
El problema de la traducción en el análisis cuantitativo de textos. Aplicación de Wordscores y Wordfish a las mociones de censura contra el lehendakari Ibarretxe¹

Rafael Leonisio y Oliver Strijbis

Wordscores y Wordfish son dos métodos de análisis cuantitativo de textos que están ganando popularidad muy rápidamente debido a que pueden ser utilizados para el análisis de posiciones ideológicas. Uno de los criterios principales para evaluar la utilidad de estos métodos es si son capaces de generar resultados comparables en diferentes idiomas. Los registros de los debates parlamentarios en el Parlamento vasco ofrecen un banco único de datos para probar la comparabilidad a través de dos idiomas muy diferentes. Esta nota de investigación valida tanto Wordscores como Wordfish, al comparar las posiciones políticas estimadas por ambos métodos, tanto en español como en euskera, en el debate de moción de censura contra el lehendakari Ibarretxe en el Parlamento vasco en 2000.

Palabras clave: Euskadi, análisis cuantitativo de textos, Wordscores, Wordfish.

INTRODUCCIÓN

El *Quantitative text analysis* o análisis cuantitativo de textos (QTA) es actualmente uno de los campos con mayor crecimiento en Ciencia Política. Eso ha sido posible en

1. Este trabajo ha sido posible gracias a la financiación del proyecto de investigación CSO2009-14381C03-01 del Gobierno español, del IT-323-07 del Gobierno vasco. Los autores quieren agradecer los útiles comentarios de Ken Benoit, que mejoraron la primera versión del texto.

parte gracias al desarrollo de dos métodos llamados *Wordscores* (Laver *et al.*, 2003, Lowe, 2008) y *Wordfish* (Slapin y Proksch, 2008). Estos métodos están adquiriendo popularidad debido a que pueden utilizarse para el análisis de posiciones ideológicas (Laver y Benoit, 2002; Laver *et al.*, 2006, Proksch y Slapin, 2006; Hug y Schultz, 2007; Debus, 2009; Klüver, 2009; Pappi y Seher, 2009), análisis que es esencial para muchas teorías en Ciencia Política, además de muy exigente metodológicamente hablando. El potencial de estos métodos y su posible falta de validación han causado un intenso debate sobre sus puntos fuertes y débiles², siendo la principal discusión si los métodos QTA aportan cálculos válidos para las posiciones ideológicas de los actores políticos (en particular Benoit y Laver, 2007; Budge y Pennings, 2007a, 2007b). Sin embargo, el actual debate sobre los pros y contras del QTA se centra mucho más en el rendimiento global que en los puntos fuertes y débiles de los elementos específicos del método y no abundan los modelos de investigación satisfactorios para probar empíricamente su rendimiento³. Uno de los criterios principales para determinar la utilidad de los métodos QTA en análisis comparativo es la posibilidad de generar resultados equiparables entre diferentes idiomas y, por lo tanto, también entre distintos países. Aunque la equiparación entre los distintos idiomas es crucial para la futura utilización de los métodos QTA, los investigadores no han prestado especial atención a los efectos de la traducción al utilizar análisis de contenido computerizado (Slapin y Proksch, 2009:12).

Esta nota de investigación pretende responder a la pregunta de si *Wordfish* y *Wordscores* son demasiado vulnerables a la traducción o si por el contrario pueden utilizarse para comparar distintas comunidades lingüísticas. Nosotros lo realizamos aplicando los dos métodos al análisis del discurso político del Parlamento del País Vasco. Al ser el euskera y el español dos idiomas muy diferentes, el análisis de los mismos discursos traducidos a ambos idiomas supone para estos métodos QTA un reto muy exigente. Por lo tanto, si se producen resultados similares en estos dos idiomas, se conseguirán conclusiones particularmente robustas⁴.

