

# ANÁLISIS ESTILOMÉTRICO DE TRES OBRAS DE ROXANA MIRANDA RUPAILAF

Enrique Nuesch<sup>1</sup>  
Gleice Angélica Queiroz Rodrigues<sup>2</sup>

**Resumen:** El artículo presenta un análisis estilométrico de textos de Roxana Miranda Rupailaf. Se comparan tres obras, *Las tentaciones de Eva* (2003), *Seducción de los venenos* (2008) y *Sumpall* (2011), haciendo medidas 1. La TTR; 2. El porcentaje de “palabras funcionales” con relación a la totalidad de palabras del texto; 3. La cantidad de sílabas métricas por verso; 4. La cantidad de palabras por verso; 5. La longitud de cada periodo. Se concluye que una variación de nivel lexical es perceptible en la obra de 2011 con relación a las dos anteriores, de 2003 y 2008.

**Palabras clave:** Estilometría; poesía; literatura femenina; mapuche.

## STYLOMETRIC ANALYSIS OF THREE WORKS BY ROXANA MIRANDA RUPAILAF

**Abstract:** The paper presents a stylometric analysis of texts by Roxana Miranda Rupailaf. Three works are compared: *Las tentaciones de Eva* (2003), *Seducción de los venenos* (2008) and *Sumpall* (2011). The aspects considered in the analysis are: 1. TTR; 2. Percentage of functional words; 3. Metric syllables per verse; 4. Words per verse; 5. Words per period. It is concluded that there is a perceivable stylistic variation in lexical use comparing the most recent work (2011) to the earlier ones (2003 and 2008).

**Keywords:** Stylometry; poetry; feminine literatura; mapuche.

---

<sup>1</sup> Doutor em Letras, docente da UNESPAR/Apucarana. E-mail: enrique.nuesch@unespar.edu.br

<sup>2</sup> Mestranda em Letras (UEM). E-mail: gleicequeiroz80@gmail.com

## Introducción

Roxana Miranda Rupailaf (1967) es una poetisa chileno-mapuche que se ha destacado en la escena literaria latinoamericana en las pasadas décadas. Como se nota a partir de la consulta a la producción académica acerca de su obra, los temas que más se han asociado a su hacer poético son el género, la decolonialidad, la memoria, el erotismo, el cuerpo y, evidentemente, las resistencia y el rescate de la cultura Mapuche.

Por otro lado, a pesar de que cuestiones relativas a la traducción (CURRIO, 2013; CALDERÓN LE JOLIFF y MORA ORDÓÑEZ, 2014; STOCCO, 2020) se hayan planteado en algunos estudios sobre (o que mencionan a) la autora, y que esta cuestión plantee consideraciones formales, no se encuentran abordajes que consideren su producción literaria en su aspecto más formal y considerando su desarrollo estilístico interno.

Lo que se propone en este artículo es, pues, contribuir con la fortuna crítica acerca de la autora, con un trabajo considerando algunos aspectos del desarrollo de su estilo en términos formales, por medio de un análisis estilométrico de tres de sus obras.

## Materiales y método

El análisis estilométrico también se puede llamar de “textométrico”. Se trata de hacer diversas medidas del material lingüístico dispuesto por el texto (SOARES JÚNIOR y DIÓGO, 2018). En nuestro caso, las medidas que tomaremos serán: 1. La TTR; 2. El porcentaje de “palabras funcionales” con relación a la totalidad de palabras del texto; 3. La cantidad de sílabas métricas por verso; 4. La cantidad de palabras por verso; 5. La longitud de cada periodo. Nuestro *corpus* de análisis es compuesto por las tres obras de Roxana Rupailaf

de que disponemos: *Las tentaciones de Eva* (2003), *Seducción de los venenos* (2008) y *Sumpall* (2011).

