

## EL INFORME DE ARBITRAJE SEGÚN LA RECOMENDACIÓN DE PUBLICACIÓN Y LA PRODUCTIVIDAD DE LOS EVALUADORES

*THE REFEREE REPORT ACCORDING TO PUBLICATION RECOMMENDATION AND REVIEWERS'  
ACADEMIC PRODUCTIVITY*

**Omar Sabaj Meruane\***; **Carlos González Vergara\*\***; **César Astudillo Zepeda\***; **Germán Varas Espinoza\*\***; **Miguel Fuentes Cortés\*\***; **Paula Cabezas del Fierro\***; **Karem Squadrito Díaz\*\***; **Álvaro Pina-Stranger\*\*\***

\* Universidad de La Serena; \*\* Pontificia Universidad Católica de Chile; \*\*\* Université de Rennes 1, Francia;

omarsabaj@userena.cl; cgonzalv@uc.cl; astudillozpedac@gmail.com; gnvaras@uc.cl; mafuentes1@uc.cl; pcabezas@userena.cl; krsquadrito@uc.cl; alvaro.pina-stranger@univ-rennes1.fr

### Historia editorial

Recibido: 01-11-2016

Primera revisión: 13-12-2017

Aceptado: 15-02-2018

Publicado: 02-07-2018

### Palabras clave

Proceso de revisión por pares

Informes de arbitraje de artículos de investigación

Escritura científica

Análisis del discurso

### Resumen

El informe de arbitraje de artículos de investigación es un género en el que se cristaliza una parte del proceso colectivo de construcción del conocimiento. Debido a su carácter confidencial, existen muy pocos trabajos que aborden empíricamente este género. A su vez, esos pocos estudios tienen una serie de limitantes, como la falta de definición de la unidad de análisis y la falta de integración de los datos discursivos con los datos sociales de quienes los producen. En este trabajo, presentamos un modelo que busca superar limitantes. Aplicamos este modelo para describir cómo varía el informe de arbitraje según dos atributos extralingüísticos, la recomendación de publicación (aceptar, condicionar o rechazar) y el número de artículos publicados por los evaluadores. Los resultados muestran que ambos atributos extralingüísticos se asocian a cambios específicos del discurso de este género, siendo la recomendación de publicación más determinante que la productividad de los evaluadores.

### Abstract

The referee report of research articles is a discursive genre in which a part of the collective process of building knowledge is crystalized. Given its confidentiality, there is little empirical work on this genre. Among these few contributions, a series of limitations can be identified, such as the lack of definition of the unit of analysis or the lack of integration between discursive and sociometric data. In this article, we present a model that contributes to overcoming these limitations. The model was applied with the purpose of analyzing the variation of referee reports according to two extralinguistic variables, i.e., publication recommendation (accepted, conditioned, rejected) and referee's number of published articles. The results showed that both extralinguistic attributes were associated with specific discourse variations, of which publication recommendation was more decisive than referee's academic productivity.

### Keywords

Peer review process

Referee report of scientific articles

Scientific writing

Discourse analysis

Sabaj Meruane, Omar; González Vergara, Carlos; Astudillo Zepeda, César; Varas Espinoza, Germán; Fuentes Cortés, Miguel; Cabezas del Fierro, Paula; Squadrito Díaz, Karem & Pina-Stranger, Álvaro (2018). El informe de arbitraje según la recomendación de publicación y la productividad de los evaluadores. *Athenea Digital*, 18(2), e2051. <https://doi.org/10.5565/rev/athenea.2051>

## Introducción

---

La ciencia es una construcción social y colectiva (Lamont, 2009; Latour, 1992; Lazega, 1992), en la que participan actores con intereses diversos. Una de las formas más directas para dar cuenta de la construcción social del conocimiento científico como una interacción es estudiar el proceso de evaluación por pares (en adelante, PEP), específicamente, a través de la descripción del 'Informe de Arbitraje' (en adelante, IA), puesto que es en este género donde se materializa la interacción discursiva entre los actores que participan en dicha construcción.

En el contexto hispanoamericano, la tarea de sustentar y desarrollar revistas académicas es especialmente desafiante. Esto porque, en general, las instituciones que albergan estas revistas no estiman de manera adecuada la carga de trabajo que dicha tarea implica. Dado el carácter confidencial de este proceso, no existe mucha información empírica que permita mejorar el sistema. En este contexto, el estudio de los informes de arbitraje puede ser un aporte importante para comprender de manera más profunda el PEP y para ayudar a visibilizar la relevante tarea que los equipos editoriales realizan en la producción, evaluación y diseminación del conocimiento científico.

En este género, de carácter anónimo, se materializan los juicios que un evaluador realiza sobre un artículo de investigación (Bolívar, 2008). Dado su carácter privado (Swales, 1996), la descripción empírica de este género ha sido muy escasa (Astudillo, Squadrito, Varas, González y Sabaj, 2016), y los enfoques que se han utilizado presentan limitaciones que parcializan de manera excesiva el objeto de estudio (Sabaj, González y Pina-Stranger, 2016).

En este contexto, en el presente trabajo exponemos, por un lado, el proceso de construcción de un modelo para el análisis de IA de artículos de investigación en el PEP que supera en parte esas limitaciones. Por otro lado, mostramos la aplicación de ese modelo a un *corpus* de 318 IA y a una submuestra de 118 informes, proveniente de tres revistas chilenas de alcance internacional.

En la primera parte del trabajo, exponemos nuestra aproximación general al PEP, como una práctica sociodiscursiva. Posteriormente, mostramos los principales trabajos que se han dedicado al análisis de los IA, especificando sus limitaciones. Luego, describimos los pasos que seguimos en la confección de nuestro modelo de análisis y sus categorías constituyentes. Por último, damos cuenta de la aplicación del modelo, los resultados e interpretaciones de esa aplicación y las conclusiones generales del trabajo.

## El proceso de evaluación por pares (PEP)

---

En términos sencillos el PEP se realiza en los siguientes pasos: a) un autor (o un grupo de autores) envía un trabajo a una revista; b) el editor selecciona evaluadores para que emitan un IA con una recomendación de publicación; c) con esas recomendaciones, el editor toma una decisión que informa al autor; d) si el trabajo es rechazado el proceso concluye; por el contrario, si el trabajo es aceptado, los autores realizan las correcciones sobre la base de las revisiones de los evaluadores y del editor y vuelven a enviar una versión para ser publicada (Bornmann, 2011; Campanario 1998a; 1998b; Garfield, 1986; 1987).

En nuestra concepción (Sabaj et al., 2016), el PEP es una práctica sociodiscursiva de carácter técnico y evaluativo que implica una coordinación epistémica entre distintos actores, a saber, los autores, los evaluadores y los editores. Esta práctica, que determina la generación y diseminación del conocimiento científico, es técnica porque depende de una experticia específica de un campo de conocimiento; es evaluativa porque implica un juicio de valor por parte de los árbitros; y, es discursiva, en tanto todo el proceso se cristaliza en textos que se intercambian entre los actores. Cada actor posee ciertas características cientométricas y sociodemográficas y produce un tipo de texto específico, que, a su vez, tiene atributos determinados. Así, es al autor al que se le asocia el manuscrito original; al editor, la solicitud de evaluación; y a los evaluadores, el IA.

### Modelos para el análisis del discurso de los IA en el PEP

El IA es un texto de carácter confidencial que se utiliza en el proceso de evaluación por pares. Justamente, por su naturaleza privada (*occluded genre*, en términos de John Swales, 1996), es un género prácticamente imposible de estudiar sin la ayuda de los editores. Por ser el primer artículo en español que aborda el tema desde un enfoque discursivo, el trabajo de Adriana Bolívar (2008) es un referente obligatorio. La autora presenta una caracterización general del IA, en la que se destaca su carácter reactivo y dependiente de otros textos (es parte de una cadena de géneros), y que su confección está determinada por el tiempo, en tanto los editores exigen que se realice en un lapso determinado. Se sostiene también que, en el IA, se evalúa de manera global el trabajo, pero también el grado de conocimiento del autor sobre una materia y sobre el género. Así también, “es evaluada la habilidad comunicativa y retórica de los autores con base en el texto que ellos han escrito” (Bolívar, 2008, p. 47).

En un detallado trabajo, Hugh Gosden (2003) hace un análisis de contenido y funcional de los comentarios presentes en 40 IA de una revista en el ámbito de las ciencias

duras. En el trabajo, se usan principalmente categorías de contenido, esto es, elementos que reflejan el objeto de la evaluación en los comentarios. Como conclusión, se establece que las categorías relativas a las afirmaciones, la discusión y las referencias son las que más frecuentemente se utilizan en los IA.

