

Modelación basada en agentes en el bachillerato: un estudio de las experiencias cotidianas y conceptos científicos

Gustavo Carreón Vázquez ¹ 

J. Enrique Hernández Zavaleta ² 

Luis Enrique Cortés Berrueco ³

Vicente Carrión Velázquez ⁴ 

Resumen

Las experiencias de aprendizaje que incluyen las vivencias de los estudiantes sobre la propagación de epidemias para construir conocimientos son limitadas. Este estudio analizó las frases cotidianas y científicas de estudiantes de bachillerato al trabajar con un simulador computacional basado en agentes sobre epidemias. Los datos incluyen videgrabaciones, sus transcripciones y producciones escritas. Utilizando minería de datos, se crearon nubes de palabras y se propuso un valor de entropía para las frases de los estudiantes respecto a las palabras que más utilizaban. Los resultados se dividen en dos: (1) el análisis de frecuencias de palabras, que mostró un incremento en el uso de palabras técnicas y científicas al avanzar las sesiones y (2) el análisis de la entropía que mostró que las explicaciones de los estudiantes contenían frases que combinaban sus experiencias cotidianas con conceptos científicos. Esta investigación aporta una metodología con una perspectiva mixta de análisis de frases que enriquece la comprensión del aprendizaje en la modelación de fenómenos complejos.

Palabras clave

Bachillerato, modelación basada en agentes, experiencias cotidianas, simulación computacional, minería de datos

¹ gcarreon@unam.mx
Instituto de Investigaciones Económicas, UNAM

² enrique_hernandez@cbu.ca
Cape Breton University, Nueva Escocia, Canadá

³ lecortesb@gmail.com
Facultad de Ingeniería, Universidad La Salle México

⁴ vcarrion@upn.mx
Universidad Pedagógica Nacional

Carreón Vázquez G., Hernández Zavaleta, J. E., Cortés Berrueco, L. E. & Carrión Velázquez, V. (2025). Modelación basada en agentes en el bachillerato: un estudio de las experiencias cotidianas y conceptos científicos. En A. Solares-Rojas, & A. P. Preciado Babb (Eds.), *La investigación en modelización matemática: un diálogo entre educadores de Latinoamérica y España* (pp. 205–229). Editorial SOMIDEM.
<https://doi.org/10.24844/SOMIDEM/S2/2025/01-09>

Abstract

Learning experiences that build knowledge on students' experiences of the spread of epidemics are limited. This study analyzed the everyday and scientific phrases of high school students when working with an agent-based epidemic's computer simulator. The data includes video recordings, transcriptions, and written productions. Using data mining, we created word clouds and proposed an entropy value for the students' phrases regarding the words they used most. The results are divided into two: (1) the word frequency analysis that showed an increase in the use of technical and scientific words as the sessions progressed and (2) the entropy analysis that showed that the students' explanations contained phrases that combined their everyday experiences with scientific concepts. This research provides a methodology with a mixed perspective of phrases analysis that enriches the understanding of learning in the modeling of complex phenomena.

Keywords

High school, agent based modeling, everyday experiences, computational simulation, data mining

Introducción

El brote y la propagación de enfermedades han sido un tema que ha interesado a investigadores y educadores en matemática educativa desde antes de la pandemia de COVID-19 (por ejemplo, Gaff et al., 2011; Gibbons et al., 2015). Estas investigaciones resaltan la importancia de interpretar gráficas y encontrar las variables que contribuyen al modelo matemático algebraico o de ecuaciones diferenciales que describen y predicen el fenómeno. La reciente pandemia revivió estos esfuerzos y sumó otros en los que investigadores y educadores matemáticos se preguntaban cómo crear conciencia sobre este fenómeno entre sus estudiantes (por ejemplo, Cantoral et al., 2020; Costa & Domingos, 2022), existiendo el acuerdo tácito de que comprender la forma en que una enfermedad se transmite incide positivamente en el desarrollo de un pensamiento crítico, respaldado por epistemologías de la matemática y las ciencias.

A pesar de este contexto, se ha prestado poca atención a las experiencias de aprendizaje en las que las vivencias de los estudiantes en términos de contagios y propagación de epidemias sean un medio para construir conocimientos científicos y habilidades críticas sobre este fenómeno. El objetivo de este capítulo es analizar las producciones lingüísticas que estudiantes de segundo año de bachillerato construyeron mientras usaban un simulador computacional, desarrollado en el ambiente de programación basado en agentes NetLogo (Wilensky, 1999), enfocado en la propagación de epidemias. Con la ayuda de un análisis de la entropía (Sivakumar et al., 2016; Wong & Yip, 2020; Zaki & Meira, 2014), nos enfocamos en encontrar patrones en la estructura de las frases expresadas por los estudiantes. De esta forma, identi-

ficamos aquellos momentos en los que los estudiantes establecieron conexiones con sus experiencias cotidianas para dar sentido a los niveles de organización (local y global) intrínsecos a la simulación computacional del fenómeno de epidemias.

En la modelación de sistemas complejos, como la propagación de una epidemia, intervienen múltiples agentes, en donde los comportamientos de la transmisión de enfermedades provienen de las interacciones locales entre ellos (Sengupta et al., 2021). En este estudio utilizamos la modelación basada en agentes y el software NetLogo para generar una actividad sobre la propagación de una epidemia que se trabajó con un grupo de quince estudiantes de segundo año de bachillerato. En las siguientes secciones describimos perspectivas sobre las experiencias fuera de la escuela y en lo científico, la modelación basada en agentes y la propuesta metodológica del trabajo.

Antecedentes

Perspectivas sobre las experiencias fuera de la escuela y los conceptos científicos

Diversos estudios respaldan la idea de que las experiencias de las y los estudiantes fuera del contexto escolar desempeñan un papel fundamental en la construcción de significados en el aula (Anhalt et al., 2018; Arrieta & Díaz, 2015; Carraher et al., 1985; Hernández-Zavaleta et al., 2023; Vygotsky, 1986; Warren et al., 2001). Estas investigaciones reconocen que las prácticas cotidianas influyen en la manera en que el estudiantado comprende y aplica conceptos científicos en la escuela. En la investigación educativa, se ha abordado la relación entre lo científico y lo cotidiano como una dicotomía disociable en donde son opuestos que difícilmente se encuentran en el proceso de aprendizaje. Esta aproximación se caracteriza por asumir cualidades opuestas de cada término. Por ejemplo, “improvisación, ambigüedad, informalidad y subjetividad” denotan lo cotidiano, mientras que lo científico se denota comúnmente como una actividad caracterizada por la “racionalidad, precisión, formalidad y objetividad” (Warren et al., 2001, p. 530). Sin embargo, existen propuestas como la de Vygotsky (1986) en la que se hace una distinción entre conceptos *científicos* y *cotidianos*, en su naturaleza y en su forma de desarrollo, pero se enfatiza que evolucionan y operan juntos en un sistema interfuncional. Como consecuencia, no pueden ser tratados de forma independiente, entendiendo que en las diferentes etapas del aprendizaje surgen patrones coordinados de ambos (Hernández-Zavaleta et al., 2023). Por su parte, Warren et al. (2001) proponen que las experiencias cotidianas no solo sirvan como contexto para comprender el fenómeno, sino también como una oportunidad para percibir y abordar dicho fenómeno de manera diferente. De esta forma, el poder de las experiencias cotidianas y el uso de

lenguaje informal desempeñan un papel fundamental en la creación de significado, al referirnos a la comprensión de un fenómeno y los conceptos científicos inherentes a él.

La modelación basada en agentes en Educación Matemática

La modelación basada en agentes (MBA) es una metodología ampliamente utilizada en el estudio de sistemas complejos (Epstein, 2006; Resnick, 1994; Wilensky & Rand, 2015). Estos modelos se conforman por colecciones de agentes que interactúan con el entorno a través del tiempo desde un enfoque *de abajo hacia arriba*, es decir, se comienza por definir los comportamientos locales y su interacción con el entorno para posteriormente definir formas de análisis de su estructura global.

