet.unirioja.es/>

- descarga/articulo/4593111.pdf
- BENASSAR, Bartolomé. *Historia de la tauromaquia: una sociedad del espectáculo*. Valencia: Editorial Pre-Textos, 2000.
- BLAKEMORE, Sarah y FRITH, Uta. *Cómo aprende el cerebro: las claves para la educación*. Barcelona: Ariel, 2011.
- BOAS, Franz. (1928). *Anthropology and Modern Life*. Abingdon: Taylor Francis. Ed. de 2004.
- BOAS, Franz. (1911). *Introduction. Handbook of American Indian Languages*. Lincoln, University of Nebraska Press. Ed. de 1966.
- BORREGUERO ZULOAGA, Margarita y LUQUE, Luis (eds.). *Linguistica italiana in Spagna, linguistica spagnola in Italia. Studi italiani di Linguistica Teorica ed Applicata* XL (3). Roma: Paccini, 2011.
- BRIGHT, Jane y BRIGHT, William. "Semantic Structures in Northwestern California and the Sapir-Whorf Hypothesis". En: *Cognitive Anthropology*, New York: Holt and Winston, 66-77, 1970.
- BRIZ GÓMEZ, Antonio. *El español coloquial en la conversación: esbozo de pragmalinguística*. Barcelona: Ariel, 1998.
- BUYSE, Kris. "¿Qué corpus en línea utilizar para qué fines en la clase de ELE?". Del texto a la lengua: *La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE: 277-289*. Asociación para la Enseñanza del Español como Lengua Extranjera, 2011. <https://dialnet.unirioja.es/servlet/articulo?codigo=54193622>
- CABALLERO BENAVENTE, Mónica. *La variabilidad lingüística nativa y no nativa en escenarios comunicativos. La función de transacción en una situación cotidiana: corpus y descripción para el español (L1/LE)*. Tesis doctoral dirigida por la Dra. Lourdes Díaz Rodríguez y la Dra. Maria Taulé Delor. Barcelona: Universitat de Barcelona, 2015.
- CANNADA BARTOLI, Vincenzo. *Antropologia e linguistica*. Roma: Carocci, 2009.
- CARDONA, Giorgio Raimondo. *La linguistica antropologica*. Roma: Parole e metodi, 1973.
- CARDONA, Giorgio Raimondo. *I sei lati del mondo. Linguaggio ed esperienza*. Roma-Bari: Laterza, 1985.
- CARDONA, Giorgio Raimondo. *La trasmissione del sapere; aspetti linguistici e antropologici*. Roma: Bagatto, 1989.
- CARDONA, Giorgio Raimondo. *Introduzione all'etnolinguistica*. Novara: De Agostini Scuola SpA, 2006.
- CARNAP, Rudolf. *Meaning and Necessity. A Study in Semantics and Modal Logic*. University of Chicago Press, 1956.
- CASADO VELARDE, Manuel. *Lengua y cultura. La etnolingüística*. Madrid: Síntesis, 1988.
- CASADO VELARDE, Manuel. *Lengua e ideología*. Pamplona: Eunsa, 1978.
- CHEN, Chaomei. *Information Visualization: Beyond the Horizon*. 2nd ed. London:

- Springer, 2006.
- CHOMSKY, Noam. *Knowledge of language*. New York: Praeger, 1986.
- CHOMSKY, Noam. *Una aproximación naturalista a la mente y al lenguaje*. Barcelona: Prensa Ibérica, 1998.
- COLLETTI, Lucio. *Ideologia e società*. Bari: Laterza, 1969.
- COMAS, José. “Escándalo en Alemania por el paso de Schröder al sector privado”. *El País.com*. Madrid: 14 de diciembre de 2005.
- CORDERO MONGE, Sergio. *Unidades pluriverbales en los diccionarios del español de Costa Rica: análisis y propuesta de descripción*. Tesis doctoral dirigida por la Dra. Emília Anglada Arboix. Barcelona: Universitat de Barcelona, 2016.
- CORPAS PASTOR, Gloria. “Acerca de la (in)traducibilidad de la fraseología”, en Gloria Corpas Pastor (ed.): *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Iberoamericana, 483-522, 2000.
- COSERIU, Eugenio. *Principios de semántica estructural*. Madrid: Gredos, 1977.
- COSERIU, Eugenio. *Gramática, semántica, universales*. Madrid: Gredos, 1978.
- COSERIU, Eugenio. *Sincronía, diacronía e historia. El problema del cambio lingüístico*. Madrid: Gredos, 1978.
- COSERIU, Eugenio. “La socio- y la etnolingüística: sus fundamentos y sus tareas”. *Anuario de Letras*, XIX: 5-29, 1981.