Siguiendo en parte el modelo de Gosden (2003), Inmaculada Fortanet (2008) realiza un análisis de lo que denomina patrones de evaluación, que corresponden a propósitos comunicativos que realizan los evaluadores en 50 IA de lingüística aplicada y del ámbito de la organización en las empresas. Entre los resultados se establece que la crítica es la función más frecuente en las dos disciplinas analizadas.

Tanto en el trabajo de Mohammad Tharirian y Elham Sadri (2013) como en el de Philippa Mungra y Pauline Weber (2010) se utilizan solo categorías de contenido, esto es, que representan los objetos evaluados. A su vez, estos comentarios son clasificados en aspectos de contenido o aspectos de forma. En el caso de Mungra y Weber (2010), se analizan 33 IA provenientes del área de la salud, específicamente, de las subespecialidades de neurología, dermatología, medicina interna y cirugía. De los resultados se concluye que la mayoría de los comentarios son asuntos de contenido, específicamente, en el área de la relevancia científica y aspectos metodológicos, seguidos de errores gramaticales.

Tharirian y Sadri (2013), por su parte, analizan 82 IA de ingeniería, medicina y ciencias sociales. Al igual que Mungra y Weber (2010), los autores llegan a la conclusión que la mayoría de los problemas que aparecen en los comentarios decían relación con las categorías asociadas al contenido.

El trabajo de Bolívar (2011) presenta un conjunto de funciones discursivas de la evaluación negativa en este género: detractiva o indicar algo que falta; correctiva o señalar lo que está mal o incorrecto; la función epistémica o indicar aquello que no está claro y la función relacionante que se refiere al modo en que los investigadores se presentan y posicionan como miembros de un grupo. Cada una de estas funciones se asocia a una serie de categorías de contenido, como la metodología, la representatividad de la muestra, los títulos, entre muchos otros.

## Cuatro problemas en el análisis del discurso de los IA

---

En esta sección, retomamos algunas críticas que hemos hecho sobre el PEP (Astudillo, 2015; Astudillo, Squadrito, Varas, González y Sabaj, 2016; Sabaj et al., 2016) y exponemos otras que justificaron la creación de un modelo para el análisis del género textual que nos ocupa.

Un primer y gran desafío que enfrenta cualquiera que desea analizar datos discursivos es la delimitación de la unidad de análisis (Paltridge, 1996; Swales, 2004). En el caso de los IA, se ha utilizado como unidad de análisis la noción de comentario, pero esta no aparece definida o bien la definición que se propone es imprecisa o parcial. Por ejemplo, en los trabajos de Mungra y Weber (2010) y de Tharirian y Sadri (2013), la noción de comentarios se limita a aquellos extractos textuales que implican una modificación por parte del autor, es decir, se dejan fuera todos los actos evaluativos que, siendo parte del género, no implican ningún cambio por parte de los autores. En el caso de Gosden (2003), la unidad de análisis comentario se equipara a la oración, pero sabemos que una oración puede contener más de un comentario. Solo en el trabajo de Fortanet (2008) se define la unidad de análisis en términos de un acto evaluativo como “cualquier unidad estructural, independiente de su configuración léxico-gramatical, que contiene tanto el (sub)aspecto sobre el cual se comenta como la acción que se recomienda” (Fortanet, 2008, p. 29).

Una segunda dificultad de los trabajos que describen el discurso del IA es la falta de integración de las distintas dimensiones que se analizan. Tal como hemos expuesto, la mayoría de los trabajos se focalizan en aspectos del contenido y funcionales. Solo en uno de los trabajos (Bolívar, 2011), se incluye además la dimensión de la valoración. Este mismo trabajo es el único que integra dos dimensiones, asociando la dimensión de la valoración a las funciones comunicativas y, de manera indirecta, a los contenidos de la evaluación, pero en términos generales no existen propuestas que permitan dar cuenta, a la vez, qué es lo evaluado, cuáles son los propósitos comunicativos o actos que se realizan al evaluar y cuál es la polaridad de los juicios (positiva, negativa o neutra) en los IA.

El tercer problema de los trabajos sobre el IA, que también puede generalizarse al campo general de los estudios del discurso, se relaciona con la ausencia de procesos de validación o transferibilidad de los modelos analíticos. Esto es sumamente relevante porque el análisis de datos discursivo es una tarea que depende fuertemente de los conocimientos previos del analista, por lo que validar el análisis se hace imprescindible. Es común que los trabajos en análisis del discurso no aborden el tema de la validez lingüística de las descripciones e interpretaciones que se realizan de los datos textuales.

Una cuarta limitante de los trabajos dedicados al análisis discursivo de los IA se refiere a la falta de información extralingüística sociorrelacional de los sujetos que participan en la generación de estos discursos. En términos generales, para caracterizar a los sujetos, el análisis del discurso utiliza categorías sociológicas de alcance sociodemográfico, como el sexo, la edad, o la inclusión en un grupo étnico o socioeconómico, entre otras propiedades. A diferencia de los sociodemográficos, los atributos so-

correlacionales permiten determinar la posición que ocupa un sujeto en relación con todos los miembros de su comunidad, a partir de un parámetro que permite ubicar a un individuo en un ranking. Por ejemplo, la cantidad de artículos, la cantidad de proyectos adjudicados, las citas recibidas, entre otros, son rasgos que nos permiten saber quién es quién en una comunidad disciplinar específica. La razón por la cual estos datos son relevantes se basa en la idea de que tanto las características sociodemográficas, pero también y, sobre todo, las sociorrelacionales determinan la naturaleza de los discursos de los IA (Varas, 2015). En este sentido se esperaría, por ejemplo, que la evaluación realizada por un investigador senior sea sustancialmente distinta de la de un investigador en formación.

En síntesis, algunas críticas de los modelos analíticos sobre el informe de arbitraje son la falta de delimitación de la unidad de análisis, la no integración de las distintas dimensiones (contenido, función y valoración), la ausencia de criterios de validación y la falta de información sociorrelacional.

### Etapas de la construcción de un modelo para el análisis de los IA en el proceso de evaluación por pares

Con el objetivo de superar esas críticas, construimos un modelo para el análisis y descripción del discurso de los IA. Este proceso se realizó en dos etapas:

**i) Construcción de un modelo inicial:** El modelo inicial se construyó a partir del análisis incremental de 15 IA, cuya recomendación de publicación era rechazar los artículos (Astudillo, 2015). En esta fase del proceso participaron los ocho investigadores que trabajaron en parejas en la construcción, ajuste y aplicación del modelo. Tanto para la construcción del modelo, como para su aplicación a la muestra se utilizó el Software Atlas Ti V.7.6.2. Primero, se realizó un listado con las categorías de análisis de otros trabajos (Ardakan, Mirzaie y Sheikhshoaei, 2011; Bolívar, 2011; Fortanet, 2008; Gosden, 2003; Mungra y Weber, 2010), las cuales fueron aplicadas a un *corpus* reducido de 5 informes (los mismos) por cada una de las cuatro parejas.

Dada la dificultad de aplicar ciertas categorías y del hecho de que partes de los IA quedaban sin análisis, se realizaron ajustes, siguiendo los preceptos de la teoría empíricamente fundada (Strauss y Corbin, 1997), para crear, separar o fundir algunas de las categorías. Tras los ajustes se volvieron a aplicar las nuevas categorías a los informes ya analizados y se incorporó el etiquetaje de otros 5 informes. El proceso se reiteró hasta que las cuatro parejas habían analizado 15 informes. Ese modelo inicial que contenía 65 categorías.

**ii) Construcción del modelo final:** La confección del modelo final tuvo como propósito ampliar la aplicación del modelo a IA con otras decisiones (condicionados y aceptados). En esta etapa se analizaron de manera incremental 25 informes de artículos condicionados y 25 de artículos aceptados, ajustando, de la misma forma descrita anteriormente, el modelo de análisis. El modelo final quedó constituido por 71 categorías. Todos los cambios realizados en esta fase final y el modelo completo con todas sus categorías aparecen en el Anexo 1.

La validación del modelo se realizó para dos versiones descritas, el modelo inicial y el final. En la primera validación, se calculó la confiabilidad del modelo, usando el Alfa de Cronbach. Para ello se solicitó a los ocho investigadores que registraran su grado de acuerdo en relación con la aplicación de cada una de las 65 categorías a un ejemplo. El valor mínimo aceptable para el coeficiente Alfa de Cronbach es 0,70. El valor de alfa calculado fue de 0,904 lo que permitió establecer que el modelo inicial era confiable.

