

La importancia del acuerdo entre codificadores para el análisis de contenido

The importance of intercoder agreement for content analysis

Macarena Peña y Lillo

University of Illinois at Urbana-Champaign
penayli2@illinois.edu

Resumen

El análisis de contenido es una técnica extensamente utilizada por los investigadores en medios de comunicación. Generalmente, los investigadores que usan el análisis de contenido cuantitativo cuentan con varios codificadores para la entrada de datos. Una pregunta clave que enfrentan es si es posible asegurar que las decisiones de los codificadores son consistentes a lo largo del análisis. Es esperable que distintas personas tengan distintas interpretaciones de los textos, lo que amenaza la confiabilidad de los resultados obtenidos y las conclusiones del estudio. Con el fin de evitar esta amenaza, los investigadores deben evaluar y reportar el acuerdo entre codificadores. En este ensayo se presentan argumentos a favor de esta práctica y se reflexiona sobre la experiencia de su uso en un caso de estudio¹.

Palabras clave: Acuerdo entre codificadores, análisis de contenido, confiabilidad, metodología cuantitativa.

Abstract

Content analysis is a widely utilized technique by researchers on mass media. Generally, researchers in quantitative content analysis rely on several coders to enter the data. A key question they face is whether it is possible to assure that the coders' decisions are consistent across the analysis. It is expected for different people to have different interpretations of the texts, which threatens the reliability of the obtained results and the study's conclusions. In order to avoid this threat, researchers must evaluate and report intercoder agreement. In this

1 El presente artículo ha sido posible gracias al aporte de CONICYT, Programa de Información Científica, Concurso Fondo de Estudios sobre el Pluralismo en el Sistema Informativo Nacional 2011, Proyecto (PLU1104) "Publicidad privada y sus implicancias para el ejercicio del periodismo y las libertades de expresión y de prensa".

essay, arguments for this practice are presented and a reflection about its use on an empirical case study is offered.

Key words:

Intercoder agreement, content analysis, reliability, quantitative methods.

El análisis de contenido es una técnica ampliamente utilizada por los investigadores en comunicación, en especial aquellos que centran su trabajo en los mensajes de los medios. La técnica es usada tanto por investigadores cualitativos como cuantitativos y en todos los casos su uso se basa en la necesidad de reducir grandes cantidades de texto (escrito, audio o audiovisual) a una cantidad de variables más limitada con el fin de establecer relaciones entre ellas y desprender significados o interpretaciones.

Para los investigadores cuantitativos, una de las principales preocupaciones al momento de utilizar el análisis de contenido es poder afirmar que todas las personas involucradas en el análisis están viendo lo mismo y, por ende, son capaces de llegar a las mismas conclusiones. Si un análisis no es capaz de asegurar aquello, las conclusiones de éste no son válidas (Neuendorf, 2002). En este breve ensayo revisaremos la lógica detrás del acuerdo entre codificadores en el contexto del análisis de contenido cuantitativo y su relevancia metodológica, y ejemplificaremos con la experiencia del análisis de prensa realizado sobre el caso de la colusión de las farmacias, cuyos resultados se presentan en esta revista.

Análisis de contenido

El análisis de contenido cuantitativo se utiliza para responder preguntas de investigación acerca de la cantidad o frecuencia de ocurrencia de ciertos patrones que pueden ser significativos desde un punto de vista comunicativo (Reinard, 2008). A diferencia de otros tipos de métodos cuantitativos cuyo foco está en la predicción o el control, el análisis de contenido cuantitativo es eminentemente descriptivo; su rol es caracterizar, por ejemplo, la cobertura de un hecho, y a partir de esa caracterización proveer explicaciones a fenómenos (Reinard, 2008).

Por lo general, el procedimiento para realizar un análisis de contenido cuantitativo comienza con la elaboración de un instrumento que contenga las variables que se van a codificar. Ese instrumento puede ser una ficha, donde se incluyan las preguntas que los codificadores deberán responder al momento de la entrada de datos. Una explicación más detallada sobre cada una de las variables puede ser provista en un libro de códigos, aunque la necesidad de éste se puede determinar de acuerdo a la complejidad de las variables en juego.

Una vez que se cuenta con el instrumento, los codificadores deben familiarizarse con él y recibir entrenamiento. Como parte del testeo del instrumento, se evalúa el nivel de acuerdo entre codificadores con una submuestra del contenido (generalmente el 10%), y una vez que

esta evaluación se considera satisfactoria, se procede al ingreso de la totalidad de los datos de la muestra. Cuando los datos están codificados, los investigadores realizan el análisis que, por lo general, incluye tablas de frecuencias y otras estadísticas de mayor complejidad.