LA TRADUCCIÓN EN WORDSCORES Y WORDFISH

Wordscores y *Wordfish* consisten en dos algoritmos diferentes, ambos basados en la comparación de frecuencias de palabras en los textos. *Wordscores* compara las frecuencias

2. *Electoral Studies* dedicó un volumen completo al cálculo las estimaciones de las posiciones ideológicas de los partidos políticos (Marks, 2007).

3. Las pruebas empíricas realizadas hasta la fecha confían en comparaciones con otros métodos de cálculo (análisis de contenido manual, encuestas de expertos, miembros del Parlamento y votantes y análisis de las votaciones en el parlamento) más que en experimentos naturales.

4. Sin embargo, lo contrario no es cierto. La desviación en los resultados entre idiomas puede deberse a varios motivos, y no todos ponen en duda la conveniencia del uso de palabras como unidades de medida.

relativas de palabras utilizadas en textos de referencia (elegidos por el investigador), para saber qué probabilidad hay de que una palabra aparezca en un texto. Esta probabilidad aporta peso a los valores definidos a priori para una dimensión política en todo texto de referencia. El total de las palabras con peso encontradas en los “textos vírgenes”, es decir, los textos para los que el investigador quiere calcular los valores en una determinada dimensión política, conduce a la posición estimada de dichos textos vírgenes⁵.

Wordfish analiza frecuencias de palabras de documentos de texto mediante la distribución de Poisson y el uso de un algoritmo de maximización de probabilidad. De este modo, Wordfish genera cálculos de probabilidad máxima de modo iterativo para los parámetros de las palabras (pesos que captan la importancia de una palabra para diferenciar las distintas posturas) y los parámetros de los documentos (valores que captan las posturas de los textos). Al incluir efectos fijos en el modelo, Wordfish contabiliza las diferentes frecuencias con las que aparecen las palabras y las distintas longitudes de los documentos⁶.

La evidencia empírica de que Wordscores y Wordfish funcionan independientemente de los idiomas elegidos se basa principalmente en probar que producen resultados válidos con textos en diferentes idiomas⁷. Sin embargo, hasta donde sabemos, solo se ha analizado una vez el mismo texto traducido a dos idiomas con ambos: Wordscores y Wordfish. Hug y Schultz (2007: 325) informan en una nota a pie de página de que comprobaron la precisión de utilizar textos traducidos, comparando los resultados de Wordscores de un texto que tenían en francés y alemán. Descubrieron que “las posturas calculadas diferían, pero se encontraban dentro de los límites de los respectivos intervalos de confianza”. En una contribución reciente, Slapin y Proksch (2009) compararon los resultados de Wordfish para los mismos discursos parlamentarios en inglés, francés y alemán⁸. Encontraron correlaciones altas (entre 0.86 y 0.93) para los resultados de las diversas lenguas, pero se enfrentaron a una interesante paradoja: los cálculos de Wordfish de los textos traducidos a dos idiomas eran más similares entre las dos versiones de idiomas que si los resultados de Wordfish del texto traducido se comparasen con los resultados de un análisis de los textos en su versión original. Esto significa, por ejemplo, que un discurso traducido del español a francés y alemán genera más valores similares al comparar la versión francesa y alemana del análisis de Wordfish que si los resultados de Wordfish para los textos originariamente dados

5. Para la descripción del algoritmo, véase Laver *et al.* (2003) y Lowe (2008). Para el manual y el código de implementación Stata, véase http://www.tcd.ie/Political_Science/wordscores/.

6. Para la descripción del algoritmo, véase Slapin y Proksch (2008). Para el manual y el código fuente R, véase www.wordfish.org.

7. Benoit y Laver (2003) utilizan Wordscores para textos en alemán e inglés, Klemmensen *et al.* (2007) para textos en danés y Slapin y Proksch (2008) utilizan Wordfish para textos en alemán, todos con resultados válidos.