En educación matemática y en el aprendizaje de las ciencias, la MBA ha mostrado ser un ambiente propicio para la elaboración y comprobación de conjeturas, ofreciendo, al mismo tiempo, la posibilidad de comprender conceptos disciplinares específicos (Sengupta et al., 2013; Wilensky & Reisman, 2006). Su relación con la asignatura de Matemáticas se focaliza en las oportunidades que ofrece la MBA: (1) la apropiación de conceptos de probabilidad y estadística (Abrahamson et al. 2006; Abrahamson & Wilensky, 2005); (2) las descripciones matemáticas que se desarrollan con la programación en paralelo de múltiples agentes (Resnick, 1994); (3) el proceso de modelización matemática inmerso en el entendimiento de sistemas complejos ecológicos, sociales, entre otros (Sengupta & Wilensky, 2011; Wagh & Wilensky, 2012); y (4) la creación y descripción de representaciones de un sistema complejo; por ejemplo, la gráfica de crecimiento de poblaciones en un sistema presa-depredador (Clark & Sengupta, 2019).

Varios autores han sugerido algunas características y propiedades de estos sistemas que pueden ser estudiados para promover aprendizajes matemáticos (Brady & Yarnes, 2020; Levy & Wilensky, 2008; Resnick, 1994; Weintrop et al., 2016). La característica que se destaca es la descentralización, mientras que las propiedades son la de emergencia y aleatoriedad. Resnick (1994) sugiere que la descentralización es inherente a los sistemas en los que no existen líderes que influyan en los comportamientos de colectivos de agentes (por ejemplo, en una parvada no hay un pájaro guía). Como forma de estudio, sugieren introducir niveles de organización (Levy & Wilensky, 2008; Weintrop et al., 2016) que se refieran a que los sistemas pueden ser entendidos desde un nivel microscópico (local) hasta uno macroscópico (global), considerando al sistema como un todo. Las interacciones de los componentes en el nivel microscópico generan patrones emergentes en un nivel macroscópico (por ejemplo, el comportamiento en parvada es un nivel de organización y el comportamiento individual de cada pájaro es otro nivel). La propiedad de emergencia se relaciona con la aparición de patrones de comportamiento en

el sistema, producto de la interacción colectiva entre elementos del sistema. Levy y Wilensky (2008) y Sengupta et al. (2021) mencionan que estos patrones son frecuentemente contraintuitivos e inesperados debido a la aleatoriedad intrínseca a este tipo de sistemas.

Por otro lado, estas propiedades y características están entrelazadas con el proceso de modelización inherente al entendimiento de los sistemas complejos. De acuerdo con Colella et al. (2001), existen dos formas de generar entendimientos en un sistema descentralizado: 1) la creación de entendimientos sobre reglas y elección de variables para los agentes que componen el sistema en un nivel microscópico (de abajo hacia arriba) y 2) observar el comportamiento del colectivo de agentes para elaborar supuestos sobre las reglas y los estados basados en la observación (de arriba hacia abajo).

Las preguntas de investigación que plantea este trabajo son:

1. ¿Cuáles son las palabras y frases representativas que utilizan los estudiantes al describir los comportamientos del fenómeno de epidemias a través del simulador de NetLogo?
2. ¿De qué manera se entrelazan el lenguaje cotidiano con los conceptos científicos de los estudiantes a medida que usan el simulador del fenómeno de epidemias a través del software NetLogo?

En las siguientes secciones se describe el diseño metodológico de las actividades mostradas a los estudiantes, considerando abordar, en primera instancia, la escala local y, posteriormente, abarcar la escala global con la ayuda de diferentes componentes del simulador.

Metodología

Datos demográficos

Este estudio se llevó a cabo con quince estudiantes de segundo año de bachillerato de un Colegio de Ciencias y Humanidades de la Universidad Nacional Autónoma de México, en el contexto de la clase de Matemáticas. Las sesiones se llevaron a cabo a través de la plataforma Zoom en abril de 2021, periodo en el que México se encontraba en situación de pandemia y las clases se llevaban a cabo vía remota. Los datos comprenden la grabación de cuatro sesiones vía Zoom, de aproximadamente 2 horas cada una, en las que los estudiantes interactuaron con el profesor y compartían sus pantallas. El trabajo de los estudiantes fue orientado a través de cuatro formas de Google donde reportaban sus respuestas.

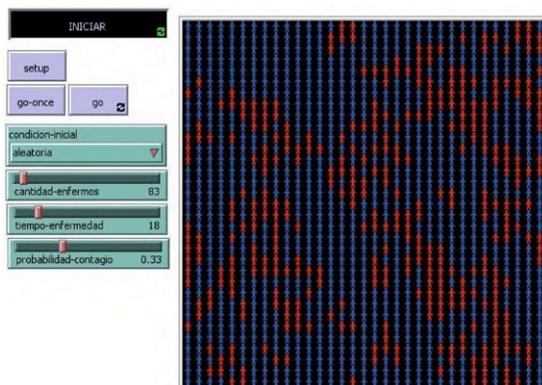
El simulador del fenómeno de epidemias

El simulador del fenómeno de epidemias en NetLogo fue desarrollado por los Autores como parte de un conjunto de actividades del proyecto Conahcyt: *Construcción de significados en procesos de modelación matemática*.

El modelo basado en agentes se fundamenta en el modelo Sano-Infectado-Recuperado con la posibilidad de infectarse nuevamente. La interfaz básica incluyó los botones *setup* (necesario para iniciar la simulación), *go-once* (para ejecutar la simulación por una unidad de tiempo) y *go* (para ejecutar la simulación por múltiples unidades de tiempo). Además, se mostró un menú de condiciones iniciales y los controles de cambio de parámetros *cantidad-enfermos*, el cual establece el número de enfermos iniciales; *tiempo-enfermedad*, el tiempo que transcurre enfermo y *probabilidad-contagio*, el cual es un mecanismo de contagio de los agentes sanos en función de los agentes enfermos en su vecindad. Los agentes azules se encuentran en estado sano, y los rojos en estado infectado. (ver Figura 1).

Figura 1

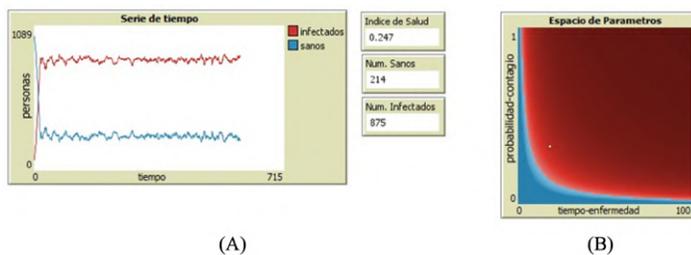
Interfaz básica del simulador



En las sesiones se incluyó la representación gráfica de las series de tiempo de agentes infectados y sanos, además de mostrar en una caja de diálogo su número en un tiempo determinado. Cada punto en el *espacio de parámetros* es un par coordenado (tiempo-enfermedad, probabilidad-contagio), el color azul muestra todos los valores para los que la enfermedad se extingue y el rojo la combinación de valores para los que la enfermedad prevalece (ver Figura 2).

Figura 2

(A) Gráfica "Serie de tiempo" y (B) Espacio de parámetros



Existen casos de comportamientos interesantes en la zona que se marca como una hipérbola, por ejemplo, la enfermedad puede oscilar entre estados de sanos y enfermos indefinidamente. Finalmente, el índice de salud es una medida entre cero y uno que indica la salud de la población de agentes. Si converge a uno indica que la enfermedad se ha erradicado, mientras que cero indica que la población total está infectada.

Descripción de las actividades

Las actividades con las que trabajaron los estudiantes se distribuyeron en cuatro sesiones: *Corporización*, *Conociendo la interfaz del simulador*, *Exploración de series de tiempo* e *Identificación del índice de salud*. La intención de la secuencia fue que los estudiantes partan de una actividad corpórea que los introdujera a la simulación desde una perspectiva local (desde lo individual), y avanzaran hacia un nivel de organización global que les permitiera dar sentido al comportamiento colectivo del fenómeno (Colella et al., 2001).