- COSERIU, Eugenio. *Teoría del lenguaje y lingüística general*. Madrid: Gredos, 1982.
- COSERIU, Eugenio. *El hombre y su lenguaje*. Madrid: Gredos, 1991.
- CRUZ PIÑOL, Mar. *Lingüística de corpus y enseñanza de español como L2*. Madrid: Arco Libros, 2012.
- CUENCA, Maria Josep. “La reformulació i la exemplificació en el discurs científic: contrast gramatical y retòric”. Ponencia pronunciada en la II Jornada sobre Coneixement, Llenguatge i Discurs especialitat. Valencia: 2001.
- CUENCA, Maria Josep y HILFERTY, Joseph. *Introducción a la lingüística cognitiva*. Barcelona: Ariel, 1999.
- DAVIDSON, Richard y BEGLEY, Sharon. *El perfil emocional de tu cerebro*. Barcelona: Destino, 2012.
- DAWKINGS, Richard. *The selfish Gene*. Oxford: Oxford University Press, 1976.
- DEL CAMPO MARTÍNEZ, M<sup>a</sup> Amparo. “Los refranes en la enseñanza de español como lengua extranjera”. Online: 2001 [http://dspace.uah.es/dspace/bitstream/handle/10017/7448/refranes\\_campo\\_REALE\\_2001.pdf?sequence=1](http://dspace.uah.es/dspace/bitstream/handle/10017/7448/refranes_campo_REALE_2001.pdf?sequence=1)
- DE MAURO, Tullio. *Lessico di frequenza dell'Italiano Parlato (LIP)*. Con la colaboración de Federico Mancini, Massimo Vedovelli y Miriam Voghera, 1993. Disponible en BADIP <http://badip.uni-graz.at/it/descrizione>
- DE MAURO, Tullio. *La fabbrica delle parole. Il lessico e i problemi di lessicologia*. Torino: UTET Libreria, 2005.
- DE MAURO, Tullio. *Guida all'uso delle parole*. Roma: Editori Riuniti, 2007.

- DE MIGUEL, Amando. *Los españoles, sociología de la vida cotidiana*. Madrid: Temas de hoy, 1990.
- DE MIGUEL, Amando. *Los españoles y la religión*. Madrid: Temas de hoy, 2006.
- DE SAUSSURE, Ferdinand. *Curso de lingüística general*. Madrid: Alianza, 1983.
- DEVOTO, Giacomo y OLI, Gian Carlo. *Il Devoto-Oli Vocabolario della lingua italiana*, a cura di Luca Serianni e Maurizio Trifone, Firenze: Le Monnier, 2014.
- DI GESÙ, Floriana. *Linguistica contrastivo-percettiva di lingue tipologicamente affini: italiano e spagnolo*. Palermo: Palermo University Press, 2016.
- DORNYEI, Zoltan. *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press, 2007.
- DOUGHTY, Catherine y WILLIAMS, Jessica. *Focus on form in classroom second language acquisition*. New York: Cambridge University Press, 1998.
- DURANTI, Alessandro. *Antropología del lenguaje*. Roma: Meltemi, 2000.
- DURANTI, Alessandro. *Culture e discorso. Un lessico per le scienze umane*, Roma: Meltemi, 2001.
- EAGLETON, Terry. *Ideology. An introduction*. London: Verso, 1991.
- FERNÁNDEZ GUIZZETTI, Germán. “La etnolingüística: del mundo del idioma al mundo de la cultura”, en *Revista de antropología*, Sao Paulo, 5, 1: 75-93, 1957.
- FERNÁNDEZ GUIZZETTI, German. *Guillermo de Humboldt, padre de la etnolingüística. Esquemas para una historia de la etnolingüística*. Buenos Aires: “Cuadernos del Instituto Nacional de Investigaciones Folklóricas”, 1: 229-245, 1960.
- FOLEY, William. *Anthropological linguistics: an introduction*. Malden: Blackwell, 1997.
- FORMENT FERNÁNDEZ, Mar. “La fraseología metafórica en la didáctica de segundas lenguas.” *Rilce* 14 (2): 225–41, 1998. <http://search.proquest.com/openview/ae-b1063806274e8fdcffd6cec2d06e0/1?pq-origsite=gscholar&cbl=1816843>
- GALLESE, Vittorio y Alvin GOLDMAN. “Mirror neurons and the simulation theory of mind-reading”. *Trends in Cognitive Sciences*, 2: 493-501, 1998.
- GARCIA-BORRÓN, Juan Pablo. *Semántica de la palabra*. Barcelona: Universitat de Barcelona, 2013.