8. Todos los discursos, excepto los de los parlamentarios de habla germana, francesa o inglesa, se traducen previamente (véase Slapin y Proksch, 2009: 13).

en español se comparan con sus valores en una versión traducida. Slapin y Proksch (2009: 13) señalan que: “los discursos de hablantes nativos alemanes contienen un conjunto de palabras que no aparece en la traducción alemana de los discursos de las partes que no hablan alemán. En otras palabras, los traductores no utilizan las mismas palabras que los hablantes nativos al traducir al alemán.” Aunque este efecto parece ser más débil para los discursos pronunciados originariamente en francés e inglés, es cierto que plantea la cuestión de si la traducción a dos idiomas crea más similitud entre los textos traducidos. Esto significaría que los resultados similares de los textos traducidos a dos idiomas son hasta cierto punto algo artificial debido a cuestiones de traducción. Esta artificialidad podría responder a la utilización de un lenguaje más simple y estandarizado en los textos traducidos que en los originales. Desde este punto de vista, una comparación de resultados de Wordfish de los textos traducidos en una sola dirección podría ser una prueba más exigente que una comparación de resultados del análisis de textos traducidos de un tercer idioma.

Hay varios motivos por los que la traducción puede resultar problemática para los métodos QTA. El más evidente es que el significado de las palabras varía dependiendo del idioma y del contexto. A nivel de palabras individuales: sinónimos y homónimos, palabras vacías, distintos componentes léxicos, la conjugación y la declinación podrían afectar a los resultados del análisis QTA⁹. Su diferente uso en cada uno de los idiomas no cambia el contenido de las palabras si estas se toman como unidades de datos individuales, pero afecta a la matriz de frecuencia de palabras, que es la base de Wordscores y Wordfish.

Las diferencias en la matriz de datos de los textos traducidos provocan dos efectos en el cálculo de Wordfish y Wordscores. En primer lugar, afectan a la “unicidad” de las palabras en el sentido de ser específicas para un texto en concreto. En segundo lugar, afectan a la frecuencia con la que puntúan las palabras. En Wordscores, la menor unicidad de las palabras y la menor frecuencia con la que dichas palabras aparecen da lugar a que las “puntuaciones brutas” (*raw scores*) de los textos vírgenes se agrupen en posiciones muy cercanas. Por lo tanto, Laver *et al.* (2003) proponen la conversión de puntuaciones brutas a “puntuaciones transformadas” (*transformed scores*) (véase también Martin y Vanberg, 2007; Lowe, 2008). Mientras el procedimiento para establecer una nueva escala corrige las diferencias de la unicidad de palabras y la frecuencia de puntuación, las desviaciones debidas a la traducción, sin embargo, no pueden excluirse por las puntuaciones transformadas. En Wordfish, las frecuencias con las que puntúan las palabras se mantienen constantes, mientras que la unicidad de la palabra puede afectar al parámetro de la misma.

9. Las palabras vacías se utilizan frecuentemente y no aportan significado sustancial, generalmente son preposiciones, artículos y pronombres. Los componentes léxicos (*tokens*) son símbolos que dividen el texto en unidades.

COMPARACIÓN DE DISCURSOS PARLAMENTARIOS EN EUSKERA Y ESPAÑOL

Tal y como se presenta más arriba, exponemos Wordfish y Wordscores a un marco cuasi-experimental. El análisis de los mismos textos en dos idiomas nos permite mantener todo constante, excepto del uso de distintas de palabras entre idiomas y de la calidad de la traducción. Como el servicio de traducción oficial del Parlamento vasco realiza la traducción, cabe esperar que esta cumpla los estándares más altos que pueden esperarse para la traducción de otros textos políticos en otros países. El análisis proporciona una prueba muy exigente porque el español y el euskera son dos idiomas muy diferentes. Mientras que no es necesario exponer las raíces latinas del español, debe mencionarse que el euskera está clasificado como idioma aislado, el último idioma preindoeuropeo de Europa (Trask, 1996). A diferencia del español, el euskera cuenta con una gran variedad de desinencias debido a su tipología aglutinante (e.g. Alegria *et al.*, 1996)¹⁰.

