

# Análisis de errores en el uso del artículo en textos de aprendientes japoneses de español L2. Estudio de corpus

*Cuadernos CANELA*, 33, pp. 71-89

Recibido: 30-IX-2021

Aceptado: 21-XII-2021

Publicado, versión impresa: 10-V-2022

ISSN 1344-9109

Publicado, versión electrónica: 10-V-2022

ISSN 2189-9568

© Las autoras 2022

canela.org.es

## Kazuko Yonekawa

Universidad Nacional de Educación a Distancia (UNED), Madrid, España

Programa de doctorado de Filología: Estudios Lingüísticos y Literarios, Teoría y Aplicaciones

## Alicia San-Mateo-Valdehíta

Universidad Nacional de Educación a Distancia (UNED), Madrid, España

### Resumen

El objetivo de esta investigación es identificar los usos correctos y erróneos del artículo en la producción escrita de 49 aprendientes japoneses de español —subcorpus de CEDEL2, de 5763 palabras— con el fin de analizar los errores más frecuentes según el nivel de dominio de la L2 y comparar los resultados con los de otros estudios realizados también con hablantes de L1 carentes de artículo. El análisis se basa en Corder (1971) y se ha evaluado la corrección/incorrección de cada caso con un criterio descriptivo: error por omisión, adición o selección del artículo (Alexopoulou, 2006). Se han detectado 1109 casos: 925 correctos y 184 erróneos (16,59 %). El porcentaje de error disminuye con el avance del nivel de dominio: en A, 28,33 %; B, 21,65 % y C, 9,06 %, y la diferencia entre los niveles es estadísticamente significativa. Los errores de selección son los más numerosos (64,13 %), seguidos de los de omisión (33,15 %) y adición (2,72 %). La selección errónea del artículo determinado en la 1.ª mención del sustantivo destaca en todos los niveles. Los resultados corroboran la dificultad que presenta en español la identificación del referente del sustantivo para los aprendientes cuya L1 carece de artículo, como el japonés, el chino, el esloveno, el serbio y el turco.

### Palabras clave

Análisis de errores, corpus de aprendientes, artículo español, aprendientes japoneses de EL2

### Introducción

El artículo se considera uno de los temas más problemáticos para los aprendientes de español como segunda lengua (EL2)<sup>1</sup> cuya L1 carece de esta categoría gramatical, tal y como lo confirman las investigaciones sobre Análisis de errores (AE), en general, y sobre el artículo, en particular, realizadas con hablantes de chino (Lin, 2003; Jie, 2012; Coto, 2015; Zhang, 2016), esloveno (Santiago, 2009), croata (Kočman, 2011), serbio (Rubinjoni, 2017), turco (García y Seoane, 2015; Nergiz, 2017), polaco (Zimny, 2017), y de japonés (Saito, 2005; Hamamatsu, 2008; Nakagawa, 2012; Kobashi, 2016). El estudio de Fernández (1988), centrado en la adquisición del artículo, fue el primero en el ámbito del AE que indicó la dificultad para los hablantes de japonés, y las investigaciones de Nakagawa (2012) y Kobashi (2016) han seguido indagando sobre el reto que supone para los japoneses el dominio del artículo. La mayoría de estos análisis se ha llevado a cabo partiendo de una prueba de juicio de gramaticalidad o con corpus recopilados por los investigadores.

Santos y Alexopoulou (2021), tras haber analizado las 95 tesis doctorales sobre AE defendidas en las universidades españolas en los últimos 30 años, enfatizan la relevancia de los corpus de aprendientes. Son instrumentos imprescindibles para desarrollar estudios de análisis de errores e interlengua contrastables, comparables y fiables (Baralo, 2010).

Esta investigación, cuyo objeto de análisis es un subcorpus de textos escritos por informantes japoneses extraído del Corpus Escrito del Español L2 (CEDEL2) (v. 2.0, sept. 2020) (Lozano y Mendikoetxea, 2013; Lozano, 2021), pretende ampliar la información sobre la adquisición del artículo español con el análisis de la producción real de los estudiantes. Partiendo de los trabajos anteriores realizados con aprendientes japoneses y de otras L1 que también carecen de artículo, nuestros objetivos<sup>2</sup> son: (1) identificar los usos correctos y erróneos del artículo registrados en el corpus; (2) determinar y describir los errores más frecuentes relacionados con el uso del artículo según el nivel de EL2; y (3) comparar los resultados con los de otros trabajos sobre aprendientes cuyas L1 tampoco poseen artículo.

## 1. Marco teórico

### 1.1. Concepto de error en L2

El error está presente en todo el proceso de la enseñanza y aprendizaje de segundas lenguas (L2). Durante muchos años una de las funciones del docente ha sido evitar que el discente cometiera errores: eran elementos que debían desaparecer por completo del panorama de la enseñanza.

Desde que las teorías innatista y cognitivista se empezaron a aplicar al campo de investigación de la Adquisición de Segundas Lenguas (ASL), el aprendizaje de la L2 se define como un proceso creativo, como la puesta en marcha de un mecanismo interno capaz de construir la gramática de una lengua dada a partir de los datos a los que está expuesto el aprendiente (Fernández, 1995, p. 207). Así el error dejó de ser un «pecado» y se convirtió en algo necesario para el aprendizaje.

Corder (1967, p. 167) distingue entre *error*, *falta* o *equivocación*. El error se caracteriza por la frecuencia y la sistematicidad. La frecuencia se mide en dos términos: *frecuencia absoluta* —número de veces que aparece un error— y *frecuencia relativa* —ocasiones que se detecta en proporción con la extensión de una producción—. La *sistematicidad* se traduce en que una regla no se ha asimilado todavía y se expresa de forma errónea o incompleta en un estadio continuo de aproximación al sistema de la L2. El error forma parte de la *competencia* del aprendiente, mientras que la *falta* o *equivocación* pertenece a la *actuación*: unas veces, expresa algo correctamente y, otras veces, de manera errónea. Por último, el *lapsus*, *fallo*, *descuido* u *olvido* es una disfunción de la actividad cerebral causada por una falta de concentración o de memoria, o por la fatiga. No es exclusivo de los estudiantes de L2 sino de cualquier hablante; por lo tanto, queda excluido del análisis. En resumen, solo los errores frecuentes y sistemáticos constituyen el objeto de estudio en la metodología de AE, puesto que de ahí se podría reconstruir el proceso de adquisición y la competencia transitoria del aprendiente.

Según Corder (1967, p. 168), los enunciados correctos no siempre son evidencia de que la persona domina las estructuras de la L2: puede estar repitiendo frases que ha oído o produciendo formas de manera literal sin transmitir la intención que hay detrás de los

enunciados. Se debe contar también con que utiliza estrategias para evitar los elementos que considera difíciles con el fin de no cometer errores (Schachter, 1974).

Las producciones lingüísticas del aprendiente se basan en su gramática y las estrategias que usa para la producción. Se expone al *input*, formula hipótesis, produce *output* basándose en sus hipótesis y, a la vez, las verifica. Por lo tanto, los errores son inevitables y necesarios como herramienta del estudiante para la comprobación y reformulación de las hipótesis confirmadas o refutadas. Al mismo tiempo, son necesarios para el docente para conocer cuál es el estadio de aprendizaje por el que pasa el estudiante, además de conocer cuáles son las estrategias que utiliza y orientarlo sobre lo que debe enseñar. Y, al investigador, los errores le aportan información sobre el proceso de aprendizaje y los procesos psicológicos subyacentes (Santos, 2004).

## 1.2. Análisis de errores del uso del artículo español

### 1.2.1. Análisis de errores del uso del artículo de aprendientes japoneses

El español y el japonés son lenguas tipológicamente distantes. En japonés los sustantivos sin determinantes poseen capacidad referencial; es decir, el hablante puede utilizar un nombre escueto para aludir por primera vez a un objeto y para referirse de nuevo a él. De manera que el sistema de referencia principalmente se establece a través del contexto oracional, discursivo o situacional, y «en última instancia el conocimiento del mundo y del lenguaje, [es] el que aporta la información necesaria para establecer la referencia oportuna en caso de haber alguna» (Montero, 2011, p. 3); no obstante, también existen recursos lingüísticos para expresar la información nueva o consabida.