En la sesión 1, *Corporización*, los estudiantes analizaron videos en los que se les presentó un salón de clases donde veinticinco estudiantes juegan el *juego del contagio*, utilizando como reglas los juegos de piedra, papel y tijeras y volados¹. En el video se inicia el juego con diferentes posiciones de estudiantes en el salón, en los que uno o más empiezan el juego en estado enfermo. Las personas que perdían el juego contra un contagiado eran contagiadas. Los enfermos utilizaban tapabocas para diferenciarse de los sanos. Esta sesión no se incluyó en el análisis, ya que este capítulo se enfoca en el uso del simulador computacional.

La sesión 2, *Conociendo la interfaz del simulador*, sirvió de introducción a los controles y representaciones de la interfaz del simulador. La actividad 1 se focalizó en indagar sobre las funciones de los botones, *setup*, *go-once* y *go*; y en la actividad 2 se les pidió explorar cómo el uso de diferentes condiciones iniciales afectaba el comportamiento del mundo de agentes.

La sesión 3, *Exploración de series de tiempo*, inició con la actividad 3, la cual se centró en cambiar valores específicos de los parámetros de cantidad-enfermos y *tiempo-enfermedad*; en la actividad 4 y 5 exploraron la gráfica de la serie de tiempo de cantidad de sanos y enfermos al variar de manera sistemática los parámetros de *tiempo-enfermedad* y *probabilidad-contagio*. Esto permitió a los estudiantes analizar la evolución de una epidemia utilizando una representación gráfica. La actividad 6 solo fue explicada por el profesor y no hubo interacción con los alumnos, por lo que no se incluye en el análisis.

¹ El juego de volados, o también conocido como cara o cruz, consiste en tirar una moneda al aire y alguno de los dos jugadores escoger uno de los lados de la moneda, si al caer al suelo resulta el lado escogido gana el jugador.

La sesión 4, *Identificación del índice de salud*, constó de la actividad 7 y 8, que tuvieron como objetivo la exploración del índice de salud y el espacio de parámetros respectivamente. Por un lado, el índice de salud les permitió explicar de manera global la salud de toda la población de agentes; por otro, la representación visual del espacio de parámetros les permitió establecer relaciones entre el tiempo de enfermedad y la probabilidad de contagio para construir explicaciones sobre todos los comportamientos posibles en el sistema.

Análisis desde la minería de datos

La minería de datos, y particularmente el análisis de entropía en educación, se ha utilizado como un medio para medir y predecir el logro académico o las tasas de abandono de los estudiantes (Díaz Acosta et al., 2022; Kovalenko et al., 2022; Sivakumar et al., 2016; Sokkhey et al., 2020; Wong & Yip, 2020). En otros ámbitos de la ciencia, esta medida de entropía es utilizada ampliamente para medir la redundancia y correlación de la información en un sistema dado (Cover & Thomas, 2006; Gray, 2011), además, se utiliza en el área de reconocimiento de patrones (Bishop, 2006).

En este trabajo se consideran a las frases como conjuntos de palabras que representan la unidad mínima de análisis para comprender cómo los individuos relacionan sus experiencias cotidianas con los conceptos científicos dentro del contexto de las actividades escolares que están enfrentando.

La metodología propuesta en este trabajo emplea la técnica de nubes de etiquetas, perteneciente al campo de la minería de datos (Hand et al., 2001, Tan et al., 2019; Zaki & Meira, 2014) y la entropía de Shannon, una medida para representar la probabilidad de ocurrencia de un símbolo en una señal informática (Cover & Thomas, 2006). Ambos elementos se combinan para analizar la estructura y los patrones de las frases generadas por los actores de las sesiones (estudiantes y profesores) en las diferentes actividades que se aplicaron.

Se utiliza la entropía de Shannon para proponer una medida de información de la carga de conceptos cotidianos y científicos en las frases de los estudiantes al explicar y argumentar un fenómeno complejo. Aunque las frases son estructuras heterogéneas considerando su longitud y su semántica, el método genera una medida en función de la entropía para cada frase, con lo que se puede realizar comparaciones y ordenaciones. La metodología genera información en diferentes niveles, las nubes de etiquetas son muestras generales de cada sesión y la comparación entre ellas permite ver la inclusión o el desuso de ciertas palabras; mientras que la entropía de las frases permite realizar comparaciones entre el conjunto de frases generadas en las actividades o entre los participantes, encontrar estructuras determinadas y distribuciones en función de su valor de entropía.

Se utilizó la metodología Knowledge Discovery in Databases (KDD) (Fayyad et al, 1996) para guiar el proceso de comparación entre las frases de los estudiantes. La metodología consistió en cuatro etapas: *selección del conjunto de datos*, *preprocesamiento*, *transformación* y *medidas*, y *visualizaciones*, explicadas en los siguientes párrafos.

Selección del conjunto de datos

Para obtener el conjunto de datos origen se utilizaron los videos de las grabaciones de las sesiones 2, 3 y 4 y se transcribieron en dos fases, la primera con el plugin de Google Tactiq (<https://tactiq.io/>), que generó una primera aproximación de la fragmentación de las frases por parte del profesor y de los alumnos; y la segunda, la verificación manual, donde se cotejó el conjunto completo de frases transcritas para hacerlas coincidir fielmente con lo hablado en los videos. Posteriormente, las transcripciones de cada sesión se almacenaron en el manejador de base de datos PostgreSQL para una recuperación y manipulación adecuada. Como resultado de este proceso, en la sesión 2 se obtuvieron 671 frases; en la sesión 3, 764 frases; y en la sesión 4 se obtuvieron 627.

Preprocesamiento

El conjunto de frases del profesor y de los alumnos son estructuras sintácticas y semánticas que reflejan el contexto (las experiencias) y la discusión de las actividades, por tal motivo, es importante analizar estas estructuras y buscar diferencias y semejanzas relevantes entre ellas. Se realizó un *ranking* de palabras para encontrar los sustantivos o verbos más utilizados; para lograr esto se realizó una limpieza de signos de puntuación, se fragmentaron las frases en palabras y se ejecutó un algoritmo de conteo de frecuencias. Posteriormente, se definió una *bolsa de palabras* (Hand et al., 2001) considerando dos características, una frecuencia alta de aparición de las palabras y la relevancia en el contexto de las actividades. La selección de la bolsa de palabras resulta una parte importante de la metodología, ya que se apoya en la experiencia del equipo de investigación. Se obtuvieron alrededor de 270 palabras para definir la bolsa; por *ejemplo*, *las siete más frecuentes, considerando las 3 sesiones analizadas, son tiempo, enfermedad, enfermos, probabilidad, sanos, contagio e índice*. El resultado es un diccionario de palabras que incluye las palabras habilitadas, que se aplica a cada una de las sesiones para obtener su propio conteo de frecuencias.