Los discursos que analizamos pertenecen a un debate que precede a una moción de censura (dos mociones en realidad) en Parlamento vasco el 5 de octubre de 2000. En el debate a estudiar, participaron once oradores: los dos líderes de los principales partidos de la oposición, Iturgaiz del Partido Popular (PP) y Redondo del Partido Socialista de Euskadi-Euskadiko Ezkerra (PSE-EE), que presentaban su candidatura a Lehendakari; el entonces lehendakari, Ibarretxe, del Partido Nacionalista Vasco (PNV); dos miembros más de la oposición, uno del PP y otro del PSE, que presentaban a los candidatos de sus partidos; y una persona más de cada partido presente en el Parlamento vasco: EA (Eusko Alkartasuna), IU (Izquierda Unida), UA (Unidad Alavesa), PP, PSE-EE y PNV¹¹. Algunos representantes hablaron más de una vez, pero dado que algunos discursos son muy breves, hemos decidido considerar como un único discurso todos los realizados por la misma persona¹². En la votación final, los miembros del PP, PSE y UA votaron a favor de las dos mociones de censura, mientras que los del PNV, EA y IU votaron en contra. El debate se llevó a cabo tanto en español como en euskera y cada orador habló en un idioma o en otro, a veces en ambos. Sin embargo, se hace una transcripción completa de la sesión en los dos idiomas, lo que nos permite analizar la misma sesión parlamentaria en dos idiomas diferentes de manera independiente.

Para los cálculos de Wordscores se han asignado valores a priori a los dos documentos más “extremos”. En este caso esperamos que los discursos del presidente regional Ibarretxe (PNV) y los de los líderes de la oposición, Redondo (PSE) e Iturgaiz (PP), ocupen posturas enfrentadas¹³. Estos textos de referencia deberían, tal y como Laver *et al.* (2003) recomiendan, contener el mismo léxico que los textos vírgenes, una gran cantidad de

10. Nuestra prueba es especialmente difícil porque no utilizamos ni lematización ni etiquetado, puesto que relajaría la suposición de que todas las palabras tienen la misma información (Lowe, 2008).

11. Los nacionalistas de la izquierda *abertzale*, *Euskal Herritarrok* (EH), no participaron en el debate.

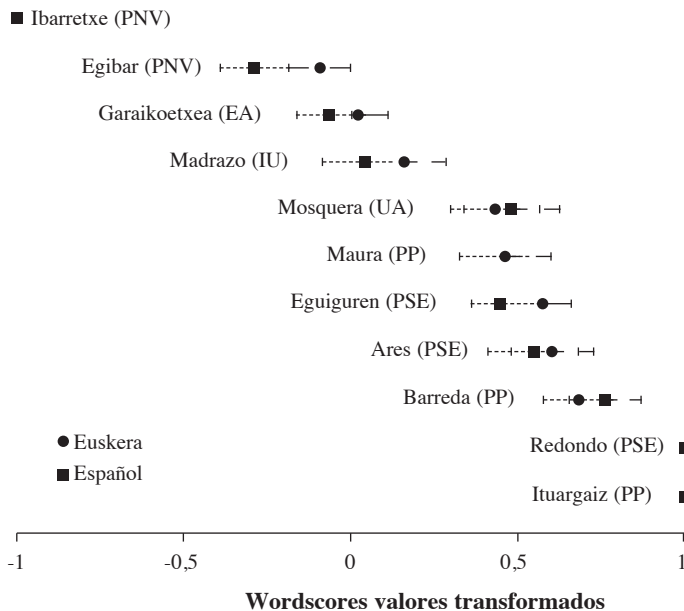
12. Los discursos se descargaron de la página web del Parlamento vasco: <http://parlamento.euskadi.net/>.

13. La misma estrategia seguida por Benoit y Laver (2003).

palabras y estar ubicados en los extremos de la escala establecida. Todas estas demandas se encuentran al comparar los discursos de la moción de censura del Parlamento Vasco¹⁴.