En una segunda instancia, se calculó la validez de la aplicación del modelo final, a través del cálculo del acuerdo inter-codificador (Fleiss, 1971). Para ello se utilizaron los datos de la aplicación de las 71 categorías a diez informes que habían sido etiquetados por todas las parejas. El resultado del cálculo del acuerdo fue de 0,65, lo que permite establecer que había acuerdo sustancial entre los observadores. Estos dos procedimientos (el cálculo de la confiabilidad y el acuerdo) permiten establecer que el modelo era consistente e independiente de los etiquetadores, lo que suple uno de los problemas indicados en la sección “Modelos para el análisis del discurso de los IA en el PEP”.

## Un modelo para el análisis de los IA del proceso de evaluación por pares de artículos de investigación

El modelo que hemos propuesto se aplica a un tipo específico de comentarios (los comentarios evaluativos). Por ello, dividimos su exposición en dos partes. En primer término, describimos algunas categorías generales que se agrupan de acuerdo a algunos criterios. En la segunda parte, definimos los comentarios evaluativos y presentamos cada una de las capas o dimensiones del modelo que utilizamos para analizarlos.

## Categorías generales

### Comentarios según el tema

Durante el proceso de construcción del modelo detectamos comentarios que no tenían por objeto el artículo o una parte de él, sino que abordaban otras materias o cumplían otras funciones. Se trata de comentarios ajenos al tópico que no versan sobre el artículo, sino que corresponden a recomendaciones generales (1), ordenadores u organizadores de los textos propios (2), comentarios sobre la experticia del autor (3), interacciones directas entre el evaluador y el autor (4) o indicaciones de partes del trabajo (5):

(1) Cuando se escribe un artículo se debe hacer de forma ordenada.

(2) **En primer lugar** me referiré a la calidad teórica del trabajo, para evaluar **luego** sus resultados.

(3) Parece la labor de un amateur.

(4) El evaluador solicita se le haga llegar el audio analizado para corroborar lo que el autor dice.

(5) Primera línea de la introducción.

### Comentarios según la presencia de elementos evaluativos: descriptivos versus evaluativos

Los comentarios evaluativos corresponden a un juicio valorativo con polaridad positiva o negativa que se expresa por medio de una oración, una cláusula subordinada o en un sintagma no clausular. Por ejemplo, en 6, aparecen dos comentarios evaluativos (marcamos con corchetes los límites de cada comentario evaluativo):

(6) [El trabajo es muy interesante] y [está bien escrito]

En los IA también aparecen otros tipos de comentarios que carecen de valoración. Estos son los comentarios descriptivos (7). Se trata de aserciones que se expresan mediante una oración, una cláusula subordinada o en un sintagma no clausular en la que se puede identificar secuencias descriptivas en las que no existen marcas de valoración.

(7) (El trabajo tiene como objetivo analizar los titulares de la prensa)

También pueden aparecer comentarios complejos compuestos por comentarios descriptivos y evaluativos (8):

(8) [La Introducción], (que aparece en la primera página), [está bien escrita].

Describimos, a continuación, de manera general, el modelo de análisis confeccionado, el cual, como hemos sostenido, se aplica a los comentarios evaluativos no ajenos, estén o no asociado a un constructo. Para esta descripción general, mostramos solo algunas categorías. El modelo completo, con cada una de las categorías por dimensión, aparece en el Anexo 1. En la Figura 1, se puede observar un esquema general con el modelo de análisis.

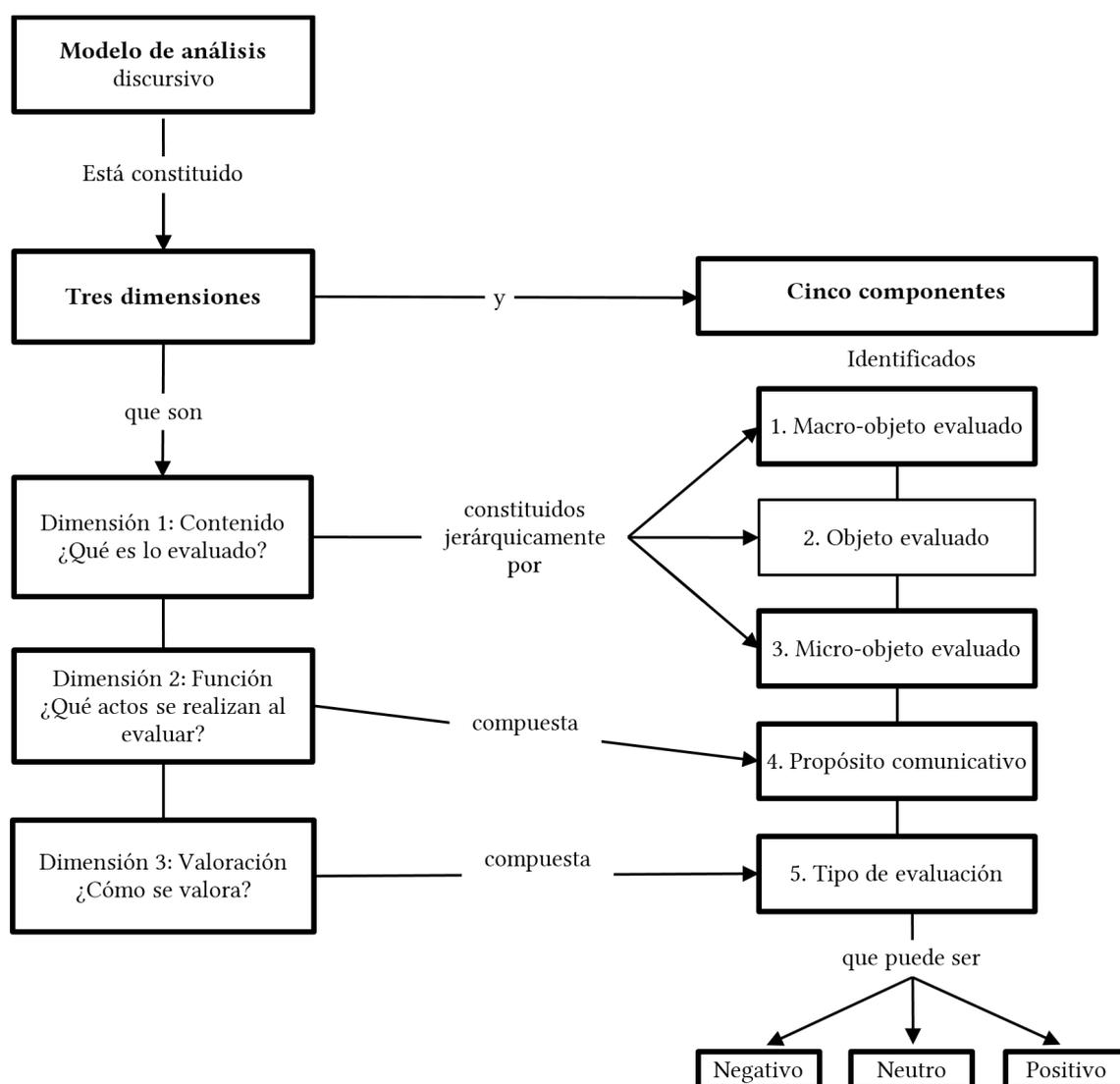


Figura 1: Dimensiones de un modelo para el análisis de los IA

A diferencia de los modelos revisados, en la clasificación que proponemos se integran, a la vez, aspectos relativos al contenido, a los propósitos comunicativos y a los

tipos de valoración. A continuación, describiremos en detalle cada una de las dimensiones y los respectivos componentes de este modelo.

## La dimensión del Contenido

Esta primera dimensión tiene por propósito describir cuáles son los objetos que se evalúan en los IA. A diferencia de otras propuestas que identifican categorías muy generales (Mungra y Weber) u otras muy específicas (Tharirian y Sadri, 2013), para identificar y tipificar los objetos evaluados, utilizamos una clasificación jerárquica, que permite analizarlos en tres niveles de generalidad/especificidad.