- GARCÍA GARCÍA, Emilio. “Neuropsicología y educación. De las neuronas espejo a la teoría de la mente.” *Revista de Psicología y Educación*, 1 (3): 69–90, 2008. <http://eprints.ucm.es/9972/>
- GARCÍA GARCÍA, Emilio; GONZÁLEZ MARQUÉS, Javier y MAESTÚ UNTURBE, Fernando. “Neuronas espejo y teoría de la mente en la explicación de la empatía.” *Ansiedad y Estrés*, 17 (2–3): 265–79, 2011. <http://eprints.sim.ucm.es/16341/>
- GARGALLO, José Enrique; TORRES TORRES, Antonio y FRANCO ANCHELERGUES, Vicente: “Líneas metodológicas de BADARE (Base de datos sobre refranes del calendario y meteorológicos de Romania)”, en *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística*. Publicaciones de la Universidad de Navarra, 207-215, 2008 y en <http://www.unav.es/linguis/simposiosel/actas/act/18>.

- pdf
- GATTO, Maristella. *The web as corpus: theory and practice*. London/New York: Bloomsbury Academic, 2014.
- GEAKE, John. *The brain at school: educational neuroscience in the classroom*. New York: McGraw-Hill Education, 2009.
- GELUSO, Joe. "Phraseology and frequency of occurrence on the web: native speakers' perceptions of Google-informed second language writing." *Computer Assisted Language Learning*, 26, 2: 144–157, 2013.
- GLEESON, Patrick y WAKEFIELD, Nancy. *Language and culture*. Columbus: Merrill and Co, 1968.
- GOLEMAN, Daniel. *El cerebro y la inteligencia emocional: nuevos descubrimientos*. Barcelona: Ediciones B, 2012.
- GÓMEZ DEL ESTAL, Mario y ZANÓN, Javier. "Tareas formales para la enseñanza de la gramática en la clase de español". En ZANÓN, Javier. (ed.) *La enseñanza del español mediante tareas*. Madrid: Edinumm: 73-99, 1999.
- GÓMEZ DÍAZ, Raquel. *La lematización en español: una aplicación para la recuperación de información*. Gijón: Ediciones Trea, 2005.
- GONZÁLEZ-LLORET, Marta. "Investigación-Acción (II): La Investigación Cuantitativa". En *DidactiRed*. Centro Virtual Cervantes, 2014. [https://cvc.cervantes.es/aula/didactired/anteriores/diciembre\\_12/17122012.htm](https://cvc.cervantes.es/aula/didactired/anteriores/diciembre_12/17122012.htm)
- GOODELL, James. *Approaching 100% by 2014: Using data-driven technology, scientifically proven practices, and cultural change to meet the no child left behind challenge*. Estados Unidos: Sound Image Inc., 2007.
- GREENFIELD, Patricia Marks y BRUNER, Jerome. "Culture and cognitive growth". *International Journal of Psychology*, 1: 89-107, 1966.
- HARPER, Douglas. *Online Etymology Dictionary*. Online, 2000. <http://www.etymonline.com/>
- HERNÁNDEZ SACRISTAN, Carlos. *Culturas y acción comunicativa. Introducción a la pragmática intercultural*. Barcelona: Octaedro, 1999.
- HINKEL, Eli. *Culture in second language teaching and learning*. Cambridge: Cambridge University Press, 1999.
- HOIJER, Harry. *The relation of language to culture*. En: KROEBER, Alfred (ed.). *Anthropology today*. Chicago: University of Chicago Press, 554-573, 1953.
- HOIJER, Harry. *Language in culture*. Chicago: University of Chicago Press, 1954.
- HOIJER, Harry. *Linguistic and cultural change*. En: HYMES, Dell (ed.). *Language in culture and society: a reader in linguistics and anthropology*. New York: Harper and Row, 455-462, 1964.
- HOHULIN, Lou. "The Absence of Lexical Equivalence and cases of its Asymmetry". *Lexicographica*, 2: 43-52, 1986.
- HORTIGUERA, Hugo y DIAZ, Adriana. "Extendiendo los muros de la clase de ELE:

- Una aproximación neurocognitiva a través de la tecnología.” *MarcoELE: Revista de Didáctica*, 19, 2014. <http://marcoele.com/una-aproximacion-neurocognitiva-a-traves-de-la-tecnologia/>
- HUDSON, Richard. *La sociolingüística*. Barcelona: Anagrama, 1981.
- HUMBOLDT, Wilhelm. VON. *On language. The diversity of human language-structure and its influence on the mental development of making*. Cambridge: Cambridge University Press. Ed. de 1999.