Dado que la cantidad de texto para investigar en un análisis de contenido cuantitativo suele ser voluminosa, por lo general los investigadores recurren a un grupo de codificadores para realizar la entrada de datos. Estos codificadores pueden ser los mismos investigadores, o bien, estudiantes que son contratados para realizar el trabajo. La pregunta que enfrentan los investigadores en cualquiera de estos casos es ¿hasta qué punto se puede asegurar que todos los codificadores involucrados en el análisis están viendo lo mismo? Detrás de esta pregunta está la inquietud respecto de hasta qué punto los distintos codificadores, enfrentados al mismo material, serán capaces de clasificar esa porción de texto dentro de una categoría y no de otra. Puesto de otra manera, el interés es saber en qué medida las decisiones tomadas por los codificadores serán consistentes a lo largo del análisis. En esta reflexión está en juego el propósito central del análisis de contenido (Lombard, Snyder-Duch, & Bracken, 2002), es decir, si el objetivo del análisis es reducir grandes volúmenes de material en un número más limitado de categorías y esas categorías posteriormente van a sustentar las conclusiones del estudio, necesitamos saber con certeza si es que las categorías se han construido sobre una base sustentable. Un elemento crucial para una base sustentable es el acuerdo entre codificadores.

¿Todos leen y entienden lo mismo?

Los codificadores humanos se enfrentan a los textos desde su propia individualidad y es esperable que los análisis que realicen sean distintos en alguna medida. Los codificadores son, a la vez, consumidores de medios y, por lo tanto, su socialización con ese tipo de textos puede ser diferente.

Quienes han investigado las formas en que las personas aprenden de los medios han postulado que existen tres procesos cognitivos involucrados en el aprendizaje (Kosicki & McLeod, 1990). El primero es el de selección, es decir, los individuos, expuestos a una enorme cantidad de mensajes, deben seleccionar a qué van a prestar atención. En segundo lugar está el procesamiento activo de la información, que es la interpretación que los lectores hacen del contenido recibido. Finalmente, está la integración reflexiva, que ocurre con posterioridad a la exposición, cuando el lector conecta la nueva información recibida con información previamente almacenada en su memoria. La integración reflexiva tiene lugar también en las interacciones con otras personas y es producto de las conversaciones que tiene el individuo con su entorno (Kosicki & McLeod, 1990). Teniendo en cuenta estos elementos es posible entender que las fuentes de desacuerdo son inherentes a la codificación realizada por personas en el análisis de contenido. No obstante, el beneficio de que sean codificadores humanos los que realicen el análisis y no herramientas computacionales, es que las personas son capaces de

entender sutilezas del lenguaje que los *software* probablemente pasarían por alto (Matthes & Kohring, 2008).