El primer nivel o componente más general, denominado *macro objeto evaluado*, da cuenta de la naturaleza general del comentario emitido por los árbitros. A partir de la revisión de otros trabajos (Gosden, 2003; Mungra y Weber, 2010; Tharirian y Sadri, 2013), se estableció que los comentarios pueden referirse a aspectos del contenido o bien a aspectos formales. Sin embargo, en nuestro modelo hemos agregado otra categoría que se utilizó cuando el comentario no se podía vincular a ninguno de los dos aspectos (*Indeterminado*).

El componente del *objeto evaluado*, el segundo en nivel de generalidad de la dimensión del contenido, proporciona información respecto del objeto textual donde el evaluador focaliza su valoración. Puede tratarse del artículo en su totalidad o de una sección específica. Cabe señalar que en ocasiones el *objeto evaluado* puede aparecer de forma implícita. El componente del objeto evaluado contiene 10 categorías.

El último componente de este nivel, el más específico de la dimensión del contenido, denominado *micro-objeto evaluado*, da cuenta de un aspecto delimitado del *objeto evaluado*, en el cual el evaluador focaliza su atención. Mientras que la presencia del *objeto evaluado* es obligatoria, aunque sea de forma implícita, la aparición del *micro-objeto evaluado* es facultativa. Este componente del modelo es el más amplio, pues contiene 34 categorías. A continuación, se muestran un ejemplo de la aplicación de esta dimensión:

(9) El tema del artículo no es original

Macro-Objeto: Contenido

Objeto: el artículo

Micro-Objeto: originalidad del tema

## La dimensión funcional

Como todo texto, los IA tienen una función general, a saber, evaluar el artículo y realizar una recomendación al editor sobre la potencial publicación del trabajo. Este propósito general se concretiza a través de propósitos comunicativos más específicos. A diferencia de otros trabajos (Gosden, 2003; Fortanet, 2008; Bolívar, 2011), que contienen funciones generales, nuestro modelo contiene funciones más específicas, entre las que se encuentran *sugerir una aclaración*, *solicitar una acción*, entre otras. Algunos ejemplos de la aplicación de esta dimensión a los comentarios de los IA:

(10) Aclarar el criterio de selección para los porcentajes de dopaje

Propósito comunicativo: Solicita acción

(11) El tema es muy pertinente

Propósito comunicativo: Destaca aspecto positivo

(12) En la introducción no deben aparecer referencias a los aspectos teóricos

Propósito comunicativo: Indica algo que sobra

## La dimensión de la valoración

Con la tercera y última dimensión del modelo de análisis, la valoración, se puede identificar la polaridad de los comentarios, a saber, negativa o positiva. Cabe señalar que también se consideró en este análisis un tipo de comentario que no es evaluativo, sino que está asociado a la categoría *Describe*. Así también, la valoración puede aparecer de forma indirecta, esto es, cuando no hay marcas explícitas de valoración, sino que estas se derivan del propósito comunicativo. Por ejemplo, en el caso (10), no aparece propiamente una marca de valoración negativa, pero del propósito comunicativo, se puede derivar que, en efecto, se trata de esa polaridad. Algunos ejemplos de esta dimensión son:

(13) El trabajo trata sobre el análisis químico de los residuos industriales.

Valoración: Neutra (se deriva del propósito comunicativo *Describe*)

(14) El tema del trabajo es muy pertinente

Valoración: Positiva

(15) No es un buen trabajo.

Valoración: Negativa

## Aplicación del modelo a un *corpus* de IA

---

En lo que sigue, mostramos la aplicación del modelo descrito a los datos de la muestra. Dos son los objetivos de esta parte empírica. Por un lado, se buscó determinar si los atributos extralingüísticos de los informes, a saber, la recomendación de publicación (aceptado, condicionado, rechazado) implicaban variaciones en la naturaleza del discurso de esos informes. Por otro lado, se exploró si el discurso de esos informes se veía afectado por las propiedades cuantitativas de los evaluadores, específicamente, en relación a su productividad, medida según el número de artículos publicados. Con estos dos objetivos, superamos en parte algunos de los problemas indicados en la sección “2.1 Modelos para el análisis del discurso de los IA en el PEP”.

## Datos

---

Como una forma de superar la falta de integración de los aspectos sociométricos y discursivos en el análisis del proceso de revisión por pares, se recolectó una muestra de 318 IA, provenientes de una población de 738 informes, correspondientes a las evaluaciones del total de artículos enviados durante los años 2008 y 2012 a tres revistas internacionales publicadas en español, *Onomázein*, *Información Tecnológica* y *Formación Universitaria* (Sabaj, Valderrama, González y Pina-Stranger, 2015). Para el desarrollo del primer objetivo se registró la recomendación de publicación del total de la muestra (318 informes).

La conformación de la muestra de IA, aparece en la Tabla 1, separada por cada revista y por la recomendación de publicación de los evaluadores.

| Recomendación    | Onomázein | Formación Universitaria | Información tecnológica |
|------------------|-----------|-------------------------|-------------------------|
| Aceptado (A)     | 14        | --                      | --                      |
| Condicionado (C) | 79        | 76                      | 58                      |
| Rechazado (R)    | 25        | 24                      | 42                      |
| TOTALES          | 118       | 100                     | 100                     |

**Tabla 1:** Conformación de la muestra de análisis

Para la consecución del segundo objetivo, se trabajó con la submuestra de 118 informes de la *Onomázein*. Para cada evaluador se registró su número de publicaciones. Para contar el número de artículos, acudiendo para ello a distintas fuentes (Google Scholar, ResearchGate, Academia, páginas personales e institucionales). Además, la in-

formación se cotejó a través de consultas de las bases de datos de Scielo, Scopus y WOS. Como una forma de cuidar la confidencialidad de los datos, los asistentes encargados de registrar el número de artículos solo tuvieron acceso al listado de evaluadores por parte del editor, sin que fuera posible vincular qué evaluador revisó qué artículo. Para la obtención de los resultados de este segundo objetivo, se agruparon los evaluadores en tres grupos de manera arbitraria, según su productividad, tal como se muestra en la Tabla 2.

| Tipo | Rango o número de artículos | N   |
|------|-----------------------------|-----|
| A    | 0                           | 12  |
| B    | 1-10                        | 41  |
| C    | Más de 10                   | 65  |
|      | Total                       | 118 |

**Tabla 2:** Número de evaluadores por número de artículos publicados

Cabe señalar que hasta donde llega nuestro conocimiento, este es el *corpus* más amplio de IA etiquetados discursivamente y con información centométrica, disponible en español.

### Procedimientos para el análisis del discurso de los IA, obtención de resultados y cálculos estadísticos

El etiquetaje de los datos discursivos fue realizado por los 8 investigadores, quienes, al igual que en el proceso de construcción del modelo discursivo de análisis, trabajaron en parejas. Específicamente, para el etiquetaje se siguieron estos pasos:

1. Identificar el límite del comentario
2. Determinar si se trataba de un comentario ajeno; en caso positivo, el análisis concluía en ese punto.
3. Determinar si el comentario era descriptivo o evaluativo. En el primer caso, se le asignaba el código correspondiente y el análisis concluía en ese punto.
4. En el caso de los comentarios evaluativos, se aplicaron, primero, las tres etiquetas correspondientes a la dimensión del contenido en sus tres niveles de especificidad (macro, objeto y micro objeto), luego se etiquetaba el propósito comunicativo y luego se asignaba la polaridad del comentario, esto es, si era positivo o negativo.

Los comentarios ajenos y los descriptivos fueron etiquetados con dos categorías, mientras que los comentarios evaluativos se les asignó seis etiquetas.

Para la obtención de los resultados se contaron las frecuencias para cada una de las categorías del modelo. En lo que sigue, se muestran los resultados relativos a tres de los componentes del modelo, a saber, *el macro-objeto*, *el objeto evaluado* y los *propósitos comunicativos*. Esta selección se basa en que la aplicación del modelo a los otros

componentes ya ha sido publicada en otros trabajos: el análisis de los *micro-objetos* puede encontrarse en César Astudillo (2015), mientras que la descripción del componente de la valoración se trató en Astudillo et al. (2016). Los resultados de la aplicación del modelo completo (las 3 dimensiones y los 5 componentes) sobre la muestra que acá se analiza están disponibles para los interesados por medio de una solicitud a los autores.