- HYMES, Dell. *Cultural focus and semantic field*. Introduction. En HYMES, Dell (ed.) *Language in culture and society*. New York: Harper, 167-170, 1964.
- HYMES, Dell. “Two Types of Linguistic Relativity”. En: BRIGHT, William. (ed.). *Sociolinguistics*. La Haya: Mouton, 114-158, 1966.
- HYMES, Dell. “On Communicative Competence” En: PRIDE, John y HOLMES, Janet (eds). *Sociolinguistics Selected Readings*. Harmondsworth: Penguin, 269-293, 1972.
- HYMES, Dell. *Models of the interaction of language and social life*. En: GUMPERZ, John y HYMES, Dell (dir.). *Directions in sociolinguistics: the ethnography of communication*. New York: Holt, Rinehart and Winston, Inc., 1972.
- IACOBONI, Marco y DAPRETTO, Mirella. “The mirror neuron system and the consequences of its dysfunction”. *Nature*, 7: 942-951, 2006.
- IACOBONI, Marco. *Las neuronas espejo*. Buenos Aires: Katz Editores, 2009.
- ILLAMOLA, Cristina. “La convergència lingüística: possible retardador del procés de gramaticalització?”. *Treballs de Sociolingüística Catalana*, 23: 81-101, 2013.
- ISIDORO, Santo, Arzobispo de Sevilla. *Etimologías*. Madrid: Biblioteca de autores cristianos, ed. de 2004.
- IZQUIERDO GIL, María del Carmen. “Los corpórea electrónicos en la enseñanza del vocabulario español: la utilidad pedagógica de las concordancias.” En: *Tecnologías de la información y de las comunicaciones en la enseñanza de la E/LE - Actas del XII Congreso Internacional de ASELE*, edited by Ana María Gimeno Sanz, 201–212. Valencia: Universidad Politécnica de Valencia – Asele, 2002. [http://cvc.cervantes.es/ensenanza/biblioteca\\_ele/asele/pdf/12/12\\_0191.pdf](http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/12/12_0191.pdf)
- JAEGER, Werner. (1923). *Aristóteles. Bases para la historia de su desarrollo intelectual*. Versión española de José Gaos, F.C.E., México, 1946.
- KÄDING, Friedrich Wilhelm (1897). *Häufigkeitswörterbuch der deutschen Sprache*. Steglitz, publicado de manera privada. La referencia a su obra aparece en el artículo de Guillermo Rojo titulado “Lingüística de corpus y Lingüística del español”, ponencia plenaria en el XV congreso de la ALFAL, Montevideo, 18-21 de agosto de 2008.
- KRAMSCH, Claire. “The cultural component of language teaching.” *Language, Culture and Curriculum*, 8 (2): 83-92, 1995. <http://www.tandfonline.com/doi/abs/10.1080/07908319509525192>
- KROEBER, Alfred y KLUCKHOHN, Clyde. *Culture: a critical review of concepts and definitions*. Harvard University, Peabody Museum of American Archeology and Eth-

- nology, papers 47, 1952.
- LANDAR, Herbert. *Language and culture*. Oxford: Oxford University Press, 1966.
- LASO MARTÍN, Natàlia Judith. *A corpus-based study of the phraseological behaviour of abstract nouns in medical english a needs analysis of a spanish medical community*. Tesis doctoral dirigida por la Dra. Isabel Verdaguer. Barcelona: Universitat de Barcelona, 2010. <http://www.tdx.cat/handle/10803/1671>
- LASO, Natalia Judith y COMELLES, Elisabet. "Usos metodológicos de corpus textuales en el aula de lexicología y morfología." En: *experiencias de innovación docente en la enseñanza universitaria de las humanidades*, editado por TRENCHS PARERA, Mireia y CRUZ PIÑOL, Mar. Barcelona: Octaedro, 55-72, 2012. <http://www.octaedro.com/OCTart.asp?libro=10183&id=es&txt=Experiencias+de+innovaci%F3n+docente+en+la+ense%F1anza+universitaria+de+las+humanidades>
- LENNEBERG, Eric Heinz. "Cognition in ethnolinguistics", *Language*, 29: 463-171, 1953.
- LEVINE, Donald. (ed). *Simmel: on individuality and social forms*. Chicago: University Press, 1971.
- LÉVI STRAUSS, Claude. *Antropologia strutturale*. Milano: Il Saggiatore, ed. de 1966.
- LIBERMAN, Alvin y MATTINGLY, Ignatius. "The motor theory of speech perception revised", *Cognition*, 21: 1-36, 1985.
- LINDSTROM, Martin. *Buyology, verdades y mentiras de por qué compramos*, Barcelona: Ediciones Gestión, 2000.