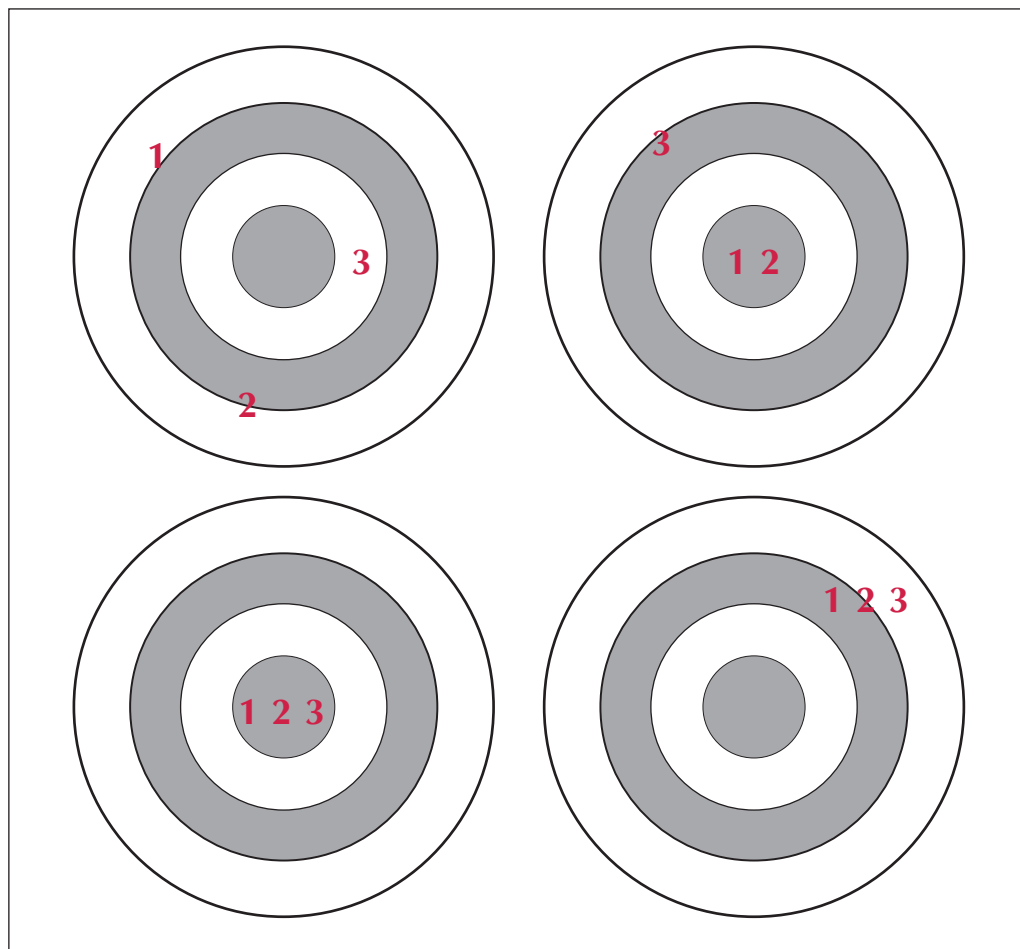
Para ejemplificar cómo el acuerdo o desacuerdo entre codificadores toma lugar en el contexto de un análisis de contenido, usaré como referencia los diagramas de la figura 1. Imaginémosnos que un investigador contrata a tres estudiantes para realizar la codificación de 500 notas de prensa. La diana en el diagrama representa el concepto al que el investigador está tratando de llegar con su análisis. Para ello construye un ítem o una serie de ítems basados en la literatura sobre el concepto y le pide a los codificadores que, luego de leer cada uno de los artículos de prensa que componen su muestra, decidan si ese concepto está o no presente en cada uno de ellos. Al centro de la diana está la realidad objetiva y certera al respecto; en otras palabras, si es que la nota contiene o no el elemento estudiado (no olvidar que el análisis de contenido cuantitativo se sustenta en el paradigma post-positivista y por ende, parte de la base de que existe una realidad externa que el investigador puede constatar²), mientras más alejado del centro, menos certera es la evaluación de los codificadores respecto del concepto y su presencia en las notas estudiadas.

La situación ideal para el investigador es que los tres codificadores, ante el mismo texto, sean capaces de ver lo mismo y que eso que vean, sea lo más cercano posible a lo que realmente ocurre. Esta situación está representada en el diagrama c), donde las observaciones de los codificadores 1, 2 y 3 están todas al centro de la diana. El hecho de que las observaciones se sitúen al centro, es decir, que los estudiantes codifiquen exactamente lo que ocurre en la realidad, está en función de la construcción del instrumento utilizado por el investigador (validez³) y no se relaciona directamente con el rol de los codificadores. Un ejemplo de eso es lo que presenta el diagrama d), donde los tres codificadores ven lo mismo, pero lo que ellos ven está distorsionado en relación a lo que efectivamente ocurre. Por lo general, sin embargo, al usar instrumentos previamente validados por otros investigadores, la amenaza de que el instrumento no mida lo que se supone debería medir se reduce considerablemente. El diagrama a) da cuenta de los tres codificadores viendo algo distinto cuando se enfrentan a un cierto texto, mientras que en b) los codificadores 1 y 2 ven lo mismo y el codificador 3 discrepa.

2 Ver una explicación más acabada al respecto en White y Marsh (2006).

3 Una explicación más acabada al respecto va más allá del alcance de este ensayo. Ver, por ejemplo, Muñiz (2005) para una explicación del concepto de validez desde la psicometría.

Figura 1. Diagrama de las posibilidades de acuerdo entre codificadores



Como se indica más arriba, las diferencias entre individuos son idiosincráticas y dado que la codificación es crucial para la investigación que se hace, cualquier tipo de desacuerdo genera distorsiones en el análisis de contenido y, por lo tanto, en sus conclusiones. De ahí la necesidad de evaluar el nivel de acuerdo entre codificadores antes de realizar la codificación de la muestra total y de reportar los resultados de esa evaluación en la publicación de las conclusiones de la investigación. El objetivo de lo primero es conocer de antemano si es que existen falencias y, si es necesario, enmendar el rumbo; ya sea revisar el instrumento, volver a entrenar a los codificadores, descartar a uno de ellos, o todas las anteriores. La importancia de reportar los resultados de la evaluación del acuerdo entre codificadores es porque con ello se entrega a los lectores información sobre la confiabilidad de los resultados reportados y las interpretaciones que se desprenden de éstos.