La figura 1 muestra los resultados de los cálculos de Wordscores. Indican las puntuaciones transformadas en euskera y en español. Tanto los valores del lehendakari Ibarretxe, como los de los líderes de la oposición, Redondo (PSE) e Iturgaiz (PP), son fijos, por lo tanto, no hay desviaciones entre las dos versiones de idiomas. En lo que respecta a los textos vírgenes los resultados de ambas versiones son muy similares: los representantes de los partidos gobernantes (PNV y EA) y los de los partidos de la oposición (UA, PSE y PP) se agrupan, mientras que la puntuación del representante de Izquierda Unida (IU) está situada entre ellos. Concretamente, la postura de Madrazo (IU) se aproxima ligeramente más a la del gobierno, lo cual refleja la cercanía de IU al gabinete de Ibarretxe en aquel momento. A excepción de los discursos de Mosquera (UA), Maura (PP) y Eguiguren (PSE), las dos versiones de Wordscores (euskera y español) colocan los discursos en el mismo orden. Esto acarrea una correlación muy alta de los textos calculados entre las dos versiones de idiomas (R=0.98), datos que reflejan claramente que la traducción apenas influye en los cálculos de Wordscores.

FIGURA 1.
RESULTADOS DE WORDSCORES EN DOS IDIOMAS (VALORES TRANSFORMADOS)

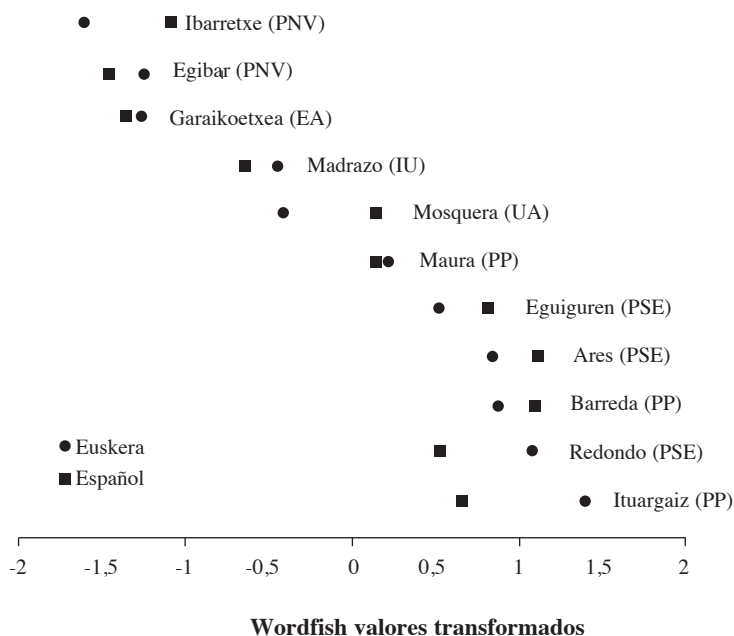


Notas: las líneas muestran intervalos de confianza al 95%.

14. Como el euskera es una lengua aglutinante, se necesitan menos palabras para producir el mismo contenido que en español. Del mismo modo, como el euskera utiliza muchos sufijos, hay más “palabras únicas” proporcionalmente. Los textos aquí analizados contienen entre 21.014 y 3.759 palabras en la versión en español y entre 14.847 y 2.435 palabras en la versión en euskera. El número de palabras únicas oscila entre 3.093 y 768 (en la versión española) y entre 4.084 y 861 (en la versión euskera).