Los resultados de la investigación de Fernández (1988) sobre el uso del artículo indican que 142 errores (de 161 usos) están relacionados con su presencia/ausencia; es decir, la adición innecesaria y la omisión donde se debería emplear, y 98 de esos errores corresponden a informantes japoneses. En el corpus de este grupo «un 20 % de los contextos posibles de aparición del artículo presenta algún problema» (1988, p. 111) y la causa aducida es la carencia de artículo en la L1, mientras que, en los informantes árabes, alemanes y franceses, que poseen el sentido del uso/omisión del artículo, el error se debe a falsas analogías de la L2 (1988, p. 112).

Los resultados de investigaciones a partir de AE de producciones escritas<sup>3</sup>, como las de Fernández (1988), Saito (2005) y Hamamatsu (2008), revelan que el artículo español es una de las principales dificultades para los japoneses. Saito (2005, p. 83) examina 31 redacciones de cuatro tipos (carta formal e informal, receta de cocina y artículo) y llega a la conclusión de que el uso u omisión del artículo y la selección entre el determinado y el indeterminado son los problemas centrales relacionados con el empleo del artículo encontrados en el corpus. Por su parte, Hamamatsu (2008, p. 75) analiza cualitativamente nueve redacciones de la prueba de expresión escrita del examen para obtener el Diploma de Español como Lengua Extranjera (DELE) de nivel B e indica que algunos de los errores relacionados con el uso del artículo se deben a la interferencia de la L1, dada la inexistencia de esa categoría, si bien en este estudio se aplica un criterio etiológico para clasificar los errores y no se contabiliza el número de errores según el tipo. Hamamatsu (2008) concluye que el aprendizaje del artículo es uno de los temas especialmente difíciles de asimilar hasta el nivel intermedio. Por otro lado, investigaciones como las de

Hashimoto (2004), Nakagawa (2012) y Kobashi (2016), en las que se emplean pruebas gramaticales, también han confirmado que el uso del artículo es uno de los mayores escollos para los universitarios japoneses.

En general, a medida que avanza el aprendizaje y se profundiza en el conocimiento de la L2, la frecuencia de los errores es menor. Sin embargo, para los japoneses, en el caso del artículo, los errores persisten incluso en los niveles avanzados. Montero (2009, p. 140), de acuerdo con Fernández (1997, p. 97), explica que es «uno de los errores más difíciles de erradicar y que ofrece una resistencia significativa y fosilizable». No obstante, Kobashi (2016, p. 505) confirma «la existencia de la relación entre el nivel de español y el grado de precisión en el uso del artículo en dicha lengua por parte de los hablantes japoneses».

Nakagawa (2012, p. 275), por su parte, explica que, «en general, resultan difíciles aquellos determinantes cuyo empleo se suele omitir en el japonés cuando la determinación de los mismos se deduce por el contexto» e indica la dificultad del empleo del artículo indeterminado<sup>4</sup> (2012, p. 254). Además, señala que hay una tendencia a evitar el artículo para no cometer errores. Kobashi (2016, p. 545) añade que «los porcentajes de aciertos varían según la estructura gramatical de cada una de las oraciones».

Por otro lado, Barreyro (2020, p. 67) señala que, en los manuales utilizados en las aulas de EL2 en las universidades japonesas, las explicaciones del uso del artículo se basan principalmente en la Teoría de la información conocida y no se extienden a más usos. Se presentan ejercicios de traducción o de rellenar huecos en oraciones y no se trata el tema desde el nivel discursivo.

### 1.2.2. Análisis de errores en el uso del artículo de aprendientes con L1 que carecen de artículo

La influencia de la L1 es uno de los principales temas de indagación en el campo de la ASL (Ortega, 2015, p. 245). Los estudios sobre el dominio de los usos del artículo por parte de estudiantes cuya L1 carece de esta categoría gramatical—como el chino (Lin, 2003; Jie, 2012; Coto, 2015; Zhang, 2016), el croata (Kočman, 2011), el serbio (Rubinjoni, 2017), el esloveno (Santiago, 2009), el turco (García y Seoane, 2015; Nergiz, 2017), y el polaco y el ruso (Dent, 2006; Zimny, 2017)— nos permiten contrastar resultados, aunque no siempre es posible realizar comparaciones totales debido a las diferencias metodológicas o a la falta de datos cuantitativos.

Lin (2003), Jie (2012) y Zhang (2016) coinciden en que la omisión del artículo es el error más frecuente y lo que difiere en sus estudios es el orden de frecuencia de los tipos de error en 2.º, 3.º y 4.º lugar. En el corpus de estudiantes taiwaneses de 80 textos expositivos (Lin, 2003, pp. 142, 168), los errores más frecuentes son los de omisión (37 %), seguidos de los de concordancia (26 %), de selección (19 %) y de adición (18 %). En el corpus de 90 redacciones de tema libre de estudiantes chinos (Jie, 2012, pp. 162-164), los errores más frecuentes son los de omisión (71 %), de selección (15 %), los ortográficos o de concordancia (9 %) y los de adición (5 %). En el corpus de 136 redacciones narrativas de estilo epistolar escritas por universitarios chinos (Zhang, 2016, p. 175), los errores más frecuentes son los de omisión (47 %), selección (31 %) y adición (22 %) (2016, p. 405).

Por su parte, Rubinjoni (2017, p. 147) recopila y analiza dos corpus de universitarios serbios: uno, en 2011, de 102 redacciones sobre un viaje soñado y otro, en 2014, de 94 textos sobre un tema académico. Los errores más frecuentes relacionados con el uso del

artículo encontrados en el corpus de 2011 fueron los de omisión (66 %), adición (21 %) y selección (13 %) y en el de 2014, los de omisión (47 %), adición (45 %) y selección (8 %); por lo tanto, el orden de los tipos de error es el mismo en ambos. En el trabajo de Santiago (2009, p. 39), centrado en el artículo determinado en un corpus de estudiantes eslovenos constituido por 110 redacciones (49 textos argumentativos, 18 narrativos, 23 informativos, 17 descriptivos y 3 trabajos fin de grado), encontramos el mismo orden y se añaden los errores de concordancia en cuarto lugar: omisión (72 %), adición (12 %), selección (10 %) y concordancia (6 %).

Por último, en el corpus de informantes turcos formado de 102 redacciones de tres tipos de textos: narrativo, descriptivo-expositivo e instructivo (Nergiz, 2017, p. 132), de todos los errores relacionados con el uso de artículo, los más frecuentes son los de omisión (45 %), adición (29 %), selección (25 %) y otros errores (1 %).

Por lo tanto, observamos que la omisión del artículo es el error más frecuente en los corpus de hablantes de chino, lenguas eslavas (serbios y eslovenos) y turco; y también coincide con el análisis del corpus de informantes japoneses de Saito (2005) en el que el orden de los errores más frecuentes ha sido: la omisión (68 %), la adición (19 %) y la selección (13 %). En definitiva, el orden de frecuencia de los tres tipos de error (omisión, adición y selección) es el mismo en los corpus de hablantes serbios, eslovenos, turcos y japoneses.

Desde el punto de vista sintáctico, Lin (2003, p. 274), Santiago (2009), Kočman (2011, p. 4), y García y Seoane (2015, p. 55) concluyen que el artículo determinado se omite erróneamente con mayor frecuencia cuando el SN desempeña la función de sujeto o de tópico, la de atributo y la de complemento del nombre. Por otro lado, respecto al indeterminado, Lin (2003, p. 274) señala que los errores de omisión ocurren cuando el artículo no se corresponde con el cuantificador numeral chino *yī* ('uno').

Rubinjoni (2017, pp. 294, 298, 304-305) indica que los errores más persistentes en los corpus de hablantes serbios son la omisión del determinado en la 1.<sup>a</sup> mención del nombre cuando está identificado por modificadores restrictivos o por el contexto situacional y en la 2.<sup>a</sup> mención del nombre con referente conocido por anáfora directa o asociativa. Al comparar los resultados con los de hablantes de otras lenguas eslavas (ruso, polaco y esloveno) y de chino, señala que el tipo de error más frecuente también es la omisión del artículo.