- LIPTON, Bruce. *La biología de la creencia*. Barcelona: Editorial Palmyra, 2007.
- LÓPEZ GARCÍA, Ángel. "La visión del mundo del otro: lenguas primitivas y categorías conceptuales". Ponencia pronunciada en el Seminario "Lenguaje y visión del mundo" celebrado en la Universidad Menéndez Pelayo, Valencia, septiembre de 1993.
- LUCY, John. *Language diversity and thought. A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press, 1992.
- LUQUE DURÁN, Juan de Dios. *Aspectos universales y particulares del léxico de las lenguas del mundo*. Granada: Método, 2001.
- LYONS, John. *Introducción al lenguaje y a la lingüística*. Barcelona: Teide, 1984.
- MALINOWSKI, Branislaw. (1923). El problema del significado en las lenguas primitivas. En: OGDEN, Charles y RICHARDS, Ivor. *El significado del significado*. Barcelona: Paidós, 310-352, 1984.
- MALINOWSKI, Branislaw. *The language of magic and gardening*. En: *Coral gardens and their magic*. Londres: Allen and Unwin, vol. 2, 4-65, 1935.
- MANCHESTER, Martin. *The philosophical foundations of Humboldt's linguistic doctrines*. Amsterdam: John Benjamins, 1985.
- MARÍAS, Julián. *El oficio del pensamiento*. Madrid: Espasa, 1968.
- MARTÍN CAMACHO, José Carlos. "Hacia una caracterización de una disciplina lingüística casi olvidada: La etnolingüística." *ELUA: Estudios de Lingüística. Universidad*

- de Alicante*, 30: 181-212, 2016. <http://revistaelua.ua.es/article/view/2016-n30-hacia-una-caracterizacion-de-una-disciplina-linguistica-casi-olvidada-la-etnolinguitica>
- MARTÍN GARCÍA, Josefa. *Los prefijos intensivos del español: caracterización morfo-semántica*. Universidad Autónoma de Madrid, 1998.
- MARTIN, Laura. "Eskimo words for snow: a case study in the genesis and decay of an anthropology example". *American Anthropologist*, 88: 418-423, 1986.
- MARX, Karl (1867). *El capital*. Barcelona: editorial Editors, ed. de 2008.
- MASON, Hilary y PATIL, Dhanurjay. *Data Driven. Creating a data culture*. O'Reilly Media, 2015.
- MASSON, Steve. "Neuroeducation: understanding the brain to improve teaching." *Neuroeducation Scientific Journal*, 1, (1), 2012. <http://www.associationneuroeducation.org/revue?category=Vol.+1+no.+1>
- MATHIOT, Madeleine. *Ethnolinguistics: Boas, Sapir, and Whorf Revisited*. The Hague: Mouton, 1979.
- McENERY, Tony y WILSON, Andrew. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996.
- MELLADO BLANCO, Carmen. "Parámetros específicos de equivalencia en las unidades fraseológicas (con ejemplos del español y el alemán)". *Revista de filología*, 33: 153-174, Universidad de Santiago de Compostela, 2015.
- MÉNDEZ-SANTOS, María del Carmen. "Gramática afectiva en contextos de instrucción formal de español como lengua extranjera (ELE)." *Hesperia. Anuario de Filología Hispánica. Nuevas perspectivas en la enseñanza del español como lengua extranjera XIX* (2): 51-84, 2016.
- MORA, Francisco. *Neurocultura*. Madrid: Alianza Editorial, 2007.
- MORA, Francisco. *Neuroeducación*. Madrid: Alianza Editorial, 2013.
- MORA, Francisco. "¿Qué es neuroeducación?". Seminario *Educación, comunicación y gramática II*, II Foro Universidad de Barcelona- SGEL, 10 de marzo de 2017.
- MORANT MARCO, Ricard y PEÑARROYA, Miquel. *Llenguatge i cultura: per a una ecologia lingüística*. Valencia: Universidad de Valencia, 1995.
- MORANT MARCO, Ricard. *El lenguaje de los estudiantes: un paseo por las aulas*. En: RODRIGUEZ GONZÁLEZ, Félix (ed.) *El lenguaje de los jóvenes*. Barcelona: Ariel, 2002.
- MORENO CABRERA, Juan Carlos. *Fundamentos de sintaxis general*. Madrid: Síntesis, 1987.
- MORENO CABRERA, Juan Carlos. *La dignidad y la igualdad de las lenguas. Crítica de la discriminación lingüística*. Madrid: Alianza, 2000.
- MOUNIN, Georges. *Lingüística y filosofía*. Madrid: Gredos, 1979.