¿Por qué es importante hacer estas medidas? Porque por medio de ellas conocemos mejor el estilo de la autora en comparación con ella misma, es decir, conocemos cómo son diferentes entre sí las obras de la autora a lo largo del tiempo según lo que ella eligió consciente e inconscientemente para escribir sus textos. Conscientemente, porque el ser humano elige su tema y busca sus palabras para decir lo que quiere decir, pero también inconscientemente porque hay cosas que no se pueden evitar, como el uso de ciertas palabras que estructuran sintácticamente las oraciones. O sea, el estilo es lo que resulta entre las elecciones conscientes e individuales de un autor y aquello que el sistema lingüístico le impone (CARVALHO, 2004). Así, cuando decimos que haremos un análisis “estilométrico”, quiere decir que mediremos cómo es que lo que resulta de esa tensión se manifiesta en los textos.

En el caso de la “densidad lexical” (TTR), usamos el *Text Analyzer* (ADAMOVIC, 2009). Según su autor, está apto para analizar textos en inglés y en otras lenguas. Como no hace clasificaciones gramaticales (distinguir verbos de adjetivos, por ejemplo), lo que el *Text Analyzer* hace es simplemente distinguir las palabras y oraciones por la presencia de los espacios y los párrafos. Por eso, en la “preparación” de los textos para este programa, se hizo necesario que se quitaran símbolos como la interrogación y la exclamación invertidas, para que expresiones como “que”, “¿qué” “¡que” no fueran contadas como palabras diferentes. Las comillas, comas, puntos, punto y coma y las interrogaciones/exclamaciones normales son identificadas por el programa y no causan confusiones como las de los símbolos antes mencionados. Después que se procesa el texto, el programa produce una lista con la frecuencia de cada una de las palabras que ocurren en el texto

y también otros datos acerca del texto, entre ellos su TTR (*type-token ratio*), lo que quiere decir que se divide la cantidad de palabras diferentes en el texto por la totalidad de palabras y se multiplica por 100. O sea “El hombre es solo un hombre” tiene 6 palabras al todo (*tokens*), pero tiene 5 palabras diferentes (*types*), puesto que “hombre” se usa dos veces (o sea, 2 *tokens* para “hombre”). Entonces la TTR de esa oración es  $5/6 \cdot 100 = 83,33$ .

Para el porcentaje de “palabras funcionales”, usamos la propia lista generada por el *Text Analyzer* y la comparamos con el listado de “palabras funcionales del español” propuesto por Hallebeek (1986), y fuimos marcando en aquella las que estaban presentes en este. Después, calculamos cuál es el porcentaje de “palabras funcionales” con relación al total de palabras de cada uno de los textos del *corpus*.

Para obtener las sílabas métricas por verso, utilizamos el programa *AOIDÓS* (NUPILL, 2016), desarrollado por el Núcleo de Pesquisas em Informática, Literatura e Linguística de la Universidad Federal de Santa Catarina. Este programa es capaz de hacer la separación de sílabas poéticas en textos de lengua portuguesa y española, incluso identificando figuras de nivel fónico, como sinalefa, dialefa, sinéresis y diéresis. En el caso de la “preparación” del texto para ser procesado en este programa, solamente fue necesario salvarlo en archivo “txt”, el cual es “subido” al sitio y procesado *online*. Cuando el procesamiento termina, tenemos una lista clasificando todos los versos por cantidad de sílabas métricas, o sea, por ejemplo, diez versos de 4 sílabas, cinco versos de 7 sílabas etc.

Para saber la cantidad de palabras por verso, nosotros mismos tomamos los textos en verso, usando el *Microsoft Word* transformamos cada una de las palabras en un número 1, y usando el *Microsoft Excel* sumamos estos números 1 de cada línea, lo que nos daba el total de palabras por línea, o sea, por verso. Por ejemplo:

Verso 1: “Yo soy el pez” > 1 1 1 1 > 4 palabras.

Verso 2: “que estaba en el agua” > 1 1 1 1 1 > 5 palabras.

## Datos obtenidos

Como dicho arriba, el *corpus* es compuesto por tres obras de momentos diferentes: *Las tentaciones de Eva* (2003), *Seducción de los venenos* (2008) y *Sumpall* (2011). Cada una de ellas será etiquetada, respectivamente, para TENTA, SEDUC y SUMP.