Además, Rubinjoni (2017, p. 302) constata que, cuando los aprendientes serbios y chinos prestan más atención al artículo, aumentan los errores de adición. Concluye que los aprendientes cuya lengua carece de marcador explícito para identificar el referente del sustantivo se enfrentan con la dificultad de asimilar un sistema lingüístico que emplea nombres con artículo, puesto que en sus L1 en la mayoría de los contextos la identificación es automática o se realiza a través de otros elementos lingüísticos como los demostrativos y los pronombres, o cambiando el orden de palabras.

### 1.3. Análisis de errores y corpus informatizados de aprendientes

El AE ha venido acompañado de la necesidad de analizar amplias muestras de la producción lingüística real de los aprendientes como fuente de información. No obstante, Santos y Alexopoulou (2021, p. 17) indican que un 41,9 % de las tesis analizadas se basa en corpus reducidos y la mayor parte de los informantes pertenecen a un grupo ya formado y normalmente constituido por los alumnos del autor-investigador. Lozano

(2021, p. 139) señala que también es preciso analizar datos de corpus, ya que contienen datos naturales y contextualizados.

Santos y Alexopoulou (2021, p. 17) consideran que «hoy en día la investigación debe canalizarse a través de corpus de estudiantes, informatizados y etiquetados». Al mismo tiempo, señalan que existen diferentes criterios para la clasificación del nivel de dominio de los participantes en los distintos estudios de AE y sugieren establecer uno común: «nosotros hemos unificado [las nomenclaturas] estableciendo equivalencias y siguiendo los niveles de referencia propuestos en 2002 para el español en el *Marco común europeo de referencia para las lenguas*» (MCER) (2021, p. 9) (Tabla 1). En esta investigación, seguimos la nivelación del MCER (Consejo de Europa, 2002).

Tradicional	DELE antiguo	DELE actual	MCER
Iniciación		DELE A1	A1 Acceso
Elemental	DELE inicial	DELE A2	A2 Plataforma
Intermedio bajo		DELE B1	B1 Umbral
Intermedio alto	DELE intermedio	DELE B2	B2 Avanzado
Avanzado		DELE C1	C1 Dominio operativo eficaz
Superior	DELE superior	DELE C2	C2 Maestría

**Tabla 1. Nomenclaturas de los niveles de referencia (Gargallo y Alexopoulou, 2021, p. 9)**

## 2. Análisis del corpus

### 2.1. Procedimiento

#### 2.1.1. Procedimiento general del Análisis de errores

El procedimiento de este estudio se basa en el modelo del AE propuesto por Corder (1971, pp. 155-156), que consiste en tres etapas: identificación, descripción y explicación de los errores (o, al menos, un intento de determinar qué los causa)<sup>5</sup>. En cuanto a la clasificación adoptada para la descripción de los errores, hay que tener presente que no existe una taxonomía estándar —lo cual, por otra parte, complica la comparación de los resultados de las diferentes investigaciones—. Vázquez (1986), Fernández (1990) y Santos (2004), siguiendo el modelo de Corder, propusieron, en su momento, varias clasificaciones para el español con el fin de describir los errores<sup>6</sup>. Más recientemente, en la misma línea, Alexopoulou (2006, p. 19) delimita el análisis en dos ámbitos: el descriptivo, que hace referencia al producto, y el etiológico, al proceso. El primero consiste en describir «los errores en términos de la diferencia física entre el enunciado del estudiante y la versión reconstruida» (2006, p. 20) y en este nivel propone cuatro categorías: *omisión* de elementos obligatorios, *adición* de elementos innecesarios o incorrectos, *selección* de elementos incorrectos y *mal ordenamiento* de los elementos. Una vez asignada una de estas categorías, los errores se analizan con más profundidad empleando un criterio lingüístico: a) fonético-fonológico-ortográfico; b) morfológico; c) sintáctico; d) léxico-semántico, e) discursivo, f) pragmático, según el objetivo de la investigación. En segundo lugar, el ámbito etiológico corresponde a la etapa de explicación de los errores, que consiste en intentar «buscar el “cómo” y el “porqué” del comportamiento lingüístico del aprendiente mediante la observación sistemática de su producción interlingüística»

(Alexopoulou, 2006, p. 23).

Esta investigación se centra en el ámbito descriptivo, cuyo proceso consiste en (1) la selección y formación del corpus, (2) la identificación y clasificación de los usos del artículo y (3) la evaluación y análisis de los casos.

### 2.1.2. Selección y formación del corpus

Actualmente hay disponibles varios corpus de producciones escritas de aprendientes de EL2 de diferentes L1, pero solo algunas cuentan con datos de informantes japoneses: CEDEL2 (Lozano, 2009) (<http://cedel2.learnercorpora.com/>), el Corpus de Textos Escritos para el Análisis de Errores de Aprendices de E/LE (CORANE) (Cestero y Penadés, 2009) y el Corpus de ELE en Japón (CELEN) (Valverde, 2020) (<https://sites.google.com/view/celen>), solo de universitarios —el 65 % de nivel A1; el 25 % de A2 y el 10 % de B1—.

Para esta investigación se ha seleccionado un subcorpus de CEDEL2 con el objetivo de analizar datos de informantes japoneses con un perfil amplio y no solo de universitarios, cuyo nivel de dominio suele ser A o B. Para la selección del corpus también fue determinante la posibilidad de acceder a los textos completos con el fin de verificar el contexto discursivo de los usos del artículo.

La recopilación de los datos de CEDEL2 comenzó en 2006 (Lozano y Mendikoetxea, 2013, p. 83) y la versión actual contiene información de 4399 participantes y 1 105 936 palabras —es el mayor corpus de aprendientes de EL2 de 11 lenguas, incluyendo el japonés; y contiene también textos en la L1 de hablantes de seis lenguas—.

En cuanto a la nivelación de los informantes según el dominio de la L2, Lozano (2009, p. 203) explica que, en principio, se pensaba utilizar el test diagnóstico Dialang, de la Universidad de Lancaster, que emplea los seis niveles de referencia estandarizados del *MCER*. Pero, en aquel momento, requería que los participantes se descargaran un *software* específico, por lo que se eligió la prueba gramatical en línea diseñada por la Universidad de Wisconsin.

La recogida de datos de CEDEL2 se lleva a cabo a través de un formulario electrónico (<http://www.learnercorpora.com/>), que incluye preguntas sobre la edad, sexo, estudios, etc., de los participantes; la experiencia de aprendizaje de la L2 y el contexto en el que se realiza la tarea. Después, se completa la prueba gramatical y la tarea escrita.

El subcorpus de 243 informantes japoneses cuenta con 23 049 palabras. Para esta investigación se han seleccionado los textos narrativos (5763 palabras) escritos a partir de unas viñetas procedentes del cuento *Frog, where are you? (A boy, a dog, and a frog)*, de Mercer Mayer, publicado en 1969 por la editorial Dial Books for Young Readers en Nueva York. Este subcorpus de 49 informantes lo forman 12 de nivel A1-A2, según el *MCER* (Consejo de Europa, 2002) (nivel elemental bajo y alto, según el test de la Universidad de Wisconsin); 19, de B1-B2 (intermedio bajo y alto) y 18, de nivel C1-C2 (avanzado bajo y alto). El rango etario es de 18 a 76 y la media de edad, 35 años; 32 son mujeres (65 %) y 17, hombres (35 %).

### 2.1.3. Identificación y clasificación de los usos del artículo

Con el fin de identificar de forma semiautomática los usos del artículo en el corpus, se ha empleado el formulario de búsqueda de CEDEL2 y la herramienta #LancsBox

6.0, creada por investigadores de la Universidad de Lancaster (Brezina, Weill-Tessier y McEnery, 2020). El formulario del corpus permite buscar por categoría gramatical (nombre: común, propio; determinante: artículo, indefinido) y recuperar los casos correspondientes. #LancsBox dispone de un sistema automático de etiquetado de palabras según su categoría gramatical adaptado a 20 lenguas, entre ellas el español. La función Key Word in Context (KWIC) (Figura 1) localiza los casos en el contexto mediante las siglas POS (*part-of-speech*) (Schmid, 1994): ART identifica los artículos (*el, la, los, las, un, uno, una, unos, unas y lo*); PAL, la contracción de la preposición *a* más el artículo determinado *el* (*al*) y PDEL, la contracción *del* (*de* más *el*). Para encontrar los sustantivos sin artículo o acompañados de otros determinantes, se han utilizado las siglas NC (nombre común), PREP (preposición), PPO (posesivo), DM (demostrativo), CARD (número cardinal) y ORD (ordinal). De esta forma, se ha detectado automáticamente la presencia y la ausencia del artículo —y de otros elementos— ante sustantivos.