Transformación y medidas

A través de la tabla de las k palabras más frecuentes de la sesión, se calculó la probabilidad de aparición de la palabra i en el conjunto de palabras con la ecuación:

$$P(\text{palabra}_i) = \frac{\text{frec}(\text{palabra}_i)}{\sum_j^k \text{frec}(\text{palabra}_j)}$$

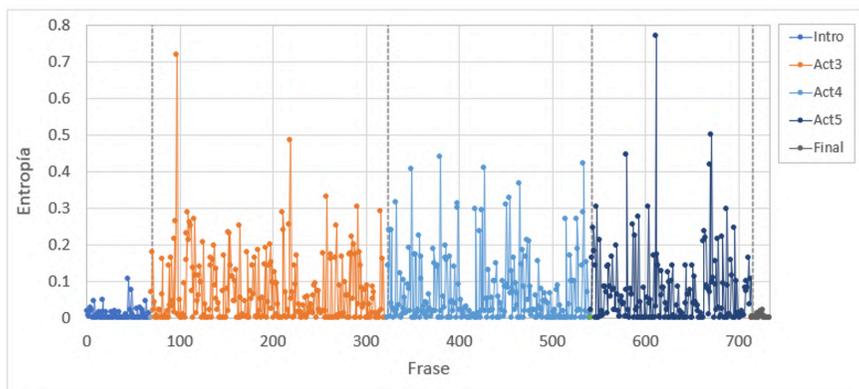
Posteriormente, con el diccionario de palabras habilitadas de cada sesión y su probabilidad de aparición se utilizó la entropía de Shannon para obtener una medida de la frase X , cuyo valor representa la carga de palabras relevantes utilizadas en la frase, y se utilizó la ecuación:

$$H(X) = \sum_i \text{habilitadas} p(x_i) \log_2 \left(\frac{1}{p(x_i)} \right)$$

Para obtener H se recorre la frase X en función del conjunto de palabras habilitadas, y con su probabilidad se calcula su entropía. Por ejemplo, la frase “establece la condición inicial en el simulador” tiene una entropía de 0.002996, ya que utiliza tres palabras habilitadas: *condición*, *inicial* y *simulador*; por lo contrario, “muchas gracias chicos, nos vemos el lunes” tiene una entropía de cero, ya que no proporciona elementos relevantes en el contexto de las actividades. De esta manera, por cada enunciado transcrito se tiene una medida que permite ser comparable. Se toma como ejemplo la sesión 3 por tener el mayor número de frases. En la Figura 3 se muestra el valor de la entropía de las frases en orden de aparición en el video.

Figura 3

Entropía de las frases de la sesión 3



En la Figura 3 se muestra la dinámica de la sesión 3, cada uno de los puntos muestra la entropía de la frase, mayor altura representa el uso de palabras con ranking alto; las líneas punteadas muestran el inicio y término de las actividades respecto al instrumento de trabajo. Hay dos frases con valores superiores a 0.7 que sobresalen respecto a las demás, éstas son explicaciones de la dinámica del simulador de epidemias que dio el profesor hacia

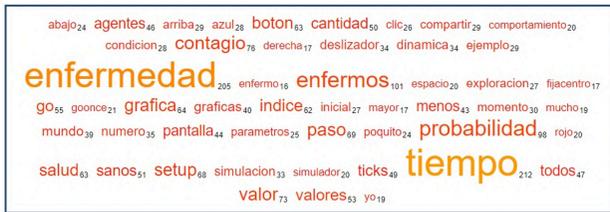
los alumnos. Las frases con entropía alta se deben a dos posibles causas, la primera es que se usan palabras con frecuencia alta y que son relevantes en el contexto, como las explicaciones proporcionadas por el profesor; y la otra causa es que puede exhibir redundancia en la idea o explicación, es decir, se usan palabras frecuentes que se repiten en la misma frase, como las proporcionadas por algunos estudiantes.

Visualizaciones

Respecto a las tablas de frecuencias se crean nubes de etiquetas con las primeras *k* palabras más frecuentes ordenadas alfabéticamente. Estas representaciones visuales contienen el número de apariciones de las palabras junto con un tamaño de fuente proporcional a su frecuencia. Por ejemplo, en la Figura 4.A se muestra la nube de etiquetas con las 50 palabras más frecuentes para el rol del profesor, y en la 4.B la de alumnos, considerando para ambas la información de las sesiones 2, 3, y 4.

Figura 4

(A) Nube de etiquetas del profesor y (B) nube de etiquetas de los alumnos



(A)



(B)

La nube de etiquetas del profesor muestra el uso de palabras formales como “mundo”, “agentes”, “espacio”, “parámetros”, además de nombres de componentes utilizados en el simulador, como “go”, “goonce”, “ticks”, “deslizador”, mientras que los estudiantes usan varios adverbios de lugar y modo para hacer descripciones: “abajo”, “arriba”, “encima”, “lento”, “rápido”, “muchos”, “pocos”, “menor”, “mayor”; en ambos casos, “tiempo” y “enfermedad” son las de mayor frecuencia.

Resultados

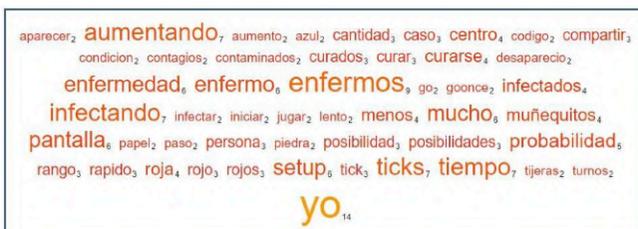
Se presentan los resultados en dos secciones de acuerdo con nuestro análisis; en la primera sección, las nubes de palabras representan una medida global de la información durante las tres sesiones, con el fin de ubicar diferencias y semejanzas de las palabras más utilizadas. En la segunda sección, nos enfocamos en un análisis más detallado de las frases realizadas por los estudiantes, utilizando la medida de entropía como recurso principal.

Explorando nubes de palabras: transición entre las experiencias cotidianas y los conceptos científicos

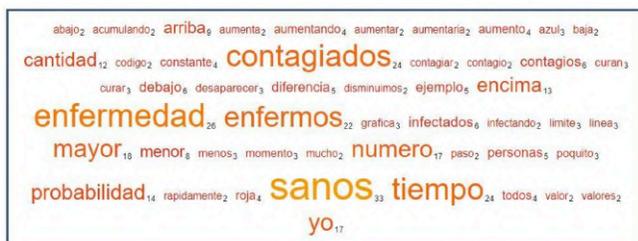
A través de las nubes de etiquetas, considerando solo el conjunto de información generado por los alumnos, se compararon las palabras con mayor frecuencia en las sesiones 2, 3, y 4 (ver Figura 5).

Figura 5

Nube de etiquetas de la sesión 2, 3 y 4 respectivamente



(A)



(B)



(C)

En la sesión 2 (Figura 5.A) se observa el uso de palabras cotidianas para describir lo que se observa en la simulación, por ejemplo, “muñequitos” o

“personas” para referirse a los agentes del modelo; se hace referencia al juego de “piedra, papel y tijeras”, que se mostró en la sesión 1 para referirse a la dinámica de contagio; se utiliza la palabra “turnos” para referirse a los cambios que pueden suceder en el agente; se utiliza el término posibilidad(es) como sinónimo de probabilidad; y se utiliza el concepto de “contaminados” y “curados” para referirse a la transición de infectados a sanos.

En la Figura 5.B crece en frecuencia de uso la palabra “contagiados”, lo que muestra un intento por describir la dinámica adecuadamente; se usan varios adverbios de modo para describir la velocidad de cambio de estados del modelo, como “rápidamente”, y conjugaciones del verbo “aumentar”. En la Figura 5.C se utilizó una medida para describir la cantidad de infectados en la población, se le llamó “índice de salud”, cuyas palabras se encuentran bien representadas en la nube de etiquetas; se encuentran en desuso las palabras “personas” y “muñequitos”, y ahora se encuentra la palabra “agentes”, que dentro de las frases se liga con la palabra “mundo”, esto es importante porque hace referencia al componente principal del simulador. En todos los casos se utilizan los colores “rojo(s)” y “azul(es)” para describir el estado de los agentes, ya sea azul para sanos o rojo para infectados.

En el gráfico de nube de palabras de las tres sesiones (Figura 5) se observa una tendencia al uso del pronombre personal yo. Se notó el uso muy productivo del yo en el lenguaje ordinario, y usos diferentes en cada aserción lingüística. Un uso del pronombre fue atraer la atención del interlocutor (el grupo), como también informativo, y no tuvo una función homogénea y hermética, pues dependió del contexto de cada oración, en ocasiones lo usaron como un yo autoconsciente, un yo reflexivo, en otras como un yo de expresión de actividad.