## La TTR

Lado a lado, las TTR de cada texto del *corpus* figuran así:

**Tab. 1** – TTR de los textos

Medida/Texto	TENTA	SEDUC	SUMP
Type	782	959	785
Token	1904	2641	2578
TTR (Ty/To*100)	41,07	36,31	30,45

Fuente: Los autores

## El porcentaje de palabras funcionales

Lado a lado, los porcentajes de palabras funcionales de cada texto del *corpus* figuran así:

**Tabla 2** – Porcentaje de “palabras funcionales” para palabras totales de los textos

Medida/Texto	TENTA	SEDUC	SUMP
Palabras funcionales	1034	1438	1524
Palabras totales	1904	2641	2578
%Pfun/Ptot	54,30	54,44	59,11

Fuente: Los autores

## La cantidad de sílabas métricas por verso

Los versos de cada texto variaron en cantidad de sílabas métricas entre 1 y 22. Esto no quiere

decir que en los tres textos todas estas cantidades están presentes. Por ejemplo, TENTA no tenía versos con más que 17 sílabas, mientras que SEDUC y SUMP sí las tenían. Abajo, ponemos lado a lado las medidas para cada texto. “Tipo” es la cantidad de sílabas por verso, “Frec” es la frecuencia en la que aparece cada tipo.

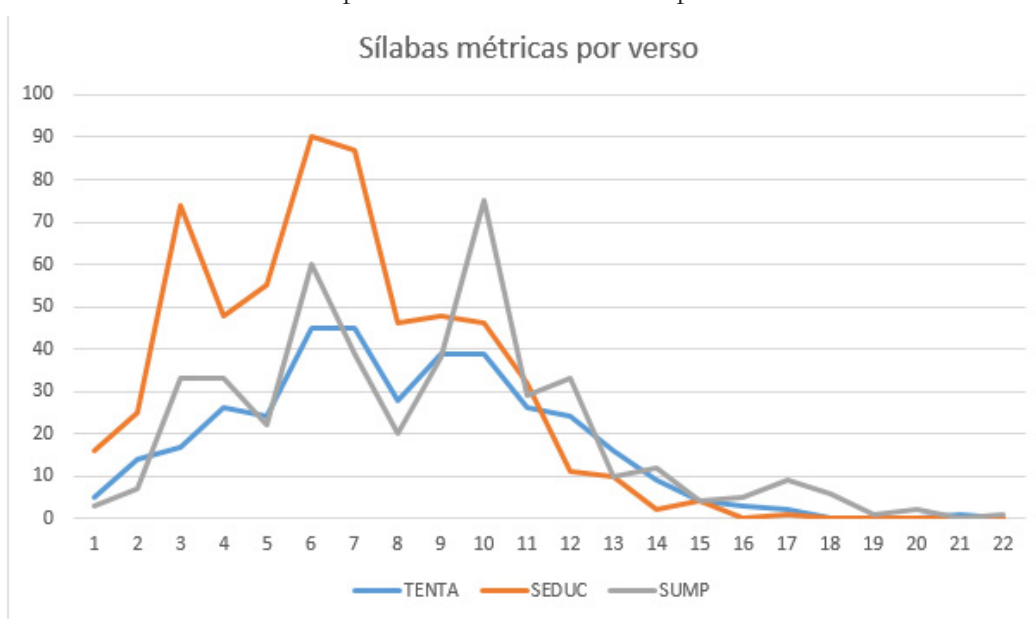
**Tabla 3** – Frecuencia de los tipos de verso considerados por su cantidad de sílabas métricas

Tipo/Frec	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Total
TENTA	5	14	17	26	24	45	45	28	39	39	26	24	16	9	4	3	2	0	0	0	1	0	367
SEDUC	16	25	74	48	55	90	87	46	48	46	32	11	10	2	4	0	1	0	0	0	0	0	595
SUMP	3	7	33	33	22	60	39	20	38	75	29	33	10	12	4	5	9	6	1	2	0	1	442

Fuente: Los autores

Puestos en un gráfico, estos datos se ven así:

**Gráfico 1** – Frecuencia de los tipos de verso considerados por su cantidad de sílabas métricas



Fuente: Los autores

## La cantidad de palabras por verso

Los versos de cada texto variaron en cantidad de palabras por verso entre 1 y 15. Así como en la medida anterior, esto no quiere decir que los tres textos tuvieron todos los tipos de cantidad de palabras por verso. Por ejemplo, TENTA y SUMP tuvieron un verso de 13 palabras cada uno, mientras que SEDUC no tuvo versos con esta cantidad de palabras. Abajo, ponemos lado a lado las medidas de cada texto. “Tipo” es la cantidad de palabras por verso y “Frec” es la frecuencia en que aparece cada tipo.

**Tabla 4** – Frecuencia de tipos de verso considerados por cantidad de palabras

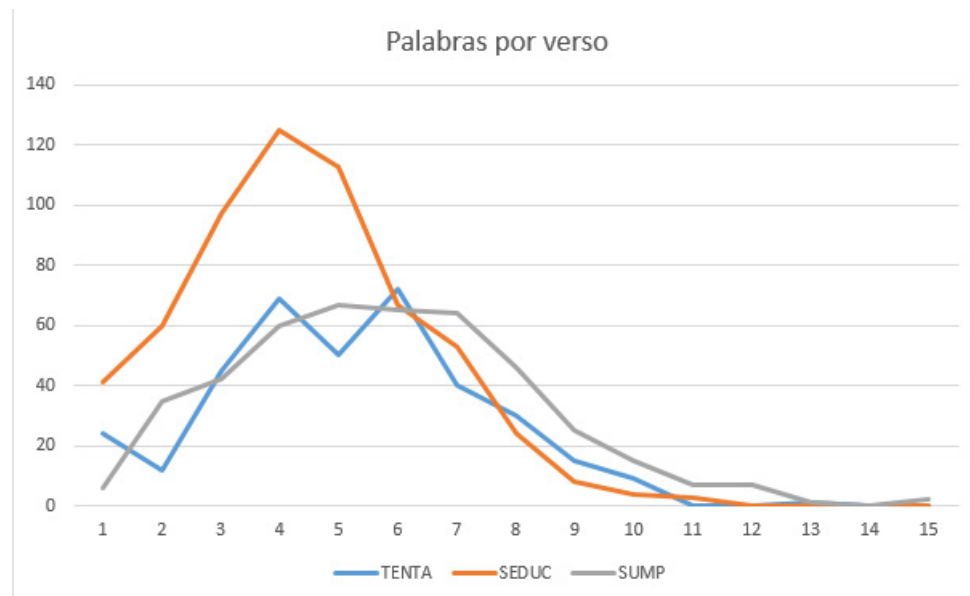
Tipo/Frec	TENTA	SEDUC	SUMP
1	24	41	6
2	12	60	35
3	45	97	42
4	69	125	60
5	50	113	67
6	72	67	65
7	40	53	64
8	30	24	46
9	15	8	25
10	9	4	15
11	0	3	7

12	0	0	7
13	1	0	1
14	0	0	0
15	0	0	2
Total	367	595	442

Fuente: Los autores

Puestos en un gráfico, estos datos se ven así:

**Gráfico 2 – Frecuencia de palabras por verso para cada texto**



Fuente: Los autores

## Discusión de los datos

A seguir se discuten los datos recién presentados y se indican qué pueden significar sobre el estilo de la autora.

### TTR y “Palabras funcionales”

Sobre las TTR, los datos nos muestran que fueron disminuyendo de una obra a la otra a lo largo de los años: TENTA – 2003 = 41,07; SEDUC – 2008 = 36,31; SUMP – 2011 = 30,45. En un principio, esto quiere decir que de una obra a otra hubo menos variedad lexical, o sea, la proporción entre palabras repetidas y palabras que no se repiten dentro de cada texto fue mayor.