Index	File	Left	Node	Right
1	JP_WR_41_1	perro al bosque, en su camino encontró	una	rana perdida. El niño decidió llevarla a
2	JP_WR_41_1	los árboles, arriba de las rocas, en	una	de ellas un venado estaba durmiendo y
3	JP_WR_41_1	que regresar pero se puso triste y	una	de las ranas quiso acompañar lo. Sheet1
4	JP_WR_41_3	día, el niño y su perro trajeron	una	rana a la casa para criarla como
5	JP_WR_41_3	rana a la casa para criarla como	una	mascota. Sin embargo, mientras que los dos
6	JP_WR_41_3	la rana. Y el niño subió a	una	pedra grande agarrando una lama de árbol,
7	JP_WR_41_3	niño subió a una piedra grande agarrando	una	lama de árbol, para buscar la rana
8	JP_WR_41_3	obstante, lo que él agarró no era	una	lama, sino era un ciervo que se
9	JP_WR_41_4	ha estado en un parque y encontró	una	rana. La ha traído a casa y
10	JP_WR_41_4	a dormir pero se le olvida ponerle	una	tapa al bote. Al día siguiente, cuando
11	JP_WR_41_4	partes pero no la encuentra. Mira, hay	una	roca grande. ¡Tal vez de allí arriba
12	JP_WR_41_4	árbol sino un ciervo que estaba echando	una	siesta! Asustado, el ciervo se levanta tan
13	JP_WR_41_4	familia cantando de alegría. Los bebés pasaron	una	noche muy preocupados por su padre que
14	JP_WR_42_5	JP Un día,	una	rana se escapó de jaula cuando un
15	JP_WR_42_5	cama. El chico tiene un perro y	una	rana, que son muy amigos. Siempre juega
16	JP_WR_42_5	unos días. Un día el subió a	una	colina y gritó el nombre de rana
17	JP_WR_43_4	día, un niño y su perro atraparon	una	rana y la metieron en una botella.
18	JP_WR_43_4	atraparon una rana y la metieron en	una	botella. Mientras dormía el niño, la rana
19	JP_WR_43_4	un ciervo, este los hizo caer en	una	laguna. Sin embargo, encontraron a la rana
20	JP_WR_43_4	JP Un día Alberto trajo a casa	una	rana que había encontrado en el bosque.
21	JP_WR_43_4	bosque, subiendo a un árbol o a	una	roca. De repente apareció un ciervo y
22	JP_WR_43_4	día el niño encontró en el bosque	una	rana preciosa del color azul verdoso, y
23	JP_WR_43_4	un cuenco de vidrio y tapó con	una	tabla, y la observaba encantado junto con

Figura 1. Ejemplo de la búsqueda con la función KWIC en #LancsBox 6.0

Los casos del artículo localizados en el corpus se han clasificado según los usos descritos en la *Nueva gramática de la lengua española* (RAE y ASALE, 2009), la *Gramática descriptiva de la lengua española* (Bosque y Demonte, 1999), la *Gramática comunicativa del español* (Matte, 1995) y Morimoto (2011). A continuación, se ha evaluado la corrección/incorrección de cada caso siguiendo el criterio descriptivo: error por *omisión, adición* o *selección* del artículo (Alexopoulou, 2006) —es la misma taxonomía empleada por Zhang (2016)—. Si el artículo determinado/indefinido está bien seleccionado, los errores en las formas contractas (*al* y *del*) —p. ej., *\*a el* o *\*de el bosque*— o de concordancia de género y número —p. ej., *\*un rana* en la 1.ª mención del nombre— no se han tenido en cuenta. El juicio de gramaticalidad lo han determinado dos correctores: un hablante nativo de japonés y de español como L2, y otro, nativo de español —al igual que en Zhang (2016, p. 175)—, de manera que se ha realizado una revisión doble. Después, se ha analizado el contexto de los errores con criterio sintáctico y discursivo; y, finalmente, los cálculos estadísticos<sup>7</sup> se han efectuado con el programa IBM SPSS Statistics para Windows (IBM Corp. 2020) y han permitido averiguar si la diferencia entre aciertos y errores es significativa.



## 2.2. Resultados

En el corpus se han detectado 1109 casos de uso del artículo: 925 correctos y 184 erróneos, lo que significa un porcentaje general de error de un 16,59 % (Tabla 2). Los resultados del análisis estadístico [ $t(1108) = 14,846$   $p = 0,000$ ;  $d$  de Cohen = 0,4] confirman que la diferencia entre el número de aciertos y errores es significativa y no se debe al azar.

En los textos de nivel A, se han identificado 180 usos del artículo: 129 correctos y 51 erróneos: el 28,33 % de los usos son incorrectos. En el nivel B, 388 usos del artículo: 304 correctos y 84 erróneos (porcentaje de error: 21,65 %). En el nivel C, se han localizado 541 usos del artículo: 492 correctos y 49 erróneos (porcentaje de error: 9,06 %). El análisis estadístico indica que la diferencia entre los resultados obtenidos en cada nivel es significativa; por lo tanto, el nivel de dominio incide en el uso del artículo ( $Chi^2 = 47,3$ ; sig. = 0,000 //  $V$  de Cramer = 0,21; sig. = 0,000). No obstante, el coeficiente  $V$  de Cramer, al estar más cerca de 0 que de 1, muestra que la asociación entre variables es pequeña.

Nivel MCER	N.º de informantes	Palabra (token)	Vocablo (type)	Type/Token ratio %	Lema	N.º total de casos	Frecuencia de uso del artículo	N.º total usos correctos	N.º total usos erróneos	% de errores
A	12	753	298	39,58 %	275	180	23,90 %	129	51	28,33 %
B	19	1951	612	30,39 %	546	388	19,89 %	304	84	21,65 %
C	18	3059	845	27,62 %	734	541	17,69 %	492	49	9,06 %
Total	49	5763	1755	30,45 %	1555	1109	19,24 %	925	184	16,59 %

**Tabla 2. Información general del corpus por niveles, según el MCER**

En cuanto al tipo de error, el 64,13 % se debe a la selección incorrecta del artículo; el 33,15 %, a la omisión y el 2,72 %, a la adición: los errores de selección oscilan entre el 60 y 67 % y los de omisión, entre el 28 y 39 %; los de adición, alrededor del 4 %, pero solo se registran en textos de nivel B y C.

### 2.2.1. Subcorpus de textos de nivel A

De los 51 errores localizados en los textos de nivel A (Tabla 3), 20 (39,22 %) se deben a la omisión del artículo y 31 (60,78 %), a la selección del mismo. No se han identificado errores por adición.

	Uso del artículo	Omisión	Adición	Selección	N.º total de errores
Determinado	Anáfora directa	7	0	3	10
	Deíctico espacial	2	0	0	2
	Con modificador	2	0	0	2
Indeterminado	1.ª mención	8	0	28	36
	Evaluativo con modificador calificativo	1	0	0	1
	<b>N.º total de errores</b>	20	0	31	51
	%	39,22 %	0 %	60,78 %	100 %

**Tabla 3. Errores en el subcorpus de textos de nivel A según los usos del artículo**

Los errores por omisión del artículo determinado ocurren cuando se trata de una anáfora directa: en la posición de sujeto preverbal (1), en una enumeración después de la conjunción *y* (2), en un complemento preposicional (3) y en un objeto directo (4). En

los ejemplos los artículos y los sustantivos están subrayados y se mantiene la ortografía del original.