Los estándares

Existen ciertos estándares en cuanto a la información que los autores deben reportar cuando se da cuenta de resultados de evaluación de acuerdo entre codificadores. Lombard, Snyder-Duch y Bracken (2002) revisaron 200 artículos de revistas de comunicación en habla inglesa entre 1994 y 1998 que utilizaron el análisis de contenido como metodología y encontraron que solamente el 69% de los artículos revisados daba cuenta de algún cálculo de acuerdo entre codificadores. Los autores encontraron que de los artículos que reportaban algún indicador de acuerdo entre codificadores, muchos omitían información relevante, como el tamaño de la submuestra sobre la cual se calculaba el porcentaje de acuerdo entre codificadores, el número de codificadores, el nivel de acuerdo para cada una de las variables por separado (en vez del acuerdo global que puede esconder variables con muy bajo nivel de acuerdo), la cantidad de entrenamiento que los codificadores recibieron y si es que las discrepancias entre los codificadores habían sido resueltas con más entrenamiento u otras medidas. Todos estos datos son relevantes al momento de juzgar la calidad del análisis y de las conclusiones de éste.

Entre los métodos para calcular el acuerdo entre codificadores (Lombard et al., 2002) se cuentan el simple porcentaje de acuerdo, y el método de Holsti, que es una corrección del simple porcentaje de acuerdo a través de una serie de variables. Estas dos son las mediciones más comúnmente reportadas por los estudios que realizan el cálculo de acuerdo, aun cuando son consideradas muy liberales e incluso infladas, pues no dan cuenta del acuerdo entre codificadores alcanzado sólo por azar (Krippendorff, 2004; Lombard et al., 2002). Las estadísticas que consideran el efecto del azar, y por ende son consideradas más conservadoras y confiables son el Pi de Scott, el alpha de Cohen y el alpha de Krippendorff. El problema de estos tres últimos métodos es que requieren un arduo trabajo si es que se calculan a mano o un amplio conocimiento de *software* estadísticos si se elige utilizar un computador para realizar el cálculo.

Es claro que la sofisticación estadística de los métodos propuestos para calcular y reportar el acuerdo entre codificadores reviste una importante barrera de entrada para muchos investigadores que no tienen mayor entrenamiento en esta área, pero eso no debería desalentarnos para dar cuenta de la rigurosidad del trabajo. El porcentaje de acuerdo entre codificadores es una medición al alcance de la mano de cualquiera con un manejo intermedio de Excel u otro *software* de cálculo. Esta medición simplemente da cuenta del porcentaje de veces en que los codificadores ven lo mismo enfrentados al mismo material. En otras palabras, el porcentaje de veces que, ante la misma pregunta y ante el mismo texto, la respuesta de los codificadores coincide.

A pesar de que es considerado muy liberal, el porcentaje de acuerdo es informativo, toda vez que entrega una guía general acerca de las divergencias que la codificación presenta y es especialmente relevante para tomar decisiones cuando se encuentran niveles de acuerdo considerablemente bajos. Esto es importante, pues aun cuando los investigadores han puesto

mucho esfuerzo en el diseño de instrumentos, del libro de códigos y en el entrenamiento de los codificadores, las diferencias intrínsecas entre los individuos juegan malas pasadas. Al menos esa es la experiencia que tuvimos al evaluar el acuerdo entre codificadores en el análisis de contenido sobre el caso de la colusión de las farmacias reportado en este número.

La experiencia en el caso farmacias

La muestra original de noticias sobre el caso colusión de las farmacias era de aproximadamente 500. Fueron reclutados tres codificadores para el trabajo, dos de ellos eran estudiantes de periodismo y la tercera, egresada de la carrera en proceso de titulación. Los codificadores fueron entrenados en dos sesiones presenciales con el fin de que se familiarizaran con los textos que debían analizar, el tema y la fecha de codificación. La ficha estaba disponible en *SurveyMonkey*, un servidor *online* para encuestas. Decidimos utilizar ese formato, pues permitía exportar los datos recolectados inmediatamente a *Excel* y así evitar una segunda codificación que pudiese dar lugar a errores en la entrada de datos. Además, era un formato amigable con el que los codificadores se familiarizaron rápidamente. La ficha había sido desarrollada por el equipo de investigadores durante varios meses de trabajo y en base a la revisión de literatura disponible.