- MURA, Angela. *La fraseología del desacuerdo: los esquemas fraseológicos en español y en italiano*. Tesis doctoral dirigida por la Dr. Leonor Ruiz Gurillo y la Dr. María Teresa Zurdo Ruiz-Ayucar. Madrid: Universidad Complutense de Madrid, 2012.

- NAVARRO, Carmen. “Fraseología contrastiva del español y el italiano (Análisis de un corpus bilingüe).” *Tonos Digital: Revista Electrónica de Estudios Filológicos*, 13, 2007. <https://digitum.um.es/xmlui/handle/10201/43284>
- NOMDEDEU RULL, Antoni. “Diccionario de fútbol”. *Anexos de Revista de Lexicografía*. A Coruña: Universidade da Coruña, 2009.
- O’CONNOR, William. “What can neuroscience teach us about teaching?”. *International Conference on Engaging Pedagogy*, 2010. <https://ulir.ul.ie/handle/10344/4456>
- OLMSTED, David. “Ethnolinguistics so far”. *SiL, Occasional Papers* 2: 1-16, 1950.
- ORTONY, Andrew; CLORE, Gerald y COLLINS, Allan. *La estructura cognitiva de las emociones*. Madrid: Siglo XXI, 1996.
- OSBORN, Alex Faickney. *Imaginación aplicada*. Madrid: Ediciones Velflex, 1957.
- PALMER, Gary. *Lingüística antropológica*. Madrid: Alianza, 2000.
- PIETRZAK, Justyna Beata. *Binomios fraseológicos en el italiano contemporáneo*. Tesis doctoral dirigida por la Dra. Julia Sevilla Muñoz y la Dra. Maria Antonella Sardelli. Universidad Complutense de Madrid, 2015 <http://eprints.ucm.es/38285/>
- PISANI, Vittore. *L’etimologia*. Brescia: Praideia Editrice, 1967.
- PITKOWSKI, Elena Fabiana y VÁSQUEZ GAMARRA, Javier. “El uso de los corpus lingüísticos como herramienta pedagógica para la enseñanza y aprendizaje de ELE.” *Tinkuy: Boletín de investigación y debate*, 11: 31–51, 2009. <https://dialnet.unirioja.es/servlet/articulo?codigo=3303856&info=resumen&idioma=SPA>
- POIRIER, Jean. *Ethnologie générale*. París: Gallimard, 1968.
- PORZIG, Walter. *El mundo maravilloso del lenguaje*. Madrid: Gredos, 1964.
- PRADERA, Máximo. *¿De qué me suena eso? Paseo informal por la música clásica*. Madrid: Santillana, 2005.
- PUSTEJOVSKY, James y BOGURAEV, Branimir. “Introduction: Lexical Semantics in Context”. *Journal of Semantics*, 12: 1-14, 1995.
- PUYAL ORTIGA, Joaquim. *Terminología futbolística*. Tesis de licenciatura dirigida por el Dr. Francisco Marsà Gómez. Barcelona: Universitat de Barcelona, 1972.
- RAFEL i FONTANALES, Joaquim y SOLER i BOU (2001). *El procesamiento de corpus*. En: MARTÍ, Maria Antonia (Coord.), *Tecnologías del lenguaje*, Barcelona: Editorial UOC, 2003.
- RAMACHANDRAN, Vilayanur. “Mirror neurons and imitation learning as the driving force behind the Great Leap Forward in human evolution”. *Edge*, 69, 29 de junio de 2000.
- RAMACHANDRAN, Vilayanur. *Lo que el cerebro nos dice: los misterios de la mente humana al descubierto*, Barcelona: Paidós, 2012.
- REAL ACADEMIA ESPAÑOLA. *Diccionario de la lengua española. Edición del tricentenario*. Espasa Libros, 2014. <http://www.rae.es/obras-academicas/diccionarios/diccionario-de-la-lengua-espanola>
- RESTREPO, Félix. *El alma de las palabras*. Diseño de semántica general. Barcelona: Im-

- prenta editorial Barcelonesa, 1917.
- REUS BOYD-SWAN, Francisco. "El léxico taurino en la vida cotidiana". *Revista de la Sociedad Española de Estudios Literarios de Cultura Popular*, 4: 239-263, 2004.
- RIZZOLATTI, Giacomo y CRAIGHERO Laila. "The mirror neuron system". *Annual Rev. Neuroscience*, 27: 169-192, 2004.
- RIZZOLATTI, Giacomo; FOGASSI, Leonardo y GALLESE, Vittorio: "Mirrors in the mind". *Scientific American*, 11, vol. 295: 54-61, 2006.
- RIZZOLATTI, Giacomo y SINIGAGLIA, Corrado. *Las neuronas espejo: los mecanismos de la empatía emocional*. Barcelona: Paidós, 2006.