- (1) Rana se escapó mientras el niño está durmiendo. (n.º 13; informante 2)
- (2) El chico y perro fueron su casa. (n.º 75; 4)
- (3) Cuando él cao en lago, [...]. (n.º 93; 5)
- (4) Ellos buscan ranas. (n.º 192; 11)

También en el uso deíctico espacial (5) y con modificador (6).

- (5) [...] la rana hue de botella. (n.º 86; 5)
- (6) Finalmente el niño encuentro la rana [en] parted[e] un árbol. (n.º 1417; 7)

Los errores por omisión del artículo indeterminado se hallan en la 1.ª mención del sustantivo (7) y en el uso evaluativo (8).

- (7) [...] y el chico llega a estanque. (n.º 50; 3)
- (8) La rana tenía familia Fuelle. (n.º 78; 4)

Los errores de selección del artículo se observan en la 2.ª mención del sustantivo donde se da un uso anafórico (9) y en la 1.ª mención (10).

- (9) [...] está entre en una botella de vidrio. Escapa de una botella de vidrio. (n.º 4; 1)
- (10) Un día el niño había mantenido la rana en la botella. (n.º 10, 11 y 12; 2)

En la 2.ª mención se debería utilizar el artículo determinado: en (9) *la botella de vidrio*; y en la 1.ª, el indeterminado: en (10) *un niño, una rana y una botella*.

### 2.2.2. Subcorpus de textos de nivel B

En los textos de nivel B, se han localizado 84 errores relacionados con el uso del artículo (Tabla 4), de los cuales, 27 (32,14 %) se deben a la omisión; 3 (3,57 %), a la adición; y 54 (64,29 %), a la selección del determinado por el indeterminado o viceversa.

	Uso del artículo	Omisión	Adición	Selección	N.º total de errores
Determinado	Anáfora directa	7	0	9	16
	Anáfora asociativa	1	0	1	2
	Deíctico espacial	2	0	0	2
	Deíctico temporal	4	0	1	5
	Con modificador	1	0	2	3
	Genérico	1	0	0	1
Indeterminado	1.ª mención	9	0	41	50
	Genérico	1	0	0	1
	Expresión fija	1	0	0	1
∅	Complemento preposicional	0	1	0	1
	Vocativo	0	2	0	2
	<b>N.º total de errores</b>	27	3	54	84
	%	32,14 %	3,57 %	64,29 %	100 %

**Tabla 4. Errores en el subcorpus de textos de nivel B según los usos del artículo**

Los errores de omisión del artículo determinado en los textos de los informantes de nivel B aparecen en la 2.ª mención del sustantivo (anáfora directa) detrás de la conjunción y (11), en la posición de sujeto (12) y en el complemento preposicional (13).

- (11) Por la mañana siguiente el niño y perro [...]. (n.º 328; 19)
- (12) [...] a un lugar que están ranas. (n.º 652; 31)
- (13) Nosotros hemos uso a mi casa con perro y rana. (n.º 416; 23)

Se localizan también en el uso de anáfora asociativa (14), deíctico espacial (15), temporal (16), con modificador (17) y en el uso genérico (18).

- (14) [...] la rana quiera en agua. (n.º 278; 15)
- (15) En camino, se encontraron un ciervo. (n.º 458; 25)
- (16) Próxima mañana, encontró la caja vacía. (n.º 367; 21)
- (17) [...] una rana trataba de escapar desde interior de una jarra. (n.º 595; 30)
- (18) El y su perrito fueron a dentro de selva. (n.º 1418; 25)

Los errores de omisión del indeterminado se detectan en la 1.ª mención (19), en el uso genérico (20) y en expresiones fijas (21).

- (19) El metió le en bote por la noche. (n.º 262; 15)
- (20) y comí desayuno. (n.º 235; 13)
- (21) Luego dieron paseo. (n.º 562; 28)

Solo hay tres casos de adición del artículo determinado en complementos preposicionales (22).

- (22) [...] llamándome con la voz alta. (n.º 687; 32)

Se localizan dos errores de adición en un vocativo (23a, 23b).

- (23a) Él la llamó "La rana, ¿dónde estás? [...]". (n.º 643; 31)
- (23b) [...] y probó a gritar. "La rana. Mi amiga rana, ¿Dónde estás?". (n.º 647; 31)

Los errores de selección del artículo indeterminado por el determinado se registran en la 2.ª mención del sustantivo (anáfora directa) en la posición de sujeto (24), en el deíctico temporal (25), con presencia de modificador en la posición de sujeto (26), en el complemento preposicional (27) y en la anáfora asociativa (28).

- (24) [...] ellos ven a un animal. Un animal grande hace algo [...]. (n.º 288; 16)
- (25) Pero por una noche, huyó [...]. (n.º 325; 19)
- (26) Una familia de las ranas estaba [...]. (n.º 549; 27)
- (27) Cuando llegaron a un borde de un precipicio, la cierva dijo [...]. (n.º 655; 31)
- (28) Llegaron hasta un borde y [...]. (n.º 494; 26)

Los errores de selección del artículo determinado aparecen en la 1.ª mención del nombre en la posición de sujeto (29, 30a, 30b), en el complemento preposicional (30c) y el objeto directo (31).

- (29) Un día, el chico atrapó una rana. (n.º 259; 15)
- (30a) El chico y (30b) el perro viven con (30c) la rana. (n.º 279, 280, 281; 16)
- (31) Tenía la rana amallia. (n.º 388; 22)

### 2.2.3. Subcorpus de textos de nivel C

De los 49 errores localizados en los textos de nivel C (Tabla 5), 14 (28,57 %) se deben a la omisión del artículo; solo 2 (4,08 %), a la adición; y 33 (67,35 %), a la selección.

	Uso del artículo	Omisión	Adición	Selección	N.º total de errores
Determinado	Anáfora directa	2	0	5	7
	Anáfora asociativa	1	0	1	2
	Con modificador	2	0	3	5
	Expresión fija	2	0	0	2
Indeterminado	1.ª mención	5	0	24	29
	Genérico	1	0	0	1
	Expresión fija	1	0	0	1
	Predicado con verbo soporte	0	1	0	1
∅	Complemento preposicional	0	1	0	1
	<b>N.º total de errores</b>	14	2	33	49
	%	28,57 %	4,08 %	67,35 %	100 %

**Tabla 5. Errores en el subcorpus de textos de nivel C según los usos del artículo**

Los errores de omisión del determinado se identifican en la 2.ª mención del sustantivo (anáfora directa) en el objeto indirecto (32) y en anáfora asociativa (33); en la construcción con el cuantificador *todo* con modificador (34), y en expresiones fijas (35).

(32) [...] le dijo a ranita [...]. (n.º 1396; 50)

(33) Cuando llegaron a suelo [...]. (n.º 895; 37)

(34) La buscó y buscó por toda parte del bosque. (n.º 1058; 42)

(35) Juanito le dio gracias a la familia [...]. (n.º 1082; 42)

Los errores de omisión del indeterminado se presentan en la 1.ª mención en complemento preposicional (36), en el uso genérico (37) y en expresiones fijas (38).

(36) Un día, una rana se escapó de jaula. (n.º 1242; 47)

(37) La rama que cogió Antonio no era rama sino [...]. (n.º 1103; 43)

(38) Raúl y Jack se despidieron a Puff con sonrisa. (n.º 865; 36)

Se han localizado dos casos erróneos de adición en el predicado de un verbo soporte (39) y en el complemento del sustantivo (40).

(39) Y cuando el niño retomó el aire y miró a su alrededor [...]. (n.º 1387; 50)

(40) [...] como si fuera una obra del arte. (n.º 717; 33)

Los errores de selección del artículo indeterminado en vez del determinado se detectan en el uso de anáfora directa (41), asociativa (42) y en el no anafórico con modificador (43).

(41) En el gran bosque un niño y Jack seguían [...]. (n.º 831; 36)

(42) Después de una aventura, ellos volvieron a casa juntos. (n.º 744; 33)

(43) [...] cayó en un lago desde un borde del lago. (n.º 939; 38)

Los errores de selección se identifican en la 1.ª mención del sustantivo en la posición de sujeto preverbal (44), y en el complemento preposicional (45), donde debería emplearse el artículo indeterminado.