En resumen, el análisis de nubes de palabras reveló una transición en el lenguaje utilizado por los estudiantes a lo largo de las sesiones. Durante las primeras sesiones, se observó el empleo de expresiones cotidianas, como el uso de términos familiares como “muñequitos” y “personas” para describir los agentes en el simulador. A medida que la sesión avanzaba, se evidenció una transición hacia un lenguaje más técnico y formal, marcada por el desuso de palabras coloquiales. Además, consideramos que la versatilidad en el uso del pronombre yo expresó autoconsciencia y reflexión de los participantes en este entorno de simulación. En la siguiente sección se muestra un análisis más detallado de las frases de los estudiantes mediante el cálculo y comparación de la cantidad de información existente durante las sesiones

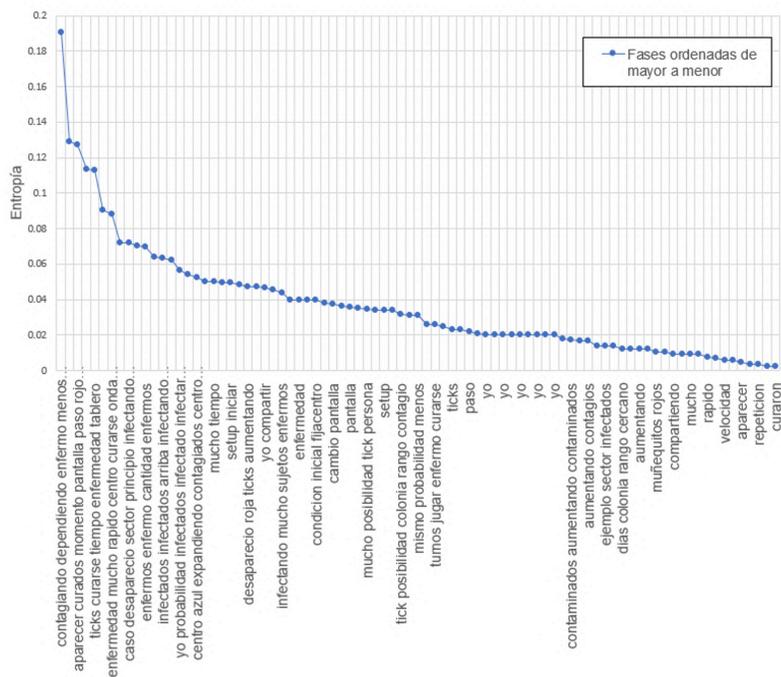
Explorando la entropía de las frases

Para un análisis más detallado se recurrió a las frases y a su valor de entropía. En la sesión 2, que comprende la actividad 1 y 2, los estudiantes generaron 146 frases, de las cuales 85 tuvieron una entropía mayor a cero. En la Figura

6 se muestra la gráfica de frases ordenadas de mayor a menor en función de la entropía, las frases con entropía cero no se consideraron. Es notable cómo decae el valor de la entropía, lo que implica que hay muchas frases con un uso relativamente bajo de palabras frecuentes, el 75% de las frases mostradas tiene una entropía menor a 0.049. En el eje horizontal se muestra el conjunto de palabras frecuentes usadas en las frases, de las cuales sobresalen: yo, enfermos, aumentando, ticks e infectados.

Figura 6

Frases ordenadas de mayor a menor en función de su entropía en la sesión 2



Por lo regular, las frases con un valor de entropía alta tienen una redundancia en la idea o explicación que tratan de realizar los estudiantes, pues se usan palabras frecuentes de manera repetitiva. En la Tabla 1 se muestran las frases con mayor entropía y algunos casos particulares.

La columna *Tipo de pensamiento en niveles* está fundamentada en las formas de generar entendimiento en un sistema descentralizado, propuesto por Colella et al. (2001). Las explicaciones que dan los estudiantes respecto a la dinámica que observan en el simulador tienen tres posibles connotaciones: de manera *local*, cuando su argumentación se basa en cómo actúan los agentes y ponen énfasis en procesos alrededor de su entorno inmediato (vecindad); *local-global*, cuando las explicaciones se entrelazan con descripciones del

agente y comportamientos generales de la dinámica; y *global* cuando solo hacen referencia a aspectos generales y no mencionan comportamiento del agente o de su vecindad.

Tabla 1

Ejemplos de frases proporcionadas por los estudiantes durante las actividades de la sesión 2

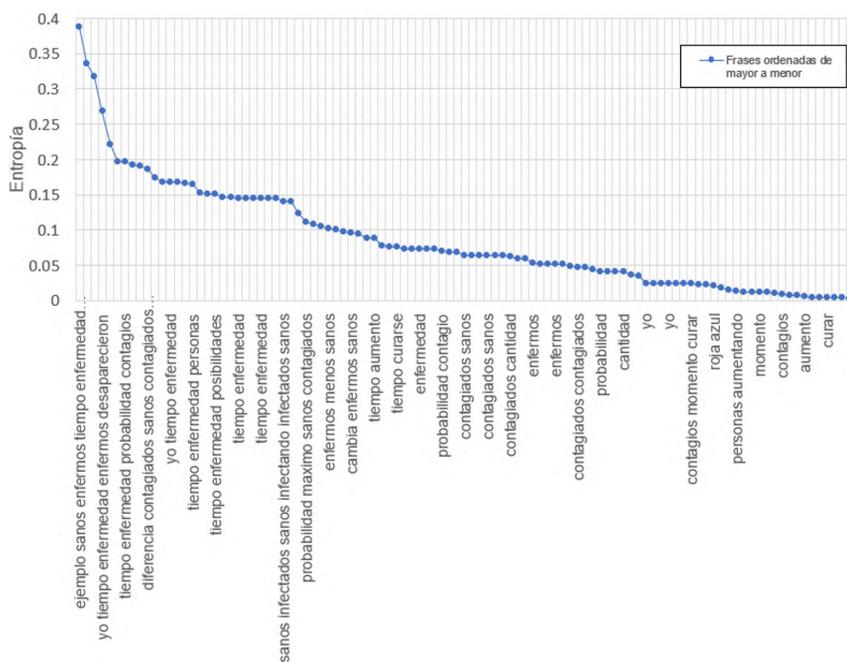
Posición	Frase	Tipo de pensamiento en niveles	Uso de palabras cotidianas
1	pues si por qué entre más tiempo este enfermo, tiene más posibilidades de ir contagiando, mientras más corto sea su tiempo que dure enfermo, tendrá menos posibilidades de ir contagiando, aumentan o decrecen las posibilidades dependiendo su tiempo de curación o lo que dure enfermo	local	posibilidad, curación
2	en todas las corridas que he hecho de poner eso de setup y go, en todos los muñequitos van cambiando a rojo, no como ahorita, a la vi hacía creo en la primera, los muñequitos se iban infectando, y ya en la segunda ya todos se pusieron en azul	global	muñequitos
9	igual como en el caso anterior, hubo uno que desapareció su sector al principio, y ya después, los que sí continuaban, continuaban infectando, fueron expandiendo la enfermedad, por eso ya se ve por todo el recuadro monitos rojos,	local-global	sector, monitos
41	sin embargo, si su tick fuera más largo, entonces ya no solamente habría posibilidad de que estuviera solo en su colonia, no, sino que vaya no sé, esté de vacaciones en tal lado o vaya a ver a la abuelita o vaya para allá entonces así también su rango de contagio se expandiría	local	posibilidad, colonia, vacaciones, abuelita
75	este es más fácil que se propague, a que, si pasas, no sé, 5 minutos, no creo que, que se corra mucho riesgo,	local	riesgo

La columna *Uso de palabras cotidianas* muestra las palabras que el estudiante usó para formar una idea o una descripción con base en su experiencia, tratando de ligar hechos de su conocimiento con lo que observa en el simulador. La columna *Posición* indica el lugar del ranking de la frase (ver Figura 6), un aspecto importante a notar es en la frase 41, donde el estudiante trata de explicar el impacto de contagio con la movilidad que puede tener un agente infectado, como ejemplo dice “ver a la abuelita” o “esté de vacaciones”; de la misma forma, la duración del movimiento del agente lo asocia con que el “tick fuera más largo”, mientras que el tick representa una unidad de tiempo.