Para realizar evaluación de acuerdo entre codificadores tomamos una submuestra de alrededor de un 10% de esta muestra siguiendo los lineamientos descritos por Neuendorf, (2002). Algunas notas tuvieron que ser descartadas del análisis y finalmente terminamos revisando 46. Los tres ayudantes codificaron cada una de las notas que componían la submuestra; es decir, se analizaron las decisiones tomadas por cada uno de los codificadores para cada nota que componía la submuestra.

A esas alturas, la ficha de codificación tenía 167 ítems, de los cuales siete requerían ingreso de texto y no fueron consideradas en el análisis debido a que es imposible cuantificarlas. Las restantes variables eran en su mayoría preguntas que se respondían No (= 0), Si (= 1), o No se puede determinar (= 2). Durante 10 días los ayudantes codificaron las notas. Una vez que terminaron, la hoja de cálculo con sus respuestas fue descargada y organizada con el fin de realizar el análisis. La figura 2 muestra la manera en que la hoja de cálculo fue organizada.

En primera instancia, las tres entradas correspondientes a una misma nota fueron agrupadas. Luego, se agregó una fila al final de las entradas para cada nota para indicar si es que los tres codificadores habían marcado lo mismo. Se contabilizaron las veces en que existía acuerdo y se obtuvo un porcentaje para cada variable en razón de la cantidad de notas analizadas.

Figura 2. Formato de los datos para el análisis de acuerdo entre codificadores

NOTA	CODIFICADOR	VARIABLE 1	VARIABLE 2
Nota 1	Codificador 1	1	1
	Codificador 2	1	2
	Codificador 3	1	0
Acuerdo nota 1		(¿son 1,2,3 iguales? =SI)	(¿son 1,2,3 iguales? =NO)
Nota 2	Codificador 1	2	
	Codificador 2	2	1
	Codificador 3	2	0
Acuerdo nota 2		(¿son 1,2,3 iguales? =SI)	(¿son 1,2,3 iguales? =NO)

En esta primera evaluación el porcentaje total de acuerdo a través de las 160 variables analizadas (sin incluir siete preguntas donde se requería ingresar texto) fue de un 73.07%. En 60 ítems el porcentaje de acuerdo se encontraba sobre el 90%. En 23, sin embargo, el nivel de acuerdo era menor al 50%, es decir, los codificadores coincidieron menos de la mitad de las veces en esas respuestas.

Algunos de los ítems con bajos niveles de acuerdo eran aquel en que preguntábamos si la nota daba cuenta de un conflicto entre el Estado y empresas (4.35%), la pregunta sobre si las partes en conflicto se reprochaban mutuamente (6.52%), si la nota hablaba de un conflicto entre empresas y consumidores (8.70%), y si es que la nota entregaba prescripciones de cómo comportarse (17.39%). Ante estos resultados, decidimos que no estábamos en condiciones de seguir adelante con la codificación sin antes revisar nuestro instrumento y volver a capacitar a los codificadores, ya que los bajos niveles de acuerdo no nos permitirían saber con claridad si es que lo que estábamos concluyendo era realmente confiable.

Se decidió evaluar nuevamente el acuerdo entre codificadores después de una revisión del instrumento y dos nuevas jornadas de capacitación con nuestros codificadores, donde resolvimos dudas y explicamos cada una de las preguntas que componían el esquema de codificación.

Luego de la revisión del instrumento, descartamos una serie de ítems y dejamos una ficha de 147 ítems más los siete en los que los codificadores debían ingresar texto. La entrada de datos en esta segunda oportunidad tomó 12 días y la submuestra estuvo compuesta de 69 notas. El nivel de acuerdo global en el segundo intento no cambió sustantivamente, de hecho, se redujo a 72.39%. Al examinar más detalladamente los resultados, nos dimos cuenta de que la tercera codificadora discrepaba sistemáticamente de los otros dos codificadores (una situación similar a la que grafica la imagen b) de la figura 1) y que al remover las observaciones de esta codificadora del análisis, el acuerdo global entre los codificadores subía a 87.07%, por lo que decidimos excluir a la tercera ayudante de la codificación del total de la muestra y destinar su trabajo a otras áreas del proyecto. No nos sorprendió que la tercera codificadora tuviese

respuestas distintas a las de los otros dos codificadores, pues los codificadores 1 y 2 tienen más elementos comunes entre sí (similar edad, mismo curso en la carrera, etc.) mientras que la codificadora 3 tiene otro contexto (ya egresada, con experiencia laboral en los medios).