- ROHLFS, Gerhard. *Estudios sobre el léxico románico*. Madrid: Gredos, 1979.
- ROSSINI FAVRETTI, Rema. *Language, culture, institutions*. Bologna: Pitagora, 1993.
- ROSSINI FAVRETTI, Rema. *La linguistica applicata. Aspetti, percorsi, problema*. Bologna: Pàtron, 1998.
- ROSSINI FAVRETTI, Rema. *Un'introduzione alla linguistica applicata*. Bologna: Pàtron, 2002.
- RUFFINATTO, Aldo. *Sobre textos y mundos. Ensayos de filología y semiótica hispánicas*. Murcia: Editum, 1990.
- RUFFINATTO, Aldo. *Las dos caras del Lazarillo: texto y mensaje*. Madrid: Castalia, 2000.
- RUIZ GARCÍA, Daniel. *Todo está bien*. Barcelona: Tusquets Editores, 2015.
- SALAZAR, Danica y VERDAGUER, Isabel. "Polysemous verbs and modality in native and non-native argumentative writing: A corpus-based study". *International Journal of English Studies (IJES)*, vol. 9: 209-219, 2009.
- SAN VICENTE, Félix. *La lengua de los nuevos españoles*. Zaragoza: Libros Pórticos, 2001.
- SAN VICENTE, Félix. *Il progetto e-learning AlfaCert del CLIRO dell'Università di Bologna*. En: *Atti del 3° Convegno Nazionale A.I.C.L.U. (6° seminario A.I.C.L.U. Messina), I Centri Linguistici Italiani: Approcci, progetti e strumenti per l'apprendimento e la valutazione*. Trieste: Centro Stampa del Dipartimento di Scienze del Linguaggio, 619-629, 2005.
- SAN VICENTE, Félix. "Tradición e innovación en la lexicografía bilingüe italoespañol: a propósito de la relación entre diccionario y gramática". *Perfiles para la historia y crítica de la lexicografía bilingüe del español*, Monza: Polimetrica International Scientific Publisher, 89-114, 2007.
- SAPIR, Edward. *The selected writings of Edward Sapir*. Berkely: University of California, 1949.
- SAPIR, Edward. *Cultura, linguaggio e personalità. Linguistica e antropología*, Torino: Einaudi, ed. de 1972.
- SCHAFF, Adam y VELEN Victor. "Problems of ethnolinguistics." *Diogenes*, 12 (46): 125-50, 1964.
- SEGRE, Cesare. *Lingua, stile e società. Studi sulla storia della prosa italiana*. Milano: Feltrinelli, 1963.

- SEPÚLVEDA, Gastón. “Algunos problemas de la etnolingüística.” *Revista Documentos Lingüísticos y Literarios UACH*, 4: 47-67, 2016. <http://www.revistadll.cl/index.php/revistadll/article/view/36>
- SHA, Guoquan. “Using Google as a super corpus to drive written language learning: A comparison with the British National Corpus.” *Computer Assisted Language Learning* 23 (5). <http://www.tandfonline.com/doi/abs/10.1080/09588221.2010.514576>, 2010.
- SILVA-FUENZALIDA, Ismael. “Ethnolinguistics and the study of culture.” *American Anthropologist*, 51 (3): 446–56. <http://www.jstor.org/stable/664540>, 1949.
- SINCLAIR, John McHardy. “Corpus and text – Basic principles”. En WYNNE, Martin: *Developoing Linguistic Copora: A guide to good practice*. Oxford: Oxbow Books, 2005. <http://ota.ox.ac.uk/documents/creating/dlc/chapter1.htm>
- SINCLAIR, John McHardy. *How to use corpora in language teaching*. Colección Studies in Corpus Linguistics, 12, John Benjamins, 2004.
- SINCLAIR, John McHardy. *Trust the text: Language, corpus and discourse*. Londres: Routledge, 2004.
- SOMMERS, Mitchell y BARCROFT, Joe (2007). “An integrated account of the effects of acoustic variability in first language and second language: evidence from aptitude, fundametal frequency, and speaking rate variability”. *Applied Psycholinguistics*, 28 (2): 231-249. Cambridge: Cambridge University Press, 1 de marzo de 2007.
- SOUSA, David. *Neurociencia Educativa. Mente, cerebro y educación*. Madrid: Narcea Ediciones, 2014.
- SPENCE, Robert. *Information visualization: design for interaction*. 2nd ed. Harlow: Addison Wesley, 2007.
- SWEETSER, Eve. *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press, 1990.
- TERRACINI, Benvenuto. *Conflitti di lingue e di cultura*. Venezia: Neri Pozza, 1957.