Para cada componente discursivo se aplicaron cálculos estadísticos para determinar si existían diferencias en las proporciones del uso de una categoría según la recomendación de publicación de los evaluadores (ver Tablas 3, 4 y 5) y del número de artículos publicados (ver Tablas 6, 7 y 8). Dado que algunas categorías tenían una frecuencia menor a cinco no fue posible usar Chi cuadrado, por lo que se utilizó el Test de K proporciones (Alfa= 5%), que permite saber si hay diferencias significativas en las proporciones en que ocurre una variable  $n$ . Para identificar entre qué par de grupos había diferencias significativas entre las proporciones se utilizaron dos métodos: el Método Monte Carlo con 5000 iteraciones (Evans y Rosenthal, 2005) y el procedimiento de Marascuilo (Berenson, Levine y Krehbiel, 2001).

Para efectos de la presentación de los resultados, utilizamos porcentajes. Para el análisis de los propósitos comunicativos, se eliminaron de las tablas de porcentajes aquellas categorías que en todos los grupos no superaran el uno por ciento, por lo que, en esos casos, la suma total de la columna, en rigor, no alcanza el 100 por ciento. Para efectos de la discusión de los resultados, agregamos a todas las tablas una columna con el porcentaje total de ocurrencia de cada categoría, independiente de las variables desagregadas. Esto, con el objetivo de focalizar esa discusión a las categorías más frecuentes y a los casos donde existen diferencias significativas.

Los datos fuente de la investigación están disponibles en el Anexo 2. Si bien, como ya hemos expuesto, nuestro *corpus* está compuesto por 318 IA y una submuestra de 118 de ellos, los datos de la investigación están constituidos por más de 8000 (en el primer caso y de 3000 para la submuestra) fragmentos textuales de esos informes, etiquetados con las categorías del modelo.

## El informe de arbitraje según la recomendación de publicación de los evaluadores

La Tabla 3 muestra el porcentaje de aparición de los comentarios según su naturaleza, agrupados según la recomendación de publicación.

| Macro-Objeto  | A          | C          | R          | Total      | Test de K proporciones (alfa =5 %) |
|---------------|------------|------------|------------|------------|------------------------------------|
| Contenido     | 75         | 58         | 64         | 61         | Hay diferencias en todas           |
| Forma         | 22         | 32         | 26         | 29         | A y C; C y R                       |
| Indeterminado | 3          | 10         | 10         | 10         | A y C; A y R                       |
| <b>Total</b>  | <b>100</b> | <b>100</b> | <b>100</b> | <b>100</b> |                                    |

**Tabla 3:** Porcentaje de macro-objetos evaluados según la recomendación de publicación de los evaluadores  
**Leyenda:** A= Aceptados; C= Condicionados; R= Rechazados

Según estos datos, independientemente de la recomendación de publicación, en los IA priman los comentarios que apuntan al contenido de los artículos, sin embargo, la proporción de estos en relación con los otros tipos, difiere de acuerdo con la recomendación y, de hecho, existen diferencias significativas en todas las proporciones de los grupos. La predominancia de los comentarios de contenido ya había sido constatada por Tharirian y Sadri (2013) y por Mungra y Weber (2010) para distintas disciplinas, y también por Astudillo (2015), en este último caso de manera específica para Informes que recomendaban rechazar la publicación de los manuscritos. En los datos que presentamos, este patrón también se evidencia en los otros tipos de recomendación (condicionado y aceptados).

Es interesante destacar que los aspectos formales aparecen en todos los tipos de recomendación, siendo más relevantes en el caso de los artículos condicionados. En los comentarios sobre la forma, se pueden observar diferencias significativas, por una parte, entre los artículos aceptados y los condicionados, y entre estos y los rechazados, sin diferencias entre los aceptados y los rechazados. Estos resultados son similares a los encontrados por Betty Samraj (2016), quien estableció que la recomendación de los evaluadores supone diferencias discursivas en los informes de arbitraje. Esto fortalece una interpretación basada en el esfuerzo que hacen los evaluadores según sea la recomendación que realizan, lo que se plasma, en última instancia, en la naturaleza discursiva de los comentarios de los IA. Así, si los evaluadores van a aceptar o rechazar el trabajo, abordarán en menor medida la forma, porque hacerlo implica un mayor esfuerzo que no está justificado por la decisión que recomiendan.

Los comentarios indeterminados son muy poco frecuentes en los artículos aceptados y tienen una frecuencia relativa igual en los artículos condicionados y rechazados. En estos casos, existen diferencias significativas entre los artículos aceptados y los condicionados, por una parte, y entre los aceptados y los rechazados, por otra.

En la Tabla 4, se muestran los porcentajes de comentarios asociados a los objetos evaluados según la recomendación de publicación:

| Objeto evaluado      | A          | C          | R          | Total      | Test de K proporciones (alfa =5 %) |
|----------------------|------------|------------|------------|------------|------------------------------------|
| <b>Título</b>        | 3          | 2          | 1          | 2          | <b>C y R</b>                       |
| <b>Resumen</b>       | 6          | 4          | 5          | 4          | No hay diferencias                 |
| <b>Introducción</b>  | 1          | 4          | 3          | 4          | <b>A y C; C y R</b>                |
| <b>Marco teórico</b> | 1          | 2          | 1          | 2          | <b>C y R</b>                       |
| <b>Metodología</b>   | 1          | 8          | 15         | 11         | <b>Hay diferencias en todas</b>    |
| <b>Resultados</b>    | 4          | 8          | 8          | 8          | No hay diferencias                 |
| <b>Conclusión</b>    | 5          | 3          | 3          | 3          | No hay diferencias                 |
| <b>Referencias</b>   | 10         | 9          | 7          | 9          | <b>C y R</b>                       |
| <b>Artículo</b>      | 68         | 59         | 56         | 58         | <b>A y R; C y R</b>                |
| <b>Total</b>         | <b>100</b> | <b>100</b> | <b>100</b> | <b>100</b> |                                    |

**Tabla 4.** Porcentaje de objetos evaluados según la recomendación de publicación de los evaluadores  
**Leyenda:** A= Aceptados; C= Condicionados; R= Rechazados

Para observar cuáles son los objetos evaluados más prominentes, nos centramos en la columna donde aparece el total. Ahí, podemos observar que, al evaluar un artículo, los árbitros realizan una valoración holística, lo que se manifiesta en que el objeto evaluado predominante (58%) sea el artículo en su totalidad, esto de manera transversal a los distintos tipos de recomendación. Cuando los artículos son aceptados este porcentaje es mayor (68%) que en los otros tipos de recomendación. Esto puede deberse a que, cuando el artículo es aceptado, no es necesario entrar en el detalle de sus partes. Cuando los comentarios tienen el artículo como objeto, existen diferencias significativas entre los artículos aceptados y rechazados, y entre estos y los condicionados.

La recomendación de publicación afecta de manera específica el discurso de los comentarios en los IA, cuando ellos tienen como objeto a la metodología. En el caso de los artículos aceptados, son muy escasos los comentarios sobre la metodología y, en cambio, son muy frecuentes en el caso de los artículos, cuya publicación ha sido rechazada. En este sentido, los aspectos metodológicos tienen gran incidencia en el destino de un artículo enviado a una revista. Estos datos concuerdan con lo explorado por Astudillo (2015), quien determinó que una de las secciones más evaluadas en los informes de arbitraje con la recomendación de rechazar los manuscritos es, justamente, la metodología.

Otro aspecto común a los distintos tipos de recomendaciones de publicación son los comentarios relativos a las referencias. Esto sugiere que la revisión bibliográfica, es uno de los aspectos centrales del género ‘informe de arbitraje’, es decir, es un tema en que los evaluadores focalizarán su atención sin importar la recomendación que entregarán al editor (estos datos pueden ser explicados sin embargo por otras variables que no hemos controlado, ver sección Conclusiones). En estos casos, las proporciones son

estadísticamente distintas entre los artículos condicionados y los rechazados, siendo su frecuencia relativa la menor para estos últimos. De estos datos se desprende que los artículos tienden a rechazarse más por motivos metodológicos que por cuestiones relativas a las referencias.

En los datos podemos observar que, por una parte, existen una serie de objetos evaluados con una frecuencia relativa mediana o baja que no implican diferencias en los comentarios de los informes según sea la recomendación (resumen, resultados conclusiones). Por otra parte, existen ciertos objetos evaluados con una frecuencia relativa baja, que sin embargo sí presentan diferencias significativas en las proporciones entre grupos particulares, como los comentarios relativos al título y al marco teórico que muestran diferencias significativas entre los artículos condicionados y rechazados, o los comentarios relativos a la introducción que exhiben diferencias significativas entre los artículos aceptados y condicionados, y entre estos y los rechazados.