En la sesión 3, que comprenden las actividades 3, 4 y 5, se exploraron los dos parámetros principales del simulador de epidemias, el tiempo de enfermedad y la probabilidad de contagio, y la gráfica de la serie de tiempo de sanos e infectados. En esta sesión se generaron 103 frases, de las cuales el 75% tuvo un valor de entropía menor a 0.144, el descenso de la curva es similar al de la sesión 2 (ver Figura 7), sin embargo, el valor de entropía de las frases creció. Las palabras más utilizadas en las frases fueron: sanos, enfermedad, tiempo, contagiados y enfermos.

Figura 7

Frases ordenadas de mayor a menor en función de su entropía en la sesión 3



De manera particular, en la actividad 3 se probaron distintas condiciones iniciales para que los estudiantes conocieran la evolución de la enfermedad en el mundo de agentes junto con la variación del parámetro *tiempo-enfermedad*. Las frases de los estudiantes se hicieron más especializadas respecto a la sesión 2. Se usaron algunas palabras de uso cotidiano, como “personas” para referirse a los agentes, “oportunidad” o “chances” para el mecanismo de probabilidad (ver Tabla 2) y “desnivel” para hacer referencia a la diferencia de cantidad de enfermos y sanos durante el tiempo de evolución de la enfermedad. En algunas explicaciones mencionaron la cantidad de enfermos iniciales y cómo se propagó la enfermedad en el mundo de agentes. Las

frases en las que los estudiantes hicieron explícito el número de agentes y la descripción del tránsito hacia el comportamiento final fueron consideradas un tipo de pensamiento local-global; cuando hacen referencia a una característica general, se consideraron de carácter global, y cuando hablaron específicamente del agente y la probabilidad de contagiar a otros, lo consideramos de carácter local.

Tabla 2

Ejemplos de frases proporcionadas por los estudiantes durante la actividad 3

Posición	Frase	Tipo de pensamiento en niveles	Uso de palabras cotidianas
1	también la enfermedad puede desaparecer, si al inicio todos están enfermos, independientemente del tiempo de enfermedad	global	sin usar
3	si por qué, por ejemplo, hace rato solamente iniciábamos con 2, como solamente es 0.33, la probabilidad o más bien las oportunidades de que caigan ese 0.33 son baja mientras más aumente el tiempo de enfermedad, más oportunidad o chances habrá de que el... la infección caiga como, bueno que, que acierte ese 0.33,	local	oportunidad, chances
8	si creo que es el tiempo que dura la enfermedad en esas personas y después todos se curan,	local-global	personas
26	ah, sí que se mantiene como un desnivel por, pero aún así es mayor el número de contagios,	global	desnivel
35	mmm, que esas 50 personas, pues van contagiando a los demás, pero llega el momento en el que algunos se curan y otra vez vuelven a llegarse a contagiar y así sucesivamente,	local	personas

En la actividad 4 se incluyó la gráfica serie de tiempo de sanos e infectados para tener un elemento más en el entendimiento del comportamiento del sistema y complementar lo que se visualizó en el mundo de agentes. El *tipo de pensamiento en niveles* que se observa es global, ya que, principalmente, se describen características de las series de tiempo. Las explicaciones de los alumnos disminuyeron en el uso de frases o palabras cotidianas, se puso en desuso la palabra “persona” y se refirieron a los agentes como sanos, enfermos e infectados. Tomaron en cuenta la información proporcionada por la serie de tiempo para tratar de describir el comportamiento del cambio de sanos e infectados; compararon las series entre ellas, mencionaron que las series “se mantienen constantes”, que van “a la par”, “juntitas”, “se separaron”, “ascienden” o “hay diferencia”.

Las explicaciones de las dinámicas tienen un componente de imprecisión, y en ocasiones son ambiguas, como “cada uno tocaba las orillas respectivamente”, haciendo referencia a que la serie de tiempo de sanos tenía valores alrededor del cero, y la de infectados alrededor de 1000, lo que visualmente implica que las curvas están en los extremos de la gráfica.

En la Actividad 5 se incluyó un nuevo parámetro: la probabilidad de contagio de los agentes. Las frases muestran cómo los estudiantes tratan de comprender la dinámica, pero reflejan cierta duda; no tienen la certeza de lo que están explicando, hay frases como “no a ver que hice”, “ahora no sé, a ver”, “o sea aquí ¿no?”, “pues podría, no sé”, y terminan su idea con la palabra “¿no?” para tratar de generar un asentamiento por parte del profesor. Las explicaciones ya no presentan esa seguridad como en las actividades previas. Las pruebas de las conjeturas en el simulador ya no son tan obvias, y es necesario ejecutar el simulador y experimentar.

En la Sesión 4 se realizaron las actividades 7 y 8, se registraron 69 frases, de las cuales el 75% tienen una entropía menor a 0.128, se reportan los valores de entropía más altos de las sesiones (ver Tabla 3), las palabras más utilizadas son: índice, salud, enfermos, tiempo y enfermedad.

En la actividad 7 se propone una medida de carácter global para que los estudiantes la relacionen con la cantidad de agentes sanos e infectados al transcurrir el tiempo, a la que se le llamó índice de salud de la población. Las frases generadas utilizan varias palabras técnicas que se usaron durante las actividades, cayeron en desuso palabras cotidianas y frases, recurriendo a ejemplos de la vida cotidiana o en forma de anécdota para describir la dinámica del simulador. Se estableció una relación con la variación de los parámetros, las probabilidades de contagio y lo que se obtiene con el *índice de salud*, además, se habla del mundo de agentes como el entorno donde los agentes cambian de estado de azul a rojo. El estudiante se mueve de una representación a otra en la interfaz del simulador, del *mundo de agentes*, a la gráfica *series de tiempo*, y a la medida global *índice de salud* para darle sentido a la propagación de la enfermedad y a la convergencia de un valor del número de infectados y sanos.

En la actividad 8 y última se incluye un nuevo componente en el simulador llamado *espacio de parámetros* (Figura 2.B) que se construye con base en el índice de salud. Las tonalidades de azul a rojo indican el nivel de sanos o infectados que tuvo la población al final de cada simulación. Las frases de los participantes con mayor entropía en la actividad se muestran en la Tabla 3, las frases son de carácter global y no hay uso de palabras cotidianas.

El estudiante de la posición 1 (Ver Tabla 3) trató de relacionar todos los componentes, dio una explicación considerando los valores de los botones (sliders), la serie de tiempo, el espacio de parámetros, y los valores del índice de salud. Fue una explicación plausible considerando que el espacio de

parámetros es una representación abstracta de las posibles dinámicas del modelo computacional. Las otras frases de los participantes son un tanto ambiguas, se notan las conexiones que hacen con el mundo de agentes y la serie de tiempo, pero no con el espacio de parámetros.

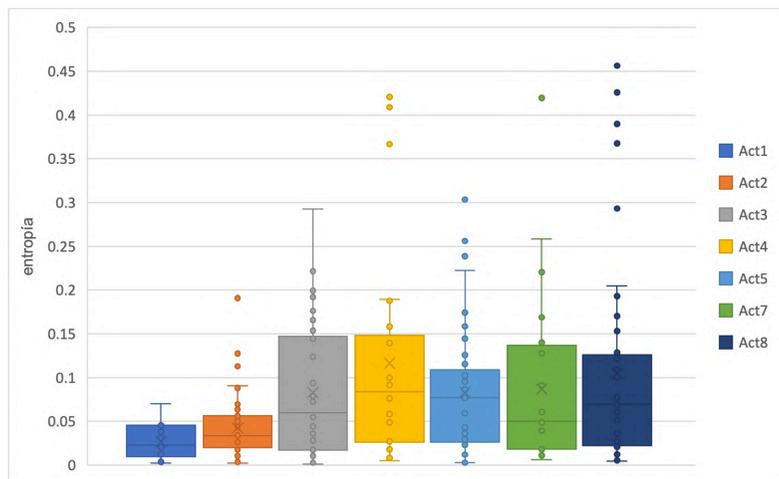
Tabla 3

Frasas con mayor valor de entropía en el estudio

Posición	Frase
1	bueno, pues ahí es la, esa línea que se ve entre el rojo y el azul es quien marca la pauta para ver si hay más, más enfermos, o más sanos en el gráfico, mientras más bajos sean los valores en los botones, más, más sanos va a haber, mientras más alto sea, más, más enfermos va a haber igual, en el índice cuanto más bajo sean los valores, el índice se va a acercarse más a al valor uno, y mientras más alto sea el índice se va a acercarse al valor 0
2	ya en esta puedo describirlo como que es como una combinación entre el primer caso y el segundo caso, o sea que ya había un 50 50 por así decirlo, ya había bastantes enfermos, pero no eran los suficientes como para compararse con los, con el cuadro del valor, de los de 100 en la probabilidad de contagio, no, en el del tiempo de enfermedad y la probabilidad de contagio en uno, o sea ya, el índice de salud podría estar por encima del 0.5 o por apenas por debajo del igual 0.5