Aun cuando al descartar a un codificador el acuerdo global mejoraba considerablemente, sabemos que lo recomendado no es dar cuenta del acuerdo general, sino que niveles de acuerdo por cada concepto. Por esto, y porque entendemos que el porcentaje de acuerdo es una medida liberal que no da cuenta del acuerdo ocurrido solo por azar, analizamos uno a uno los ítems que componían el instrumento y decidimos descartar aquellos donde el nivel de acuerdo no superaba el 80%. La versión definitiva de la ficha de codificación, que fue la que se utilizó para el análisis final de toda la muestra, consistió en 106 variables más aquellas donde los codificadores debían ingresar texto y el acuerdo general alcanzó al 94%.

Realizar las dos evaluaciones de acuerdo entre codificadores tomó al menos ocho semanas y una serie de reuniones del equipo de investigación y con los codificadores para revisar el instrumento, el material para el análisis, identificar las fuentes del desacuerdo e intentar resolverlas. A pesar de la inversión en tiempo y trabajo, la evaluación de la experiencia es positiva, ya que permitió identificar falencias que hubiesen distorsionado enormemente los posibles resultados obtenidos.

Conclusión

El acuerdo entre codificadores es esencial para el análisis de contenido cuantitativo. El análisis de contenido es una metodología que permite analizar grandes cantidades de texto, describirlo detalladamente e interpretar los patrones encontrados. La codificación llevada a cabo por personas en vez de programas computacionales permite un nivel más sutil de análisis, pero está sujeta a errores que pueden amenazar la validez de las conclusiones de la investigación. Al realizar un análisis de contenido cuantitativo, los investigadores deben preguntarse si es que todos aquellos que están llevando a cabo el análisis están viendo lo mismo. Si los investigadores no pueden asegurar que las decisiones tomadas por los codificadores son consistentes, entonces los resultados del análisis son cuestionables.

La experiencia en el análisis de contenido sobre la cobertura de la prensa chilena del caso colusión de las farmacias demuestra que, aun cuando se diseñan los instrumentos cuidadosamente y se invierte tiempo en entrenar a los codificadores, las discrepancias entre éstos no son infrecuentes. Las personas se relacionan con los textos de distintas maneras y, dado que el análisis de contenido es una técnica basada en la interpretación, las subjetividades juegan un rol relevante al momento de realizar la codificación.

Con el fin de asegurar que el análisis no tenga mayores sesgos, los investigadores deben evaluar el nivel de acuerdo entre codificadores antes de realizar el análisis total. Si bien es una inversión en tiempo y trabajo, en especial cuando los resultados iniciales no son los esperados,

como en el caso del estudio del caso colusión, es una precaución adecuada para asegurar que las conclusiones que se desprendan serán legítimas y no viciadas por ruidos imposibles de controlar. Más aun, para sustentar los descubrimientos hechos a partir de un análisis de contenido cuantitativo, es necesario que los investigadores reporten los niveles de acuerdo para cada variable al momento de publicar los resultados de sus estudios.

Referencias bibliográficas

- Kosicki, G. M. y McLeod, J. M. (1990). Learning from political news: Effects of media images and information-processing strategies. En S. Kraus (Ed.), *Mass Communication and Political Information Processing* (pp. 69-83). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Krippendorff, K. (2004). Reliability in Content Analysis. *Human Communication Research*, 30(3), 411-433. doi: 10.1111/j.1468-2958.2004.tb00738.x
- Lombard, M., Snyder-Duch, J., y Bracken, C. C. (2002). Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research*, 28(4), 587-604. doi: 10.1111/j.1468-2958.2002.tb00826.x
- Matthes, J., & Kohring, M. (2008). The Content Analysis of Media Frames: Toward Improving Reliability and Validity. *Journal of Communication*, 58(2), 258-279. doi: 10.1111/j.1460-2466.2008.00384.x
- Neuendorf, K. A. (2002). *Communication Methods and Measures*. Thousand Oaks, CA: Sage.
- Reinard, J. C. (2008). *Introduction to Communication Research* (4th ed.). New York, NY: McGraw-Hill.
- White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library Trends*, 55(1), 22-45. doi: 10.1353/lib.2006.0053