(44) Un día, el niño y su perro trajeron una rana. (n.º 1152; 45)

(45) Suben a la pequeña roca. (n.º 761; 34)

### 2.3. Discusión

Este análisis de 1109 usos del artículo revela que los casos erróneos son 184 y que el porcentaje de error disminuye con el avance de dominio de la L2: del 28,33 % en el nivel A al 9,06 % en el C, y estadísticamente la diferencia entre niveles es significativa.

En términos globales, la selección incorrecta del artículo es el error detectado más frecuente (64,13 %), seguido de la omisión (33,15 %) y de la adición del artículo inadecuado (2,72 %). Teniendo en cuenta el uso (Tablas 3-5), la selección del artículo determinado, en lugar del indeterminado, en la 1.<sup>a</sup> mención del sustantivo, es el tipo de error más frecuente: 93, en total; 28 errores en el nivel A (el 54,9 % de los errores identificados en ese nivel); 41, en el B (48,81 %) y 24, en el C (48,98 %). El segundo tipo de error más frecuente es la omisión del indeterminado en la 1.<sup>a</sup> mención del sustantivo: 22, en total; en el nivel A, hay 8 errores, que es el 15,69 % de todos los encontrados; en el B, 9 (10,71 %); y en el C, 5 (10,2 %). El tercer error más frecuente es la selección errónea del indeterminado en la 2.<sup>a</sup> mención: 17, en total; 3 errores en el nivel A; 9, en el B y 5, en el C; y, el cuarto, la omisión del determinado en la 2.<sup>a</sup> mención del sustantivo (uso anafórico): 16, en total; 7 casos en el nivel A y en el B, y 2, en el C.

Finalmente, en todos los niveles, se han localizado escasos errores del artículo determinado en el uso no anafórico con modificadores, en el deíctico temporal (*por la mañana, al día siguiente y todos los días*) y en expresiones como *al final, al fin* o *dar las gracias a*. Con el indeterminado, se identifican menos errores en el uso genérico, en el cuantificador y numérico, y en expresiones como *un instante, un momento* o  *echar una siesta*. En el grupo nominal escueto, se detectan muy pocos errores en el complemento preposicional de materia (*una botella de cristal*) y en locuciones como *a duras penas, a trancas y barrancas, de repente, por casualidad, por fin, por suerte, por supuesto* o *sin embargo*.

## Conclusiones

Tras el análisis de los 1109 usos del artículo en 49 textos narrativos (5763 palabras), extraídos de CEDEL2, escritos por estudiantes de EL2 japoneses, concluimos que el porcentaje de casos correctos (83,41 %) es muy superior al de los erróneos, y que el nivel de dominio incide significativamente en el uso del artículo, en el sentido de que, a medida que avanza, se cometen menos errores.

Nuestro segundo objetivo era determinar y describir los errores más frecuentes según el nivel de dominio de EL2. Observamos que los tipos de error más frecuentes —el uso del artículo determinado o la omisión del indeterminado en la 1.<sup>a</sup> mención del sustantivo; y el uso del indeterminado en la 2.<sup>a</sup> mención o la omisión del determinado— prácticamente se mantienen en los tres niveles.

El tercer objetivo era comparar los resultados con otros análisis de corpus de L1 sin artículo, al igual que el japonés y, en este sentido, el tipo de error más frecuente hallado en nuestro corpus se debe a la selección del artículo y no a la omisión, como ocurre en los corpus de estudiantes japoneses (Saito, 2005), de hablantes de chino (Lin, 2003; Jie, 2012; Zhang, 2016), de lenguas eslavas: serbio (Rubinjon, 2017) y esloveno (Santiago, 2009), y de turco (Nergiz, 2017).

Consideramos que la tipología narrativa de la tarea en la que se debe presentar información nueva y referirse a ella después ha podido influir en la alta frecuencia de errores de la selección del artículo en la 1.<sup>a</sup> y 2.<sup>a</sup> mención del sustantivo. Aparte, teniendo en cuenta el sistema de referencia en japonés, el hecho de que los informantes hayan empleado artículo en la 1.<sup>a</sup> y 2.<sup>a</sup> mención del sustantivo —aunque a veces erróneamente—, en lugar de omitirlo, se puede deber a que están en proceso de aproximación al sistema referencial de la L2.

La alternancia del artículo determinado e indeterminado para identificar el referente del sustantivo y el empleo del artículo como mecanismo de cohesión textual constituyen una dificultad significativa incluso para los aprendientes de nivel superior, y estos resultados coinciden con los de los hablantes de chino, serbio, esloveno y turco, que son lenguas en las que la identificación del referente del sustantivo se realiza a través de otros elementos lingüísticos o, en la mayor parte de los casos, por el contexto o por el conocimiento general (Rubinjon, 2017, pp. 304-305).

Por otro lado, el alto porcentaje de aciertos del artículo determinado como deíctico temporal y del nombre escueto en locuciones posiblemente se debe a que los aprendientes las han memorizado como fórmulas, con lo cual no precisan pensar si el sustantivo es identificable unívocamente o si se ha presentado con antelación, al contrario de lo que sucede al introducir un sustantivo como información nueva —ahí se da el mayor número de incidencias—. En la misma línea, con estudiantes chinos, Jie (2012, p. 250) concluye que «es posible plantear la hipótesis de que el empleo correcto del artículo en este tipo de expresiones dependa más de la memoria mecánica que de la memoria significativa».

En definitiva, los resultados de este análisis son relevantes porque aportan un mejor conocimiento del dominio de los usos del artículo por parte de hablantes de una lengua que carece de esta categoría gramatical. En el futuro, con el fin de seguir profundizando en el tema, este estudio descriptivo de AE se completará con el etiológico para tratar de determinar las causas de los errores y plantear propuestas didácticas. Además, consideramos pertinente analizar sintagmas nominales con otros determinantes como los posesivos (46, 47), puesto que podrían ser casos de sustitución estratégica del artículo para evitar errores (Saito, 2005, p. 45).

(46) Un día un niño sacó a pasear a su perrito al bosque, en su camino encontró una rana perdida. (informante 44)

(47) Lo que él agarró no era una lama, sino era un ciervo que se descansaba apoyando su cabeza en la piedra. (informante 45)

También se espera ampliar el análisis de tipos de textos —con otras tareas de CEDEL2 u otros corpus que incluyen informantes japoneses, como CORANE (Cestero y Penadés, 2009) o CELEN (Valverde, 2020)— y compararlos con textos escritos por nativos de español y por nativos de japonés con el objetivo de verificar el sistema empleado para identificar el referente de los sustantivos.

### Agradecimientos

Queremos agradecer a los profesores Cristóbal Lozano y Nobuo Ignacio López Sako el acceso a los textos de CEDEL2 de los informantes japoneses y toda la información facilitada sobre el corpus y su elaboración. Asimismo, agradecemos la colaboración del profesor Hiroto Ueda (Universidad de Tokio) para incluir los datos de este estudio en el sistema LYNEAL (Letras y Números en Análisis Lingüísticos), que se encuentran disponibles en GREC-UNED (archivo 2) (<https://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/index.html>).

### Referencias bibliográficas

- Alexopoulou, A. (2006). Los criterios descriptivo y etiológico en la clasificación de los errores del hablante no nativo: una nueva perspectiva. *Porta Linguarum*, 5, 17-35.
- Baralo, M. (2010). La investigación en español como segunda lengua: necesidad de un corpus de español no nativo. En *V Congreso Internacional de la Lengua Española*. Valparaíso. <https://congresosdelalengua.es/valparaiso/paneles-ponencias/lengua-educacion/baralo-marta.htm>.