**Tabla 5 – TTR de los textos**

Medida/Texto	TENTA	SEDUC	SUMP
Type	782	959	785
Token	1904	2641	2578
TTR (Ty/To*100)	41,07	36,31	30,45

Fuente: Los autores

Si se compara TENTA con SEDUC y TENTA con SUMP, hace sentido que las TTR sean bastante diferentes entre las dos comparaciones, porque según los investigadores que usan TTR, “quanto maior for um texto, menor será a TTR” (MARTINS, 2016). En términos de estilo, pues, se podría pensar que dentro del universo de la autora (pues no estamos comparando las obras de esta autora con la de otros autores), de una obra a otra la variedad del léxico empleado fue menor. Se debe mirar entonces la proporción de “palabras funcionales” dentro de cada texto. Las “palabras funcionales” son las que tienen la tendencia de repetirse más, pues son las palabras que cumplen las funciones sintácticas que estructuran las oraciones. Y ahí se ve que hubo un aumento de dicha proporción:

**Tabla 6 –** Porcentaje de “palabras funcionales” para palabras totales de los textos.

Medida/Texto	TENTA	SEDUC	SUMP
Palabras funcionales	1034	1438	1524
Palabras totales	1904	2641	2578
%Pfun/Ptot	54,30	54,44	59,11

Fuente: Los autores

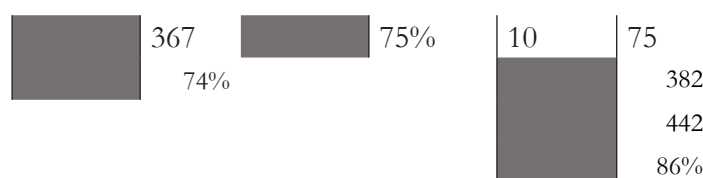
O sea, tanto las TTR como el porcentaje de “palabras funcionales” indican que hubo menor variedad lexical en SUMP que en TENTA y SEDUC. Y esto porque 1) SEDUC es un texto mayor que SUMP y sin embargo su TTR es también mayor; 2) SEDUC es un texto mayor que SUMP y sin embargo su porcentaje de “palabras funcionales” es menor.

## Sílabas métricas por verso

Si se miran la tabla y el gráfico de sílabas por verso, se nota que SUMP es el texto que más tipos de versos diferentes tiene, variando entre 1 y 22, mientras que TENTA varía entre 1 y 17 (con solamente 1 ocurrencia por encima de 17, que es de un verso con 21 sílabas) y SEDUC varía entre 1 y 15 (con solamente 1 ocurrencia por encima de 15, que es de un verso con 17 sílabas). Si separamos para cada texto los tipos de versos más frecuentes por encima de la mediana (o sea, el punto medio en las frecuencias organizadas de menos a más), tenemos la siguiente tabla:

**Tabla 7 -** Porcentaje del total de versos concentrados en los tipos más frecuentes de acuerdo con su cantidad de sílabas métricas

TENTA		SEDUC		SUMP	
Tipo	freq	tipo	Freq	tipo	freq
12	24	10	46	8	20
4	26	4	48	5	22
11	26	9	48	11	29
8	28	5	55	3	33
9	39	3	74	4	33
10	39	7	87	12	33
6	45	6	90	9	38
7	45		448	7	39
	272		595	6	60



Fuente: Los autores

Se nota, entonces que SUMP concentra 86% de sus versos en los tipos más frecuentes, mientras que SEDUC concentra 75% y TENTA concentra 74%. Por otro lado, vemos que para los tres textos, los tres tipos de versos más frecuentes incluyen los que tienen 6 y 7 sílabas métricas, que son superadas por el de 10 solamente en SUMP. En este aspecto, se nota una constancia en el uso más frecuente de versos con sílabas de 6 y 7 sílabas métricas en los tres textos.

## Palabras por verso

Mirando los datos de esta medida, vemos que de nuevo SUMP es el texto más variado, con versos variando entre 1 y 15 palabras, sin registro solamente en el tipo de 14 palabras. A su vez SEDUC varía entre 1 y 11 palabras, mientras que TENTA varía entre 1 y 10 palabras por verso, con solo 1 registro por encima de 10, en el tipo de 13 palabras por verso.