Una distribución homogénea de las categorías de la Tabla 4 puede reflejar un mayor nivel de detalle en la evaluación. Según esto, cuando los artículos se condicionan, los evaluadores distribuyen de forma más pareja sus comentarios en las distintas opciones, lo que implicaría una revisión más detallada del artículo. Estos datos reflejan la lógica general de la revisión por pares, en tanto los evaluadores no están disponibles para invertir tiempo en un artículo que consideran bueno (aceptado) ni en uno que consideran malo (rechazado), pero sí en un trabajo valioso que necesita ser modificado.

En la Tabla 5, se exponen los porcentajes de aparición de los propósitos comunicativos en los IA, agrupados según la recomendación de publicación.

Si observamos la columna del total de la Tabla 5, podemos observar cuáles son los propósitos comunicativos más frecuentes en los IA. De esos datos podemos deducir que existen unos pocos propósitos que constituyen el núcleo de este género, a saber, Describir un error, Describir, Solicitar acción, Corregir y Describir lo que falta.

Es interesante dar cuenta de la forma en que un atributo extralingüístico del proceso de revisión por pares, como la recomendación de publicación en un informe de arbitraje, determina aspectos específicos del uso del lenguaje, como en este caso, ciertos propósitos comunicativos.

La incidencia de la recomendación de publicación en el discurso de los IA se manifiesta de dos formas en estos datos. En primer lugar, se expresa en la mayor ocurrencia de ciertos propósitos para algunas de las recomendaciones. Por ejemplo, el propósito "Destaca aspecto positivo", esperablemente, es característico de los informes de los

| Propósito comunicativo              | A          | C          | R          | Total      | Test de K proporciones (alfa =5 %) |
|-------------------------------------|------------|------------|------------|------------|------------------------------------|
| Condiciona publicación              | 1          | 3          | 1          | 2          | A y C; C y R                       |
| Corrige                             | 10         | 10         | 7          | 9          | C y R                              |
| Describe                            | 3          | 12         | 15         | 13         | Hay diferencias en todas           |
| Describe error                      | 7          | 19         | 32         | 24         | Hay diferencias en todas           |
| Describe lo que falta               | 2          | 7          | 13         | 9          | Hay diferencias en todas           |
| Destaca aspecto positivo            | 44         | 11         | 4          | 8          | Hay diferencias en todas           |
| Evalúa aporte                       | 9          | 2          | 3          | 3          | A y C; A y R                       |
| Rechaza la publicación              | 0          | 0          | 2          | 1          | A y R; C y R                       |
| Solicita acción                     | 5          | 15         | 8          | 12         | A y C; C y R                       |
| Sugiere acción                      | 10         | 7          | 4          | 6          | C y R                              |
| Sugiere acción mediante pregunta    | 0          | 4          | 4          | 4          | No hay diferencias                 |
| Sugiere inclusión                   | 6          | 6          | 4          | 5          | A y C; A y R                       |
| Sugiere inclusión mediante pregunta | 0          | 2          | 1          | 1          | A y C; A y R                       |
| <b>Total</b>                        | <b>100</b> | <b>100</b> | <b>100</b> | <b>100</b> |                                    |

**Tabla 5:** Porcentaje de propósitos comunicativos según la recomendación de publicación de los evaluadores

**Legenda:** A= Aceptados; C= Condicionados; R= Rechazados

artículos aceptados. Cuando los evaluadores recomiendan la publicación del trabajo, realizan más sugerencias, en cambio, si condicionan la publicación del trabajo, entonces directamente solicitan acciones. Es interesante comparar estos resultados con los obtenidos por Brian Paltridge (2015). El autor señala que, si bien la solicitud de acciones es menos cortés que una sugerencia, estas últimas confunden a los autores, sobre todo cuando los evaluadores recomiendan rechazar o aceptar la publicación de los manuscritos.

En segundo lugar, la manera en que la recomendación afecta el uso del lenguaje en este género se evidencia en que, de las trece categorías de la Tabla 5 (recuérdese que se eliminaron de esa tabla aquellas categorías con menos de un 1% de ocurrencia), en cuatro existen diferencias entre todos los grupos, en 8 se presentan diferencias entre algunos de los grupos y, solo, un propósito comunicativo no se ve afectado por esa recomendación.

Profundizando en la discusión, existe un primer grupo de cuatro propósitos de alta frecuencia que además exhiben diferencias en el uso en todos los tipos de recomendación. Tres de estos propósitos además constituyen parte de lo que hemos denominado núcleo del género y su frecuencia relativa siempre es mayor en los informes de artículos que fueron rechazados: Describe, Describe error y Describe lo que falta. Estos propósitos comunicativos coinciden de manera directa con aquellos identificados por Bolívar (2011). Así, los distintos tipos de recomendación no implican el uso de pro-

pósitos comunicativos del todo distintos, sino que supone el uso de algunos en distinta proporción.

## El informe de arbitraje según la productividad de los evaluadores

En la Tabla 6, se muestran los porcentajes de ocurrencia de los macro-objetos evaluados, agrupados según el número de artículos de los evaluadores.

| Macro-Objeto  | A   | B   | C   | Total | Test de K proporciones<br>(alfa =5 %) |
|---------------|-----|-----|-----|-------|---------------------------------------|
| Contenido     | 64  | 46  | 66  | 59    | A y B; B y C                          |
| Forma         | 20  | 41  | 24  | 29    | A y B; B y C                          |
| Indeterminado | 16  | 13  | 10  | 12    | A y C                                 |
| <b>Total</b>  | 100 | 100 | 100 | 100   |                                       |

**Tabla 6:** Porcentaje de macro-objeto evaluados según el número de artículos de los evaluadores  
**Legenda:** A= 0 artículos; B= 1-10 artículos; C= Más de 10 artículos

Al enfrentarse a la necesidad de producir un informe de arbitraje, los evaluadores realizan de manera consciente o inconsciente una serie de elecciones discursivas. Esas opciones pueden estar, en parte, determinadas por algunos atributos extralingüísticos de los evaluadores, como, en este caso, el número de artículos publicados. Tal como se observa en la Tabla 6, todos los comentarios apuntan mayoritariamente a aspectos de contenido, independientemente de la productividad de los evaluadores. Sin embargo, la proporción en que los distintos tipos de evaluadores hacen comentarios de contenido permite diferenciar a los evaluadores sin publicaciones (A), de los evaluadores con un máximo de diez artículos (B), y a estos de los evaluados con mucha productividad (C).

Es interesante notar que, en relación con los comentarios de contenido y de forma, los evaluadores sin publicaciones (A) se comportan de manera muy similar a los evaluadores consagrados (C). La diferencia más evidente en este caso está dada por los evaluadores en vías de consolidación (B), que si bien, al igual que el resto, realizan más comentarios de contenido que de forma, son los que más atención prestan a los aspectos formales (41%). No existen otros trabajos que permitan explicar este comportamiento de manera específica. De manera general, y en específico para los manuscritos que han sido rechazados, se ha sostenido que los evaluadores utilizan razones de fondo y no de forma al evaluar los artículos de investigación (Astudillo, 2015).

Si comparamos los datos de la Tabla 6 con aquellos de la Tabla 3, podemos establecer que la predominancia de comentarios de contenido es transversal a los IA, de

manera independiente de la recomendación y del número de publicaciones de los evaluadores.

En la Tabla 7, se muestran los resultados de la aparición de los objetos evaluados.

| Objeto evaluado             | A   | B   | C   | Total | Test de K proporciones<br>(alfa =5 %) |
|-----------------------------|-----|-----|-----|-------|---------------------------------------|
| <b>Título</b>               | 1   | 2   | 2   | 2     | No hay diferencias                    |
| <b>Resumen</b>              | 4   | 5   | 4   | 4     | No hay diferencias                    |
| <b>Introducción</b>         | 6   | 3   | 2   | 3     | <b>B y C</b>                          |
| <b>Marco teórico</b>        | 3   | 3   | 3   | 3     | No hay diferencias                    |
| <b>Metodología</b>          | 16  | 6   | 12  | 10    | <b>A y B; B y C</b>                   |
| <b>Resultados</b>           | 12  | 4   | 6   | 6     | <b>Hay diferencias en todas</b>       |
| <b>Conclusión</b>           | 4   | 3   | 2   | 2     | No hay diferencias                    |
| <b>Referencias</b>          | 12  | 7   | 9   | 9     | No hay diferencias                    |
| <b>Artículo</b>             | 41  | 66  | 59  | 60    | <b>Hay diferencias en todas</b>       |
| <b>Organización interna</b> | 2   | 2   | 1   | 2     | No hay diferencias                    |
| <b>Total</b>                | 100 | 100 | 100 | 100   |                                       |

**Tabla 7.** Porcentaje de objetos evaluados según la productividad de los evaluadores  
**Legenda:** A= 0 artículos; B= 1-10 artículos; C= +10 artículos

El análisis de los objetos evaluados en los IA según el número de artículos de los evaluadores devela un patrón similar a los datos observados según el tipo de recomendación de publicación (Tabla 4): existen ciertos objetos transversales que son comunes a los tres tipos de evaluadores (el artículo en su totalidad, la metodología, los resultados y las referencias).