- Barreyro, E. (2020). *Los rasgos del artículo definido en español. Una propuesta didáctica para la enseñanza de determinantes a hablantes nativos de japonés*. TFM, U. de Estudios Extranjeros de Kobe.
- Bosque, I., y Demonte, V. (1999). *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe.
- Brezina, V., Weill-Tessier, P., y McEnery, A. (2020). #LancsBox v. 6.0.
- Campillos, L. (2012). *La expresión oral en español lengua extranjera: interlengua y análisis de errores basado en corpus*. Tesis doctoral, UAM.
- Cestero, A. M., y Penadés, I. (2009). *Corpus de Textos Escritos para el Análisis de Errores de Aprendices de E/ELE (CORANE)*, CD-rom. Alcalá de Henares: UAH.
- Consejo de Europa (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación (MCER)*. Madrid: Ministerio de Educación. Cultura y Deporte, Anaya.
- Corder, S. P. (1967). The significance of learner's errors. *IRAL*, 5(4), 161-170.
- Corder, S. P. (1971). Idiosyncratic dialects and error analysis. *IRAL*, 9(2), 147-160.
- Coto, V. (2015). Aproximación a la interlengua de los alumnos sinohablantes de ELE. En Y. Morimoto *et al.* (Eds.), *La enseñanza de ELE centrada en el alumno* (pp. 285-294). Madrid: ASELE.
- Denst, E. (2006). *Uso del artículo en español para hablantes eslavos*. Tesis doctoral, UCM.
- Fernández, S. (1988). El uso del artículo en aprendices de español lengua extranjera. En *Actas del I Congreso de ASELE* (pp. 109-118), Granada: ASELE.
- Fernández, S. (1990). *Interlengua y análisis de errores en el aprendizaje del español como lengua extranjera [en HNN de origen francés, alemán, árabe y japonés]*. Tesis doctoral, UCM.
- Fernández, S. (1995). Errores e interlengua en el aprendizaje del español como lengua extranjera. *Didáctica. Lengua y literatura*, 7, 203-216.
- Fernández, S. (1997). *Interlengua y análisis de errores en el aprendizaje del español como lengua extranjera*. Madrid: Edelsa.
- García, E., y Seoane, A. (2015). Errores comunes de aprendientes turcos de ELE: una propuesta didáctica. *Boletín ELE Anatolia*, 1, 50-63.
- Hamamatsu, N. (2008). Lingüística contrastiva y análisis de errores: aplicación al estudio de una muestra de hablantes japoneses aprendientes de español como lengua extranjera. *Hispanica*, 52, 67-86.
- Hashimoto, K. (2004). Sobre la didáctica de los artículos para la clase de gramática español: nivel elemental. *Gaikokugo Kyouiku Didáctica de la lengua extranjera*, 89-99.
- IBM Corp. (2020). IBM SPSS Statistics para Windows, Versión 27.0.1.0. Armonk: IBM Corp.
- Jie, S. (2012). *El artículo en la enseñanza de ELE: estudiantes de origen chino*. Tesis doctoral, UB.
- Kobashi, S. (2016). *Enseñanza y aprendizaje del artículo en español por parte de estudiantes japoneses*. Tesis doctoral, UCM.
- Kočman, A. (2011). Errores gramaticales comunes en alumnos croatas de español como segunda lengua en la producción escrita. *Marcoele*, 13, 1-12.
- Lin, T. J. (2003). *La adquisición y el uso del artículo de estudiantes chinos*. Tesis doctoral, UAH.
- Lozano, C. (2009). CEDEL2: Corpus Escrito del Español L2. En M. Bretones *et al.* (Eds.), *Applied linguistics now: understanding language and mind/La lingüística aplicada actual: comprendiendo el lenguaje y la mente* (pp. 197-212). Almería: UAL.
- Lozano, C. (2021). Corpus textuales de aprendices para investigar sobre la adquisición del español LE/L2. En M. Cruz (Ed.), *E-Research y español LE/L2: investigar en la era digital* (pp. 138-162). Nueva York: Routledge.

- Lozano, C., y Mendikoetxea, A. (2013). Learner corpora and second language acquisition: the design and collection of CEDEL2. En A. Díaz-Negrillo *et al.* (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 65-100). Ámsterdam: John Benjamins.
- Matte, F. (1995). El artículo. En *Gramática comunicativa del español* (pp. 199-216). Madrid: Edelsa.
- Montero, A. (2009). Errores frecuentes en la escritura de los universitarios japoneses, estudiantes de español de nivel A2. *Cuadernos CANELA*, XXI, 129-150.
- Montero, S. (2011). El artículo y otros fantasmas del nombre. *redELE*, 21, 1-28.
- Morimoto, Y. (2011). *El artículo en español*. Madrid: Castalia.
- Nakagawa, M. (2012). *El empleo de los determinantes de español por estudiantes japoneses de ELE: estudio contrastivo y análisis de errores*. Tesis doctoral, UNED.
- Nergiz, B. (2017). *Análisis de errores en textos escritos por estudiantes turcos del nivel B1 de español como lengua extranjera*. Tesis doctoral, UPF.
- Ortega, L. (2015). Second language learning explained? SLA across 10 contemporary theories. En B. Van Patten y J. Williams (Eds.), *Theories in second language acquisition: an introduction* (pp. 245-272). Nueva York: Routledge.
- RAE y ASALE (2009). *Nueva gramática de la lengua española. Morfología y sintaxis I (NGLE)*. Madrid: Espasa Libros.
- Rubinjoní, V. (2017). *Análisis de errores en el uso del artículo determinado en producciones escritas de aprendientes serbios de español como lengua extranjera*. Tesis doctoral, UAH.
- Saito, A. (2005). *Análisis de errores en la expresión escrita de los estudiantes japoneses*. TFM, USAL.
- Santiago, G. (2009). La adquisición del artículo determinado por aprendices eslovenos: análisis cuantitativo y cualitativo de los errores. *Journal for Foreign Languages*, 1(1-2), 35-51.
- Santos, I. (2004). El análisis de errores en la interlengua del hablante no nativo. En J. Sánchez e I. Santos (Eds.), *Vademécum para la formación de profesores: enseñar español como segunda lengua (L2)/lengua extranjera (LE)* (pp. 391-410). Madrid: SGEL.
- Santos, I., y Alexopoulou, A. (2021). Metaanálisis de las tesis doctorales de análisis de errores en la interlengua española a lo largo de tres décadas (1991-2019). *Marcoele*, 32, 2-36.
- Schachter, J. (1974). Error analysis: avoidance. *Language Learning*, 24, 205-214.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. En *Proceedings of International Conference on New Methods in Language Processing*. Manchester. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- Valverde, P. (2020). Diseño y creación de un corpus de aprendices de ELE en Japón (CELEN). *E-AESLA*, 6, 223-240.
- Vázquez, G. (1986). Hacia una valoración del concepto de error En *Actas de las I Jornadas Internacionales de Didáctica del Español como Lengua extranjera* (pp. 151-162). Madrid: Ministerio de Cultura.
- Zhang, Y. (2016). *Enseñanza-aprendizaje de ELE y análisis de errores de estudiantes chinos de español*. Tesis doctoral, USC.
- Zimny, Á. (2017). *Análisis de errores en la adquisición del artículo español por alumnos polacos de ELE*. Tesis doctoral, UCM.



## Notas

<sup>1</sup> Utilizamos el término *español como segunda lengua (EL2)* para referirnos al aprendizaje de español desde el punto de vista cronológico. Es decir, como una lengua que se aprende después de la primera indistintamente de si el aprendizaje se realiza o no en un lugar donde el español es lengua de comunicación, y sin necesidad de que sea la segunda lengua aprendida. En ese sentido, *EL2* abarca el término *español como lengua extranjera (ELE)*.

<sup>2</sup> El objetivo no es presentar estudios descriptivos diacrónicos sobre el artículo ni evaluar discusiones lingüísticas teóricas sobre esta categoría gramatical.

<sup>3</sup> Sobre análisis de corpus orales, véase Campillos (2012), que analiza los errores de un corpus oral de estudiantes con distintas L1. Los resultados indican que el grupo japonés destaca por el escaso uso del artículo (4,74 % de las unidades léxicas) y por el número de errores cometidos (124 de un total de 635). Japoneses, chinos y polacos son los que más errores cometen (2012, p. 227).