Si hacemos el mismo procedimiento que con la medida anterior, mirando las cantidades de palabras por verso de la mitad con más frecuencias (o sea, la porción por encima de la mediana), podemos organizar los datos de esta medida así:

**Tabla 8** - Porcentaje de concentración de los versos más frecuentes de acuerdo con su cantidad palabras

TENTA		SEDUC		SUMP	
Tipo	frec	tipo	frec	tipo	frec
7	40	2	60	3	42
3	45	6	67	8	46
5	50	3	97	4	60
4	69	5	113	7	64
6	72	4	125	6	65
	276		462	5	67
	367		595		344
	75%		77%		77%

Fuente: Los autores

Se ve que en el caso de las palabras por verso hay una uniformidad en los tres textos. Los tipos de verso (por cantidad de palabras) más frecuentes ocupan un 75% del total en TENTA, 77% en SEDUC y 77% en SUMP. O sea, esta es una medida que se presenta constante en los tres textos. Igualmente, se puede notar que entre los tres tipos más frecuentes están los versos de 5 palabras, de 5 y 4 en TENTA y SEDUC, y de 5 y 6 palabras en TENTA y SUMP.

## Interpretación de los datos

En términos de variedad lexical, SUMP se muestra el menos variado. Si, como recordamos que dicen los investigadores que usan TTR, los textos más cortos tienden a tener TTR más alto, entonces podemos hacer unas comparaciones de pares de obras. La diferencia en número de palabras entre SUMP y TENTA muestra que el primero tiene +674 palabras y -10,62 en TRR. La diferencia de número de palabras entre SEDUC y TENTA muestra que el primero tiene +737 palabras y -4,76 en TTR. O sea, la diferencia entre el número total de palabras entre SEDUC y SUMP es de solo 63 palabras a más para SEDUC, pero ellas son responsables por una distancia de TTR de 4,76 para SEDUC-TENTA y de 10,62 para SUMP-TENTA. Por su vez, SEDUC y SUMP son los textos más próximos en número de palabras, las 63 mencionadas, y esto causa una diferencia de +5,86 para la TTR de SEDUC sobre la TTR de SUMP.

**Tabla 9** - Variación de número de palabras, TTR y % de palabras funcionales entre pares de obras

	v.Pal	v.TTR	v.Pfunc/Ptot
SEDUC-TENTA	+737	-4,76	-0,14%
SUMP-TENTA	+674	-10,62	-4,81
SEDUC-SUMP	+63	+5,86	-4,67%

Fuente: Los autores

En la tabla también incluimos la variación de palabras funcionales para mostrar que la misma relación entre los pares de obras se repiten. A pesar de que haya solo 63 palabras de distancia entre SEDUC y SUMP, la segunda se aleja mucho más de TENTA que la primera.

Ahora, sobre la variedad de versos de acuerdo con su cantidad de sílabas métricas, SUMP es el más variado, con presencia en casi todos los tipos entre 1 y 22, mientras que TENTA y SEDUC tienen poca frecuencia más allá que las 15 sílabas. O sea, en su obra más reciente, la autora ha dado más

espacio a versos más largos que en las anteriores. Por otro lado, en SUMP los versos se concentran con más intensidad en un conjunto menor de metros, mientras que en TENTA y SEDUC hay menos concentración en el mismo conjunto de metros.

Por fin, sobre la cantidad de palabras por verso, se observa que es poco variable entre las tres obras.

## Conclusión

Se puede decir que, en lo que dice respecto a las características textuales aquí medidas, hay una variación de estilo que separa la obra más reciente, *Sumpall* (2011) de las dos menos recientes, *Las Tentaciones de Eva* (2003) y *Seducción de los venenos* (2008), que se manifiesta principalmente en el léxico y los tipos de verso de acuerdo a la cantidad de sílabas métricas. Estas medidas nada tienen que ver con la “calidad” de las obras en sí mismas, son solamente características textuales. Cuando se dice que *Sumpall*, por ejemplo tiene menos “variedad lexical”, esto nada tiene que ver con “calidad”. Se debe recordar que hay figuras de lenguaje que usan la repetición de palabras como recurso (por ejemplo, la anáfora, o el retruécano). O sea, todo esto se trata más de conocer las características textuales de una o varias obras, y no de hacer juicios sobre “calidad” o aun mismo “creatividad”. Otras medidas en este sentido debieran ser hechas, como las frecuencias relativas de figuras de lenguaje, por ejemplo, pero ello entraría en consideraciones semánticas, lo que no estaba dentro de los objetivos de esta breve investigación.