Es interesante notar que, si bien la categoría que hace referencia al artículo como un todo es la más frecuente en los tres grupos, su ocurrencia relativa es menor en el grupo de los evaluadores sin publicaciones (A), lo que sugiere que estos realizan una revisión más detallada y analítica que los otros tipos de evaluadores (B y C).

Así también, los evaluadores sin productividad científica son los que mayor atención les otorgan a los objetos evaluados que son comunes a los tres grupos, lo que se evidencia en la mayor proporción de comentarios asociados a la metodología (16%), a los resultados (12%) y a las referencias (12%). Estos resultados sugieren que los evaluadores neófitos son más exigentes que sus pares más productivos al momento de realizar una revisión. En parte, estos datos son consistentes con los resultados de las investigaciones de Germán Varas (2015) y de Thomas Stossel (1985). Estos autores señalan que existe una relación inversa entre el estatus de los evaluadores y la calidad de la re-

troalimentación presente en los informes de arbitraje. Los datos de la Tabla 7 sugieren, además, que el evaluador sin productividad realiza una revisión más minuciosa por la necesidad de ir legitimándose en campo disciplinar, en este caso, mostrando ante el editor que es capaz de realizar una evaluación detallada y exhaustiva.

Los datos de la última columna de la Tabla 7 dan cuenta de que los atributos extralingüísticos de los evaluadores implican diferencias significativas en los objetos evaluados de los comentarios, específicamente, cuando ellos hacen referencia a los resultados y al artículo en su totalidad. Cuando los comentarios tienen como objeto a la introducción o a la metodología, las diferencias entre los tipos de evaluadores son parciales, esto es, se dan entre grupos específicos. El resto de los objetos evaluados en los comentarios de IA no implica variaciones significativas según la productividad de los evaluadores.

Si comparamos los datos de la Tabla 7 con aquellos de la Tabla 4, podemos constatar cómo el foco del análisis da cuenta del comportamiento de esos datos y la manera en que ellos se complementan. Así, por ejemplo, hay objetos evaluados que no son sensibles ni a la recomendación de publicación por parte de los evaluadores ni a su productividad en términos del número de artículos publicados, como es el caso de aquellos comentarios que tienen por objeto el resumen o la conclusión. Por otra parte, hay categorías que exhiben diferencias (totales o parciales) en un criterio, pero no en el otro o viceversa, como, por ejemplo, los comentarios referidos a los resultados que no son afectados por la productividad de los evaluadores, pero sí por la recomendación de publicación. Por último, el uso de comentarios a ciertos objetos específicos implica diferencias (totales o parciales) en las dos formas de agrupamiento de los datos (recomendación de publicación y productividad de los evaluadores), como sucede con los comentarios que tienen como asunto la introducción, la metodología o el artículo en su totalidad.

En la Tabla 8, se muestra el porcentaje de aparición de los propósitos comunicativos según la productividad de los evaluadores.

Tal como se observa en la Tabla 8, el grupo de propósitos comunicativos nucleares tienden a ser los mismos que si analizamos los informes según la recomendación de evaluación (ver Tabla 5). De particular interés resulta el uso de algunos propósitos comunicativos que muestran una tendencia específica en relación con la productividad. Por ejemplo, los propósitos “Describir error” y “Sugerir acción” que aumentan con la productividad de los evaluadores o, el propósito de “Solicitar acción”, el que, por el contrario, es mayor en los evaluadores con menos artículos publicados. Nuevamente, los datos muestran, en general, que los evaluadores sin publicaciones realizan eva-

luaciones más directas y exhaustivas, mientras que los más productivos son más indirectos (realizan más sugerencias que directivos) y menos detallistas. Estos resultados podrían explicarse a la luz de la evidencia aportada por Germán Varas (2015) y por Thomas Stossel (1985), en el sentido que los tipos de actos que se realizan en los informes de arbitraje dependen en gran medida del estatus de los evaluadores, lo cual afectaría en la calidad de la retroalimentación.

| Propósitos comunicativos         | A   | B   | C   | Total | Test de K proporciones<br>(alfa =5 %) |
|----------------------------------|-----|-----|-----|-------|---------------------------------------|
| Condiciona publicación           | 2   | 2   | 1   | 2     | No hay diferencias                    |
| Corrige                          | 9   | 11  | 8   | 9     | No hay diferencias                    |
| Describe                         | 13  | 8   | 10  | 9     | <b>A y B</b>                          |
| Describe error                   | 23  | 30  | 31  | 30    | <b>A y C</b>                          |
| Describe lo que falta            | 10  | 7   | 7   | 7     | No hay diferencias                    |
| Destaca aspecto positivo         | 2   | 9   | 8   | 8     | <b>A y B; A y C</b>                   |
| Evalúa aporte                    | 3   | 3   | 2   | 2     | No hay diferencias                    |
| Indica lo que sobra              | 2   | 1   | 1   | 1     | No hay diferencias                    |
| Solicita acción                  | 17  | 12  | 10  | 11    | <b>A y C</b>                          |
| Sugiere acción                   | 3   | 7   | 8   | 7     | <b>A y B; A y C</b>                   |
| Sugiere acción mediante pregunta | 5   | 2   | 4   | 3     | <b>B y C</b>                          |
| Sugiere inclusión                | 7   | 7   | 8   | 8     | No hay diferencias                    |
| <b>Total</b>                     | 100 | 100 | 100 | 100   |                                       |

**Tabla 8.** Proporción de propósitos comunicativos según el número de artículos de los evaluadores  
**Leyenda:** A= 0 artículos; B= 1-10 artículos; C= +10 artículos

Según los datos de la Tabla 8, la productividad de los evaluadores no afecta de manera directa el uso de los propósitos comunicativos en los IA, en tanto, en ninguna de las categorías se generan diferencias significativas entre los tres tipos de evaluadores. A pesar de ello, podemos constatar que el uso de algunos propósitos comunicativos sí se ve afectado por las diferencias de la condición extralingüística de los evaluadores, a saber, el número de sus publicaciones. En la mayoría de los casos, estas diferencias se manifiestan entre evaluadores sin publicaciones (A) y aquellos que muestran una alta productividad (C) y, además, tienden a aparecer en el uso de propósitos comunicativos de alta frecuencia relativa en los IA, como “Describe error”, “Solicita acción” y “Destaca aspecto positivo”.

Si comparamos los resultados de la Tabla 8 con aquellos de la Tabla 5, podemos establecer que el uso de los propósitos comunicativos en los IA de artículos de investigación es mucho más sensible a la recomendación de publicación que a la productividad de los evaluadores. En efecto, solo una de las categorías analizadas según la reco-

mendación no se vio afectada por ella (ver Tabla 5), mientras que en el caso de los propósitos comunicativos son varios los propósitos comunicativos que no son sensibles a la productividad de los evaluadores.

## Conclusiones

---

A partir de cuatro críticas a los trabajos de análisis del discurso que tienen como objeto de investigación los IA, en este artículo hemos propuesto un modelo de análisis para la descripción de este género con el cual se intentó superar, en parte, esas críticas. Además, hemos aplicado tres de las dimensiones o capas de este modelo (la naturaleza del comentario, el objeto evaluado y los propósitos comunicativos) para describir un *corpus* de IA. Como una forma de integrar aspectos discursivos con información socio-relacional, agrupamos el *corpus* según dos atributos extralingüísticos, a saber, la recomendación de publicación y la productividad de los evaluadores. Del análisis de los datos, se pudo concluir que la recomendación de publicación es la que presenta una mayor variación discursiva del género IA.

Los datos de esta investigación pueden ser de utilidad para todos los actores involucrados en la actividad científica, pero especialmente, para los editores, quienes podrían seleccionar evaluadores según las características de los trabajos y de los propios atributos de los investigadores. Así, por ejemplo, si un editor considera que un manuscrito es bueno, pero la sección de la metodología necesita especial atención, podría seleccionar a un evaluador con poca productividad, los cuales, según los datos de este trabajo son los que más se focalizan en dicha sección.