<sup>4</sup> Para mayor claridad de la exposición usamos siempre los términos *artículo determinado e indeterminado* —en la bibliografía de referencia también se emplea *artículo definido e indefinido*—, si bien hay que tener en cuenta lo siguiente. El indeterminado o indefinido, para algunos gramáticos, según se recoge en la *NGLE* (RAE y ASALE, 2009), se comporta más bien como los cuantificadores, indefinidos y numerales; pero no hay completo acuerdo sobre si realmente es un artículo, un cuantificador (pronombre o adjetivo) indefinido —en contraste con *algún*— o un cuantificador numeral —en oposición a otros cardinales: *dos, tres...*— (2009, pp. 1088, 1100). Procede del numeral latino *unus/una/unum* y, en el proceso de gramaticalización, se neutralizó el rasgo de *cardinalidad* propio de los numerales y se mantuvo el de *indeterminación o indefinitud*; sin embargo, hay casos dudosos en los que el artículo se asimila a los cardinales (*Devolvió un libro, no dos*) y en los que no hay diferencia entre las dos interpretaciones (*Le entregó un sobre puede ser un sobre y no dos o uno cualquiera*) (2009, p. 1025). El artículo determinado, por su parte, procede del demostrativo latino *ille/illa/illum* —en latín no había artículo— y en el proceso evolutivo mantuvo el rasgo de *definitud* y perdió el *deíctico*. Según la naturaleza semántica, los cuantificadores se clasifican, por un lado, en (1) *fuertes, universales o definidos*: son los que delimitan la denotación del grupo nominal al que acompañan e informan de su referencia, y entre ellos están el artículo determinado, los demostrativos, posesivos, los relativos *cual* y *cuyo*, cuantificadores universales y definidos anafóricos como *uno* u *otro*; y, por otro lado, (2) *cuantificadores débiles o indefinidos*: indican que la predicación que introducen se aplica, en general, a una parte de un conjunto o de una entidad, e incluyen el artículo indeterminado, entre otros cuantificadores, como los existenciales, los numerales cardinales, los evaluativos, los comparativos y los de indistinción (2009, pp.1387-1391).

<sup>5</sup> Santos (2004) completó el proceso de AE incluyendo las aplicaciones didácticas. Su propuesta de análisis es: compilación del corpus de datos, identificación de los errores, descripción de los errores, clasificación de acuerdo con una taxonomía, explicación de los errores, evaluación de los resultados, discusión de los resultados e implicaciones didácticas.

<sup>6</sup> En primer lugar, Vázquez (1986, pp. 158-159) adopta los criterios gramatical, pedagógico, etiológico, comunicativo y descriptivo, y dentro de este último incluye errores por omisión, adición, falsa elección y falsa colocación. Fernández (1990), por otro lado, clasifica los errores en léxicos, discursivos, gráficos y fonéticos, y gramaticales, y dentro de estos analiza los valores y usos de diferentes categorías gramaticales, entre ellas el artículo, cuyos errores se dividen en: uso-omisión y elección entre las formas *el/un*. Por último, Santos (2004, pp. 404-405) aúna modelos anteriores y propone clasificar los errores según los criterios etiológico, comunicativo, pedagógico, pragmático, cultural y descriptivo/lingüístico, y este último incluye errores de adición, omisión, selección falsa, colocación falsa y yuxtaposición.

<sup>7</sup> Para determinar si la diferencia entre las medias de dos grupos es significativa se utiliza la prueba *t* de Student y se aplica cuando la variable independiente tiene dos niveles (en nuestro caso, acierto y error). La diferencia es significativa si la probabilidad (valor *p*) de que se deba al azar es de un 5 % o menor ( $p > 0,05$ ). El valor estadístico de contraste Chi cuadrado ( $\chi^2$ )

permite saber si existe o no relación entre las variables estudiadas cuando inciden dos factores (p. ej., las variantes acierto y error en relación con el nivel de dominio). Solo si el valor  $p$  asociado al estadístico es igual o menor a 0,05 se rechaza la hipótesis nula y se considera significativa la diferencia entre los resultados. Después, para determinar la intensidad de la relación entre las variables se aplica el coeficiente  $V$  de Cramer, que va de 0 (sin asociación) a 1 (máxima asociación).

### Perfil de las autoras

Kazuko Yonekawa

Es licenciada en Humanidades y Filología Inglesa, y cuenta con un máster en Enseñanza del Español como Segunda Lengua (UNED) y otro en Ciencias Empresariales (MBA). Actualmente cursa el doctorado en la UNED, en el programa de la Facultad de Filología: Estudios Lingüísticos y Literarios, Teoría y Aplicaciones. Mantiene varias líneas de investigación, entre las que destaca el análisis de errores y la competencia docente, en concreto, la formación de profesores no nativos. Imparte clases de español en la Universidad Keio SFC y clases de español para fines específicos en la Universidad Kanagawa.

Alicia San Mateo Valdehita

Es Profesora Titular en el Departamento de Lengua Española y Lingüística General de la UNED, donde imparte asignaturas sobre norma y uso de la lengua, sintaxis y enseñanza de español como L2. Durante más de diez años ha enseñado español como L2 en la Universidad de Vanderbilt. Sus principales líneas de investigación son la enseñanza de español, especialmente la del léxico, y la lingüística de corpus. Ha publicado materiales didácticos y artículos en revistas como *JSLT*, *Porta Linguarum*, *SiC* y *Signos*.

### Title

Corpus-based analysis of the errors in the use of articles made by Japanese students of Spanish as L2

### Abstract

The purpose of this research is to identify the correct and incorrect uses of articles at each proficiency level in texts written by 49 L1 Japanese learners of Spanish as a second language (subcorpus of CEDEL2; 5763 words) and to compare the results with similar research on other learners with article-less L1s. The analysis was based on Corder (1971) and each usage was evaluated as correct/incorrect, and if incorrect was then classified as either omission or overuse or the wrong choice of article (Alexopoulou, 2006). Among the 1109 article usages there were 925 correct uses and 184 incorrect uses (an error rate of 16.59%). The error rate decreases as the level of proficiency rises: at A it was 28.33%, at B 21.65%, and at C 9.06%, and this difference between each level was found to be statistically significant. Selection errors were the most frequent (64.13%), followed by omission errors (33.15%) and then addition errors (2.72%). The misuse of the definite article before the first mention of nouns was the most frequent error at each level of proficiency. These results confirm the difficulties speakers of article-less L1 (such as Japanese, Chinese, Slovenian, Serbian and Turkish) have in referencing nouns in Spanish.

### Keywords

Error analysis, learners' corpus, Spanish articles, Japanese learners of Spanish as L2

### タイトル

日本語を母語とする第二外国語としてのスペイン語学習者コーパスにおける冠詞誤用分析

## 要旨

本稿の目的は、日本語を母語とする49名の第二外国語としてのスペイン語学習者の作文コーパス(CEDEL2のサブコーパス、5763語)における冠詞使用について、習熟度レベル別に正用と誤用を特定し、冠詞のない母語を持つ学習者を対象とした類似の先行研究の結果と比較することである。分析はCorder(1971)に従い、冠詞使用を正用・誤用に分別し、欠落、過剰使用、誤選択の記述観点により分類した(Alexopoulou, 2006)。1109の冠詞使用のうち925の正用と184の誤用が認められた(平均誤用率16.59%)。習得レベルが上昇するに伴い誤用率は、Aレベル28.33%、Bレベル21.65%、Cレベル9.06%と低下し、レベル間の誤用率には統計的有意差が認められた。誤選択による誤用の頻度が最も高く(64.13%)、欠落(33.15%)、過剰使用(2.72%)がそれに続いた。初出の名詞に定冠詞を用いる誤選択が、全ての習得レベルにおいて頻度が最も高い誤用であった。日本語、中国語、スロベニア語、セルビア語、トルコ語など冠詞のない母語を持つ学習者が、スペイン語で名詞を指示する際に持つ困難さは共通していることが確認された。

## キーワード

誤用分析、学習者コーパス、スペイン語の冠詞、日本語を母語とするスペイン語学習者

## Lista de tablas

Tabla 1. Nomenclaturas de los niveles de referencia (Gargallo y Alexopoulou, 2021, p. 9)

Tabla 2. Información general del corpus por niveles, según el *MCER*

Tabla 3. Errores en el subcorpus de textos de nivel A según los usos del artículo

Tabla 4. Errores en el subcorpus de textos de nivel B según los usos del artículo

Tabla 5. Errores en el subcorpus de textos de nivel C según los usos del artículo

## Lista de figuras

Figura 1. Ejemplo de la búsqueda con la función KWIC en #LancsBox 6.0