## Referencias

ADAMOVIC, Mladen. Text Analyzer. 2009. Disponible em: <https://www.online-utility.org/text/analyzer.jsp>. Acessado em 20 jan. 2021.



- CALDERÓN LE JOLIFF, Tatiana; MORA ORDÓÑEZ, Edith. (Orgs.). *Afpunmapu Fronteras Borderlands. Poética de los confines: Chile-México*. Valparaíso: Ediciones Universitarias de Valparaíso, 2014.
- CARVALHO, C. de. A estilística e o ensino de português. *Cadernos do CNLF*, v. VIII, n. 12, Rio de Janeiro, 2004. Disponível em: <http://www.filologia.org.br/viiiicnlf/anais/caderno12-02.html>. Acesso em: 31 mai 2021.
- CURRIO, Maribel. Poesía mapuche: la instalación de una mismidad étnica en la literatura chilena. *A Contracorriente*, Raleigh, v. 10, n. 3, p. 21-53, 2013. Disponible en <https://dialnet.unirioja.es/servlet/articulo?codigo=5401200>. Aceso en 10 jan. 2021.
- HALLEBEEK, Jos. Las palabras funcionales del español”, *Boletín de la AEPE*, Santander, n. 34-35, p. 205-216, 1986. Disponível em: [https://cvc.cervantes.es/ensenanza/biblioteca\\_ele/aepe/boletin\\_34-35\\_18\\_86.htm](https://cvc.cervantes.es/ensenanza/biblioteca_ele/aepe/boletin_34-35_18_86.htm). Acesso em: 10 jan. 2021.
- MARTINS, Mário. A diversidade lexical na escrita de textos escolares. *Fórum Linguístico*, Florianópolis, v.13, n.1, p.1068-1082, 2016. Disponível em: <https://doi.org/10.5007/1984-8412.2016v13n1p1068>. Aceso em: 15 jan. 2021.
- MIRANDA RUPAILAF, Roxana. *Las tentaciones de Eva*. Santiago: Ediciones Príncipe, 2003.
- \_\_\_\_\_. *Seducción de los venenos*. Santiago: LOM, 2008
- \_\_\_\_\_. *Sumpall*. Santiago: Del Aire Editores, 2011.
- NÚCLEO DE PESQUISAS EM INFORMÁTICA, LITERATURA E LINGUÍSTICA. *Aoidós*. Disponível em: <https://aoidos.ufsc.br/>. Acesso em: 20 fev. 2020.
- SOARES JÚNIOR, Ademar Pereira; DIÓGO, Nathália Silva. Pronomes, artigos e outras palavras funcionais no discurso dos gêneros de “Grande Sertão: Veredas” (crítica algorítmica). *Mafuá*, Florianópolis, n. 29, 2018. Disponível em: [https://mafua.ufsc.br/2018/pronomes-artigos-e-outras-](https://mafua.ufsc.br/2018/pronomes-artigos-e-outras-palavras-funcionais-no-discurso-dos-generos-de-grande-sertao-veredas-critica-algoritmica/)
- palavras-funcionais-no-discurso-dos-generos-de-grande-sertao-veredas-critica-algoritmica/. Acesso em: 13 fev. 2021.
- STOCCO, Melisa. Traducción colaborativa en la poesía mapuche: un aporte a la revitalización del mapuzungun. *TRANS. Revista de Traductología*, Málaga, n. 24, p. 469-481, 2020. Disponível em: <https://revistas.uma.es/index.php/trans/article/view/6407>. Acesso em: 23 mar. 2021.

**Submissão: julho de 2021.**

**Accite: agosto de 2021.**