En la Figura 8 se realiza una comparación con diagramas de caja de las distribuciones de los valores de la entropía entre las diferentes actividades. Las actividades 1 y 2, pertenecientes a la sesión 2, tienen en su conjunto los valores más bajos de entropía de las frases; de manera particular, hay varias frases en la actividad 2 consideradas como puntos atípicos o *outliers*, que comprenden las frases con mayor entropía de la actividad (ver Tabla 1). Para la sesión 3, hay un incremento general en la entropía de las frases; en particular, la actividad 4 y 5 mantienen un incremento en su valor límite máximo respecto a su distribución, y la tendencia sigue hasta la actividad 7 de la sesión 4, además se nota un incremento en valores atípicos de algunas frases.

La actividad 8 tiene un cúmulo de *outliers* y contiene la frase con la más alta entropía del estudio (ver Tabla 3), además se nota un incremento en la media aritmética (marcadas en X dentro de la caja) en la actividad 5, 7 y 8. En general, las distribuciones de las sesiones 3 y 4 mantienen una semejanza en su rango intercuartil (RIC) y surgen frases con una carga alta de palabras frecuentes que salen de la distribución típica, mientras que en la sesión 2, el RIC es de longitud pequeña con una media aritmética menor a 0.05. En la sesión 4 (actividad 7 y 8) se mantiene un RIC similar a la sesión 3, además, hay un incremento en el uso de palabras frecuentes en las frases, las cuales se visualizan en los puntos atípicos.

Figura 8*Distribuciones de los valores de la entropía de las frases por actividad*

Discusión

El presente estudio ha explorado el papel que juegan las experiencias de los estudiantes fuera del contexto escolar en la construcción de significados sobre el fenómeno de propagación de epidemias. Al abordar esta temática, nos propusimos no solo analizar las manifestaciones del lenguaje cotidiano, sino también comprender cómo las vivencias de los estudiantes impactan en la comprensión de los niveles de organización de un sistema complejo, simulado mediante un modelo basado en agentes en el software NetLogo.

El análisis de nubes de palabras reveló una transición en el lenguaje utilizado por los estudiantes a lo largo de las tres sesiones analizadas; a medida que avanzaban, se observó un cambio hacia un lenguaje más técnico, incorporando nociones de conceptos científicos. Este cambio es indicativo de un proceso, donde los estudiantes adoptan nuevos términos y lo integran en otros para la comprensión del fenómeno estudiado.

Por otro lado, se llevó a cabo un análisis detallado mediante el uso de frases y su valor de entropía como herramientas clave. La entropía se utilizó para evaluar el contenido informativo de las frases; este método proporcionó medidas comparables para cada enunciado transcrito, facilitando una evaluación cuantitativa de la riqueza informativa en el discurso estudiado. Los ordenamientos de mayor a menor de los valores de la entropía de las frases muestran una estructura semejante en las diferentes sesiones, al inicio hay algunas frases con valor alto, y posteriormente disminuyen hacia valores cercanos a cero; sin embargo, las cotas máximas crecen conforme avanzan las sesiones. Por ejemplo, el promedio del valor de entropía de las 10 primeras frases de la sesión 2 es 0.1, la sesión 3, 0.24, y la sesión 4, 0.32, lo que muestra

un incremento progresivo en el uso de las palabras frecuentes y representativas dentro del contexto de las actividades.

Nuestros resultados mostraron que el diseño de las actividades favoreció la interfuncionalidad entre lo cotidiano y lo científico en términos de expresiones de los estudiantes. El hecho de que las frases con connotaciones científicas y cotidianas tengan una mayor entropía que las frases meramente cotidianas sugiere una transición de las palabras que utilizan los estudiantes. El análisis de las frases revela esta interfuncionalidad que va cambiando a un lenguaje más formal conforme el paso de las sesiones. Este hallazgo refuerza la idea de que la incorporación de perspectivas y expresiones diversas en el ámbito del salón de clases debe ser reconocida como un método de organización en el trabajo escolar.

A lo largo de las sesiones se identificaron frases que nos permitieron inferir la asimilación de procesos básicos de la propagación de una epidemia, como su duración y la probabilidad de contagio del agente. Los argumentos etiquetados como de carácter *local*, *global* y *local-global* nos permitieron caracterizar el tipo de argumentación que realizaban; en las primeras tres actividades se entrelazaban lo local y global, mientras que, en las siguientes actividades, la argumentación se centraba en descripciones de la dinámica global, como conteos y descripciones de las series de tiempo, y no tanto en las vecindades de los agentes.

En resumen, por un lado, el análisis de nubes de palabras nos proporcionó información sobre la evolución del lenguaje utilizado dentro de las actividades. La transición de un lenguaje cotidiano a uno técnico no solo mostró la incorporación de nociones de conceptos científicos, sino también la capacidad de los estudiantes para argumentar, de manera distinta, sus respuestas dentro de las discusiones. Por otro lado, el análisis de la entropía mostró que los estudiantes integraron términos científicos en su discurso cotidiano con una tendencia hacia un mayor uso de términos del lenguaje científico contenido en las actividades. Los resultados de la aplicación de este método cuantitativo de evaluación de frases muestran el posible potencial de la modelación basada en agentes como una herramienta educativa para fomentar la comprensión de fenómenos complejos en los estudiantes.

Conclusiones

Los aportes metodológicos de este estudio impactan directamente en la investigación en educación, adoptando una perspectiva mixta de análisis que enriquece la comprensión de escenarios educativos, que incluyen el estudio de modelos y simulaciones computacionales de fenómenos complejos.

El diseño secuenciado de las actividades en este estudio provee escenarios para el desarrollo de aprendizajes basados en la simulación computacional y la MBA. Sin embargo, para el éxito del diseño fue crucial considerar la

diversidad de ideas y experiencias fuera de la escuela de los estudiantes, lo que fomentó su participación durante todas las sesiones. Nuestro diseño permitió dar sentido a los diferentes niveles de organización que componen el sistema complejo del fenómeno de propagación de epidemias.

Una de las limitaciones de este trabajo es que los resultados se encuentran restringidos a los datos obtenidos en cuatro sesiones de trabajo con un mismo grupo. En este sentido, es esencial validar nuestros hallazgos con una población más extensa y la implementación de iteraciones adicionales del diseño.

La mayor parte de nuestra metodología se basa en procesos algorítmicos y automáticos para procesar datos, sin embargo, una parte sensible es la definición de la bolsa de palabras, donde la experiencia del investigador toma un papel importante para el refinamiento y establecimiento del contexto. Por lo tanto, es importante establecer lineamientos adecuados para evitar ambigüedades, sesgos y falta de contexto en los análisis. En miras al futuro, vislumbramos la aplicación de nuestras actividades con audiencias más extensas, junto con la integración de nuevas estrategias que promuevan una participación entre los estudiantes en un salón de clases.

Agradecimientos

Agradecemos al Consejo Nacional de Humanidades, Ciencias y Tecnologías CONAHCYT por el apoyo al proyecto A1-S-33505 “Construcción de significados en procesos de modelación matemática. Una aproximación basada en el uso de herramientas de simulación computacional desde una perspectiva semiótica”.