La figura 2 muestra los resultados de los cálculos de Wordfish. En este caso se calculan los valores de todos los documentos. Como ahora no se impone la misma escala de cálculo a priori, esta prueba es aún más exigente. Sin embargo, los cálculos de Wordfish también muestran una fuerte correspondencia entre idiomas. De nuevo observamos dos grupos de partidos, los del gobierno y los de la oposición, y el representante de Izquierda Unida (IU) situado entre ellos. Solo los cálculos para los textos de los líderes del gobierno y de la oposición, que son fijos en la versión de Wordscores (y que contienen la mayoría de las palabras) se desvían un poco entre los dos idiomas. Sin embargo, se establece una alta correlación ($R=0.92$) entre los valores de las dos versiones de idiomas, confirmado así que la traducción tiene un impacto pequeño en el resultado.

FIGURA 2.
RESULTADOS DE WORFISH EN DOS IDIOMAS (PUNTUACIONES ESTANDARIZADAS)



Al comparar los resultados del análisis de Wordscores y Wordfish, podemos concluir que producen resultados muy similares (véase cuadro 1). En ambas versiones se refleja la división gobierno-oposición (que en este caso coincide con la oposición entre nacionalistas vascos y los entonces llamados constitucionalistas) y la postura ambivalente de Izquierda Unida (IU). La similitud de los resultados de Wordscores y Wordfish confirma la elección de la división gobierno-oposición como principal dimensión que subyace a la diferencia entre los discursos. También muestra que la utilización correcta de Wordscores y Wordfish produce resultados similares y verosímiles.

CUADRO 1.
CORRELACIONES ENTRE MÉTODOS E IDIOMAS

	Wordscores euskera	Wordfish español	Wordfish euskera
Wordscores español	0.99**	0.78**	0.71*
Wordscores euskera	-	0.72*	0.68*
Wordfish español		-	0.92**

Nota: *mismo método*; *mismo idioma*, *p<0.05, **p<0.01.

CONCLUSIONES

Hemos sometido a Wordscores y Wordfish, dos prometedores métodos de análisis cuantitativo de textos para la estimación de posturas ideológicas, al exigente reto de demostrar que producen los mismos resultados aplicados a textos equivalentes en dos idiomas muy diferentes: el español y el euskera. Los resultados muestran la capacidad de Wordscores y Wordfish para comparar entre idiomas y confirman totalmente los resultados obtenidos por Slapin y Proksch (2007). Pero no solo hay una correlación muy alta *intramétodos* (es decir, un método produce resultados similares en los dos diferentes idiomas) sino también *intermétodos* (es decir, ambos métodos en sus diferentes versiones idiomáticas producen unos resultados muy parecidos). El cuadro 1 muestra claramente que existen unas correlaciones muy altas ya bien comparemos métodos, idiomas o ambas cuestiones a la vez. Por otro lado, estos dos métodos no solo han producido resultados similares sino también válidos, ya que han agrupado los discursos de tal manera que han reproducido la dialéctica gobierno-oposición que entonces se daba, colocando en un lado a los partidos que apoyaban al gobierno y que votaron en contra de la moción de censura y en el otro a la oposición que votó a favor. Un caso que refuerza la validez de los métodos es la posición del representante de IU el cual, aunque entonces en la oposición, estaba bastante cercano a las tesis gubernamentales. Ambos métodos le sitúan más cerca de Ibarretxe que de los otros dos candidatos, cuya moción de censura votó en contra.

Son buenas noticias, por tanto, para el análisis comparativo de las posturas ideológicas de los actores políticos, puesto que demuestra la posibilidad de traducir textos a un mismo idioma antes de llevar a cabo el análisis. Esto, sin embargo, no significa que los métodos de análisis cuantitativo de textos como Wordscores y Wordfish estén lo suficientemente elaborados para generar comparaciones longitudinales entre países. Para detectar los puntos débiles del análisis cuantitativo de textos para estimar posiciones ideológicas en los textos, deberían aislarse los posibles puntos débiles y someterlos a los exámenes más exigentes posibles.

Referencias

Alegria, I., Artola, X., Sarasola, K. y Urkia, M. 1996. "Automatic morphological analysis of Basque", *Literary and Linguist Computing*, 11: 193-203.