- TERUEL SAEZ, Antonio. *Vocabulario de fútbol*. Gijón: Trea, 2007.
- TIMOFEEVA, Larissa. *Acerca de los aspectos traductológicos de la fraseología española*. Tesis doctoral dirigida por la Dra. Leonor Ruiz Gurillo. Alicante: Universidad de Alicante, 2008. <http://www.tesisenred.net/handle/10803/11756>
- TOBIAS, Philip. “L’evolution du cerveau humain”. *La Recherche*, XI: 282-292, 1980.
- TORRES TORRES, Antonio. *El español de América*. Barcelona: Universitat de Barcelona, 2001.
- TORRUELLA, Joan y LLISTERRI, Joaquim. “Diseño de corpus textuales y orales”. En BLECUA, José Manuel; CLAVERÍA, Gloria; SÁNCHEZ, Carlos y TORRUELLA, Joan (Eds.). *Filología e informática. Nuevas tecnologías en los estudios filológicos*, Barcelona: Seminario de Filología e Informática de la Universidad Autónoma de Barcelona y Ed. Milenio, 45-77, 1999.
- TRAMONTI. Francesca Romana. *Appunti di antropologia cognitiva*. Lungavilla (PV):

- Edizioni Altavista, 2010.
- ULLMANN, Stephen. *Semántica. Introducción a la ciencia del significado*. Aguilar: Madrid, 1976.
- VELÁZQUEZ-PUERTO, Karen. *Análisis fraseológico contrastivo español-alemán y su aplicación a ELE: una propuesta didáctica para la enseñanza-aprendizaje de las unidades fraseológicas desde una perspectiva cognitiva*. Tesis doctoral dirigida por la Dra. María Luisa Regueiro Rodríguez. Madrid: Universidad Complutense de Madrid, 2015.
- VENY i CLAR, Joan y PONS GRIERA, Lúdia. *Atles Lingüístic del Domini Català. 4. La família: cicle de la vida. 5. Món espiritual: l'Església. Festes religioses. Creences. 6. Jocs. 7. Temps cronològic. Meteorologia. 8. Topografia*. vol. III. Barcelona: Institut d'Estudis Catalans, 2006.
- VERDAGUER, Isabel y JUAN, María. "Generación de un diccionario especializado combinatorio bilingüe". *Anuari de Filologia*. Barcelona: XXI-XXII, 9, 69-78, 1998-2000.
- VERDAGUER, Isabel. "A corpus-based contrastive study of mental verbs". *Lengua, traducción, recepción. En honor de Julio-César Santoyo*. León: Universidad de León, 663-713, 2010.
- VILADOT i PRESAS, Ma. Àngels. "Percepció de la vitalitat i identitat social." *Treballs de Sociolingüística Catalana*, 10: 125-129, 1992. <http://www.raco.cat/index.php/TSC/article/view/224037>
- VILA i MORENO, Francesc Xavier; NOGUÉ, Marina y VILA, Ignasi. *Estudis d'implantació terminològica. Una aproximació en l'àmbit dels esports*. Barcelona: Eumo Editorial, 2007.
- VILA i MORENO, Francesc Xavier. "Algunes bases per a la recerca sociolingüística en sentit ampli". *Posar-hi la base. Usos i aprenentatges lingüístics en el domini català*, 8: 11-24, 2012.
- VIVAS HOLGADO, Jesús. *El fútbol: léxico, deporte y periodismo*. Cáceres: Universidad de Extremadura, 1999.
- VOSSLER, Karl. *Algunos caracteres de la cultura española*. Buenos Aires: Espasa-Calpe Argentina, 1941.
- VOSSLER, Karl. *Cultura y lengua de Francia*. Buenos Aires: Losada, ed. de 1955.
- WANDRUSZKA, Mario. *Nuestros idiomas: comparables e incomparables*. Madrid: Greedos, 1976.
- WARE, Colin. *Information visualization: perception for design*. 2nd ed. San Francisco: Morgan Kaufman, 2004.
- WHORF, Benjamin Lee. *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. Cambridge: The Massachusetts Institute of Technology Press, 1956.
- WIERZBICKA, Anna. *Cross-cultural pragmatics. The semantics of human Interaction*. Berlin: Mouton de Gruyter, 1991.
- WIERZBICKA, Anna. *Understanding cultures through their key words*. New York: Oxford

University Press, 1997.

ZENGIN, Bugra. "Benefit of Google search engine in learning and teaching collocations." *Eurasian Journal of Educational Research*, 34: 151, 2009. <http://connection.ebscohost.com/c/articles/36264742/benefit-google-search-engine-learning-teaching-collocations>