Referencias

- Abrahamson, D., Berland, M., Shapiro, B., Unterman, J., & Wilensky, U. (2006). Leveraging epistemological diversity through computer-based argumentation in the domain of probability. *For the Learning of Mathematics*, 26(3), 19–45. <http://www.jstor.org/stable/40248549>
- Abrahamson, D., & Wilensky, U. (2005). ProbLab goes to school: Design, teaching, and learning of probability with multi-agent interactive computer models. En M. Bosch (Ed.), *The Proceedings of the Fourth Conference of the European Society for Research in Mathematics Education* (pp. 570–579). FUNDEMI IQS – Universitat Ramon Llull - ERME. http://erme.site/wp-content/uploads/2021/06/CERME4_WG5.pdf
- Anhalt, C. O., Cortez, R., & Bennett, A. B. (2018). The emergence of mathematical modeling competencies: An investigation of prospective secondary mathematics teachers. *Mathematical Thinking and Learning*, 20(3), 202–221. <https://doi.org/nzqp>

- Arrieta Vera, J., & Díaz Moreno, L. (2015). Una perspectiva de la modelación desde la Socioepistemología. *Revista latinoamericana de investigación en matemática educativa*, 18(1), 19–48. <https://doi.org/nzqq>
- Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer. <https://link.springer.com/book/9780387310732>
- Brady, C., & Yarnes, L. (2020). Embodied participatory simulations of disease as an entry point for network analysis. En B. Tangney, J. R. Byrne & C. Girvan (Eds.), *Proceedings of the 2020 Constructionism Conference* (pp. 89–91). *Constructionism Conference*. <https://bit.ly/4boiDHo>
- Cantoral, R., Ríos Jarquín, W., Reyes Gasperini, D., Cantoral Uriza, E. A., Barrios, E., Fallas Soto, R., ... & Bonilla Solano, A. (2020). Matemática Educativa, transversalidad y COVID-19. *Revista latinoamericana de investigación en matemática educativa*, 23(1). <https://doi.org/nzqq>
- Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British journal of Developmental Psychology*, 3(1), 21–29. <https://doi.org/dv676g>
- Clark, D. B., & Sengupta, P. (2019). Reconceptualizing games for integrating computational thinking and science as practice: Collaborative agent-based disciplinarily-integrated games. *Interactive Learning Environments*, 28(3), 328–346. <https://doi.org/ggdpg4>
- Colella, V. S., Klopfer, E., & Resnick, M. (2001). *Adventures in modeling: Exploring complex, dynamic systems with StarLogo*. Teachers College Press. <https://dl.acm.org/doi/10.5555/558772>
- Costa, M. C., & Domingos, A. (2022). Promoting mathematics teaching with real life scenarios in the context of COVID-19 pandemic. En J. Hodgen, E. Geraniou, G. Bolondi, & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)* (pp. 4569–4402). Free University of Bozen-Bolzano - ERME. <https://hal.science/hal-03745362/document>
- Cover, T.M., & Thomas, J.A. (2006). *Elements of information theory*. John Wiley & Sons, Inc. <https://doi.org/djwjd6>
- Díaz Costa, E., Romero Jeldres, M., Faouzi Nadim, T., & Pardo Soto, C. (2022). Modelos predictivos de la competencia pedagógica en docentes de EMTP mediante la minería de datos educacionales. *Estudios pedagógicos*, 48(2), 179–197. <https://doi.org/nzqr>
- Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling* (STU-Student edition). Princeton University Press. <http://www.jstor.org/stable/j.ctt7rxj1>
- Fayyad, U., Piatetsky-Shapiro, G., y Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37. <https://doi.org/ggf6n8>

- Gaff, H., Lyons, M., & Watson, G. (2011). Classroom manipulative to engage students in mathematical modeling of disease spread: $1+1=A$ chool!. *Mathematical Modelling of Natural Phenomena*, 6(6), 215–226. <https://doi.org/dhpszv>
- Gibbons, K., Karp, K. S., & Dillon, F. (2015). Epidemics, exponential functions, and modeling. *Mathematics Teaching in the Middle School*, 21(2), 90–97. <https://doi.org/gdthxb>
- Gray, R.M. (2011). *Entropy and information theory* (2a ed.). Springer. <https://doi.org/dm2jfb>
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. The MIT Press. <https://mitpress.mit.edu/9780262082907/principles-of-data-mining/>
- Hernández-Zavaleta, J. E., Brady, C., Becker, S., y Clark, D. B. (2023). Vygotskian hybridizing of motion and mapping: Learning about geometric transformations in block-based programming environments. *Mathematical Thinking and Learning*, 27(1), 25–59. <https://doi.org/n8sv>
- Kovalenko, O., Shtefan, L., Yaschun, T., Bondarenko, T., & Ohdanskiy, K. (2022). Predicting the educational and cognitive activity of teaching engineers in computer science based on Mathematical Models. En M. E. Auer, H. Hortsch, O. Michler, & T. Köhler (Eds.), *Mobility for smart cities and regional development - Challenges for higher education* (pp. 616–623). Springer. <https://doi.org/nzqs>
- Levy, S. T., & Wilensky, U. (2008). Inventing a “mid level” to make ends meet: Reasoning between the levels of complexity. *Cognition and Instruction*, 26(1), 1–47. <https://doi.org/dspnfr>
- Resnick, M. (1994). *Turtles, Termites. Traffic Jams: Explorations in Massively Parallel*. The MIT Press. <https://mitpress.mit.edu/9780262680936/turtles-termites-and-traffic-jams/>
- Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, 18(2), 351–380. <https://doi.org/gc3f7b>
- Sengupta, P., Dickes, A., & Farris, A. V. (2021). *Voicing code in STEM: A dialogical imagination*. MIT Press. <https://doi.org/nzqn>
- Sengupta, P., & Wilensky, U. (2011). Lowering the learning threshold: Multi-agent-based models and learning electricity. En M. Khine, & I. Saleh (Eds.), *Models and modeling: Cognitive tools for scientific enquiry* (pp. 141–171). Springer. <https://doi.org/bbwrrnp>
- Sivakumar, S., Venkataraman, S., & Selvaraj, R. (2016). Predictive modeling of student dropout indicators in educational data mining using improved decision tree. *Indian Journal of Science and Technology*, 9(4), 1–5. <https://doi.org/ggcwzj>

- Sokkhey, P., Navy, S., Tong, L., & Okazaki, T. (2020). Multi-models of educational data mining for predicting student performance in mathematics: A case study on high schools in Cambodia. *IEIE Transactions on Smart Processing and Computing*, 9(3), 217–229. <https://doi.org/nzqt>
- Tan, P.N., Steinbach, M., & Kumar, V. (2019). *Introduction to data mining* (2a ed.). Pearson. <https://bit.ly/4itoo8L>
- Vygotsky, L. S. (1986). *Thought and language*. The MIT Press.
- Wagh, A. & Wilensky, U. (2012). Evolution in blocks: Building models of evolution using blocks. En C. Kyngios, J. Clayson, & N. Yiannoutsou (Eds.), *Constructionism: Theory, Practice and Impact. Proceedings of Constructionism* (2012) (pp. 549–554). The Educational Technology Lab, National & Kapodistrian University of Athens. <https://bit.ly/4ilq8RD>
- Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of Research in Science Teaching*, 38(5), 529–552. <https://doi.org/d2k9f2>
- Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, 25(1), 127–147. <https://doi.org/f77xtr>
- Wilensky, U. (1999). *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University. <http://ccl.northwestern.edu/netlogo/>
- Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. The MIT Press. <https://bit.ly/3F466ws>
- Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and Instruction*, 24(2), 171–209. <https://doi.org/b3w4vq>
- Wong, J. C. F., & Yip, T. C. Y. (2020). Measuring students' academic performance through educational data mining. *International Journal of Information and Education Technology*, 10(11), 797–804. <https://doi.org/nzqk>
- Zaki, M., & Meira, Jr, W. (2014). *Data mining and analysis: Fundamental concepts and algorithms*. Cambridge University Press. <https://doi.org/nzqmq>