La debilidad más evidente de este trabajo radica en no haber considerado una forma de aislar de los resultados una variable que puede incidir fuertemente en ellos o su interpretación. Básicamente, nos referimos a la manera en que el discurso de los IA se ve afectado por los formatos de evaluación utilizados por las revistas. Sabemos que dichos formatos orientan la evaluación que realizan los árbitros, lo que en definitiva puede explicar por qué ciertos objetos son más prominentes que otros y por qué los comentarios tienden más a referirse al contenido que a aspectos formales. Esta debilidad no es, sin embargo, aplicable de la misma manera al caso de los propósitos comunicativos, ya que su uso no es orientado de modo tan explícito como puede suceder para las otras categorías (la naturaleza del comentario y los objetos evaluados). Sin embargo, el hecho de que los formatos contengan secciones para realizar comentarios libres matiza en parte esta debilidad.

A pesar de esta debilidad, creemos que la propuesta de modelo de análisis y la metodología utilizada tienen el valor de superar, al menos en parte, algunas de las críticas

que hemos esbozado con relación al ámbito general del análisis del discurso. En síntesis, se espera que los resultados de esta investigación contribuyan a una comprensión más profunda de forma en que se construye colectivamente el conocimiento científico.

## Agradecimientos

---

Esta investigación fue financiada gracias al Proyecto Fondecyt 1130290: “La interacción socio-discursiva en la construcción colectiva del conocimiento científico: la dinámica interna del proceso de evaluación por pares”.

## Anexos

---

### Anexo 1

Un modelo discursivo para el análisis de los IA en el proceso de evaluación por pares:

<http://atheneadigital.net/article/view/2051/2051-a1-pdf-es>

### Anexo 2

Datos fuente para los cálculos estadísticos:

<http://atheneadigital.net/article/view/2051/2051-a2-pdf-es>

## Referencias

---

- Ardakan, Mohammad Abooyee; Mirzaie, Seyyed Ayatollah & Sheikhshoaei, Fatemeh (2011). The Peer-Review Process for Articles in Iran's Scientific Journals. *Journal of Scholarly Publishing*, 42(2), 243-261.  
<https://doi.org/10.3138/jsp.42.2.243>
- Astudillo, César (2015). *Aplicación de un modelo discursivo para el análisis de los Informes de Evaluación de Artículos Rechazados (IEAR) en el Proceso de Evaluación por Pares (PEP) de tres revistas chilenas*. Tesis de Magíster sin publicar, Universidad de La Serena.
- Astudillo, César; Squadrito, Kareem; Varas, Germán; González, Carlos & Sabaj, Omar (2016). Polaridad de los comentarios y consistencia interna en los informes de arbitraje de artículos de investigación. *Acta Bioethica*, 22(1), 119-128.  
<https://doi.org/10.4067/S1726-569X2016000100013>
- Berenson, Mark; Levine, David & Krehbiel, Timothy (2001). *Estadística para administración*. Pearson Education.

- Bolívar, Adriana (2008). El informe de arbitraje como género discursivo en la dinámica de la investigación. *Revista Latinoamericana de Estudios del Discurso*, 8(1), 41-64.
- Bolívar, Adriana (2011). Funciones discursivas de la evaluación negativa en informes de arbitraje de artículos de investigación en educación. *Núcleo* 28, 59-89.
- Bornmann, Lutz (2011). Scientific Peer Review. *Annual Review of Information Science and Technology*, 45(1), 197-245. <https://doi.org/10.1002/aris.2011.1440450112>
- Campanario, Juan Miguel (1998a). Peer review for journals as it stands today. Part 1. *Science Communication*, 19(3), 181-211. <https://doi.org/10.1177/1075547098019003002>
- Campanario, Juan Miguel (1998b). Peer review for journals as it stands today. Part 2. *Science Communication*, 19(4), 277-306. <https://doi.org/10.1177/1075547098019004002>
- Evans, Michael & Rosenthal, Jeffrey (2005). *Probabilidad y estadística*. Barcelona: Reverté.
- Fleiss, John (1971). Measuring Nominal Scale Agreement among many Rater. *Psychological Bulletin*, 76(5), 378-382. <http://dx.doi.org/10.1037/h0031619>
- Fortanet, Inmaculada (2008). Evaluative language in peer review referee reports. *Journal of English for academic purposes*, 7(1), 27-37. <https://doi.org/10.1016/j.jeap.2008.02.004>
- Garfield, Eugene (1986). Refereeing and Peer Review, Part 2: The Research on Refereeing and Alternatives to the Present System. *Essays of an Information Scientist*, 9(32), 3-12.
- Garfield, Eugene (1987). Refereeing and Peer Review, Part 4: Research on the Peer Review of Grant Proposals and Suggestions for Improvement. *Essays of an Information Scientist*, 10(5), 27-33.
- Gosden, Hugh (2003). 'Why not give us the full story?': functions of referees' comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes*, 2, 87-101. [https://doi.org/10.1016/S1475-1585\(02\)00037-1](https://doi.org/10.1016/S1475-1585(02)00037-1)
- Lamont, Michèle (2009). *How professors think: inside the curious world of academic judgment*. Cambridge, Massachusetts / London, England: Harvard University press.
- Latour, Bruno (1992). *Ciencia en acción: Cómo seguir a los científicos e ingenieros a través de la sociedad*. Barcelona: Editorial Labor.
- Lazega, Emmanuel (1992). *Micropolitics of Knowledge: Communication and indirect control in workgroups*. New York: Aldine-de-Gruyter.
- Mungra, Philippa & Webber, Pauline (2010). Peer Review Process in medical research publication: Language and content comment. *Journal of English for Specific Purposes*, 29, 43-53. <https://doi.org/10.1016/j.esp.2009.07.002>
- Paltridge, Brian (1996). Genre analysis and the identification of textual boundaries. *Applied Linguistics*, 15(3), 288-299. <https://doi.org/10.1093/applin/15.3.288>

- Paltridge, Brian (2015). Referees' comments on submissions to peer-reviewed journals: when is a suggestion not a suggestion? *Studies in Higher Education*, 40(1), 106-122. <https://doi.org/10.1080/03075079.2013.818641>
- Sabaj, Omar; Valderrama, José Omar; González, Carlos & Pina-Stranger, Álvaro (2015). Relationship between the Duration of Peer- Review, Publication Decision, and Agreement among Reviewers in three Chilean Journals. *European Science Editing*, 41(4), 87-90.
- Sabaj, Omar; González, Carlos & Pina-Stranger, Álvaro (2016). What we Still don't Know about Peer Review. *Journal of Scholarly Publishing*, 47(2), 180-212. <https://doi.org/10.3138/jsp.47.2.180>
- Samraj, Betty (2016). Discourse structure and variation in manuscript review. Implications for genre categorization. *English for Specific Purposes*, 42, 76-88. <http://dx.doi.org/10.1016/j.esp.2015.12.003>
- Strauss, Anselm & Corbin, Juliet (1997). *Grounded theory in practice*. Londres: Sage.
- Stossel, Thomas (1985). Reviewer status and review quality: experience of the Journal of Clinical Investigation. *The New England Journal of Medicine*, 312(10), 658-659. <http://dx.doi.org/10.1056/NEJM198503073121024>
- Swales, John (1996). Occluded genres in the academy: The case of the submission letter. In Eija Ventola & Anna Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp. 45-58). Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.41.06swa>
- Swales, John (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524827>
- Tharirian, Mohammad Hassan & Sadri, Elham (2013). Peer Reviewers' Comments on Research Articles Submitted by Iranian Researchers. *The Journal of Teaching Language Skills* 5(3), 107-123.
- Varas, Germán (2015). El informe de arbitraje en el proceso de revisión por pares de artículos de investigación: Niveles de retroalimentación según el tipo de evaluador (Tesis de Magíster). Universidad de La Serena. La Serena, Chile.



Este texto está protegido por una licencia [Creative Commons 4.0](https://creativecommons.org/licenses/by/4.0/).

Usted es libre para Compartir —copiar y redistribuir el material en cualquier medio o formato— y Adaptar el documento —remezclar, transformar y crear a partir del material— para cualquier propósito, incluso comercialmente, siempre que cumpla la condición de:

**Atribución:** Usted debe reconocer el crédito de una obra de manera adecuada, proporcionar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que tiene el apoyo del licenciente o lo recibe por el uso que hace.

[Resumen de licencia](#) - [Texto completo de la licencia](#)