- Benoit, K. y Laver, M. 2003. "Estimating Irish party policy positions using computer wordscoring: the 2002 election - a research note", *Irish Political Studies*, 18: 97-107.
- 2007. "Benchmarks for text analysis: A response to Budge and Pennings", *Electoral Studies*, 26: 130-135.
- Budge, I. y Pennings, P. 2007a. "Do they work? Validating computerised word frequency estimates against policy series", *Electoral Studies* 26: 121-129.
- 2007b. "Missing the message and shooting the messenger: Benoit and Laver's 'response'", *Electoral Studies*, 26: 136-141.
- Debus, M. 2009. "Analysing Party Politics in Germany with New Approaches for Estimating Policy Preferences of Political Actors", *German Politics*, 18: 281-300.
- Hug, S. y Schultz, T. 2007. "Left-Right Positions of Political Parties in Switzerland", *Party Politics*, 13: 305-330.
- Klemmensen, R., Binzer, S. y Ejnar, M. 2007. "Estimating policy positions using political texts: An evaluation of the wordscores approach", *Electoral Studies* 26: 746-755.
- Klüver, H. 2009. "Measuring Interest Group Influence Using Quantitative Text Analysis", *European Union Politics*, 10: 535-549.
- Laver, M. y Garry, J. 2003. "Extracting Policy Positions from Political Texts Using Words as Data", *American Political Science Review*, 97: 311-331.
- Laver, M., Benoit, K. y Sauger, N. 2006. "Policy Competition in the 2002 French Legislative and Presidential Elections", *European Journal of Political Research*, 45: 667-697.
- Lowe, W. 2008. "Understanding Wordscores", *Political Analysis*, 16: 356-371.
- Marks, G. 2007. "Introduction: Triangulation and the square-root law", *Electoral Studies*, 26: 1-10.
- Martin, L. W. y Vanberg, G. 2007. "A Robust Transformation Procedure for Interpreting Political Text", *Political Analysis*, 16: 93-100.
- Pappi, F. U. y Seher, N. M. 2009. "Party Election Programmes, Signalling Policies and Salience of Specific Policy Domains: The German Parties from 1990 to 2005", *German Politics*, 18: 403-425.
- Slapin, J. B. y Proksch, S.-O. 2008. "A scaling model for estimating time-series party positions from texts", *American Journal of Political Science* 52: 705-722.
- 2009. "Position Taking in European Parliament Speeches", *British Journal of Political Science*, 40: 587-611.
- Trask, R. L. 1996. *History of Basque*. Nueva York: Routledge.

Presentado para evaluación: 26 de noviembre de 2011

Aceptado para publicación: 19 de octubre de 2012

RAFAEL LEONISIO. UNIVERSIDAD DEL PAÍS VASCO

rafael.leonisio@ehu.es

Es investigador del equipo Euskobarometro en el Departamento de Ciencia Política de la UPV-EHU. Licenciado en Ciencias Políticas (2001) y diploma de Estudios Avanzados en Ciencia Política (2003). En la actualidad culmina sus estudios de doctorado en la Universidad del País Vasco. Sus principales áreas de investigación son los partidos políticos, el comportamiento electoral y la política vasca, sobre las que ha publicado artículos en revistas como *Regional & Federal Studies* o la *Revista Española de Ciencia Política*.

OLIVER STRIJBIS. UNIVERSIDAD DE HAMBURGO

oliver.strijbis@uni-hamburg.de

Es investigador postdoctoral en el departamento de Ciencia Política en la Universidad de Hamburgo. Licenciado en Ciencias Políticas (Zurich, 2007) y doctorado en Relaciones Internacionales y Economía Política (Universidad de St Gallen, 2012). Entre sus intereses investigadores se encuentra la identidad étnica y nacional, partidos y elecciones. Ha publicado artículos en *Parliamentary Affairs*, *Regional & Federal Studies*, la *Revista Española de Ciencia Política* y *Swiss Political Science Review*.