

Identificación de dimensiones de bienestar y calidad de vida a partir de modelamiento de temas del lenguaje natural de las personas

Identification of dimensions of well-being and quality of life through natural language processing modeling of people's discourse

Identificação das dimensões de bem-estar e qualidade de vida a partir da modelagem de temas da linguagem natural das pessoas

<https://doi.org/10.15332/21459169.10248>

Artículo

Guberney Muñetón Santa¹
Guillermo León Moreno Soto²

Recibido: 11/09/2024
Aprobado: 15/11/2024

Citar como:

Muñetón Santa, G., & Moreno Soto, G. L. (2024). Identificación de dimensiones de bienestar y calidad de vida a partir de modelamiento de temas del lenguaje natural de las personas. *Análisis*, 56(105), 91-116. <https://doi.org/10.15332/21459169.10248>



Resumen

Este artículo aborda la complejidad que conlleva medir la calidad de vida y el bienestar, reconociendo las limitaciones de los enfoques tradicionales que a menudo no logran capturar las dimensiones subjetivas y contextuales que realmente importan a las personas. A través de una metodología de modelamiento de temas correlacionados, se identificaron las dimensiones clave de la calidad de vida a partir de respuestas textuales de encuestas realizadas en Medellín, Colombia. Los resultados revelan mayores frecuencias en términos como “salud”, “pobre de espíritu” y “vivir bien”. Además, el modelamiento de temas sugiere que la calidad de vida está relacionada con aspectos tanto objetivos como subjetivos del bienestar. De esta manera emergen dimensiones de vulnerabilidad material, necesidades básicas, bienestar espiritual, superación personal y salud integral. La discusión se centra

¹ Profesor e investigador del Instituto de Estudios Regionales (INER), Universidad de Antioquia, Medellín, Colombia. Doctor en Ingeniería Electrónica y de Computación. Grupo de Investigación Recursos Estratégicos, Región y Dinámicas Socioambientales (RERDSA). correo electrónico: guberney.muneton@udea.edu.co. ORCID: <https://orcid.org/0000-0002-5194-1914>

² Investigador del Instituto de Estudios Regionales, Universidad de Antioquia. Magíster en Desarrollo. Grupo de Investigación Recursos Estratégicos, Región y Dinámicas Socioambientales (RERDSA) y Grupo de Investigación - Escuela de Prospectiva y Desarrollo Empresarial de la Institución Universitaria Esumer, Medellín, Colombia. Correo electrónico: gleon.moreno@udea.edu.co ORCID: <https://orcid.org/0000-0003-1400-710X>

en las implicaciones de estos hallazgos para la formulación de políticas públicas más inclusivas y efectivas, y sugiere vías para futuras investigaciones que integren enfoques cualitativos y cuantitativos en la evaluación del bienestar y la calidad de vida.

Palabras clave: calidad de vida, bienestar, modelamiento de temas, procesamiento de lenguaje natural, Medellín.

Abstract

This article addresses the complexity of measuring quality of life and well-being, recognizing the limitations of traditional approaches that often fail to capture the subjective and contextual dimensions that truly matter to people. Through a correlated topic modeling methodology, key dimensions of quality of life were identified based on textual responses from surveys conducted in Medellín, Colombia. The results reveal higher frequencies for terms such as "health," "poor in spirit," and "living well." Moreover, the topic modeling suggests that quality of life is related to both objective and subjective aspects of well-being. Dimensions of material vulnerability, basic needs, spiritual well-being, personal growth, and holistic health emerge from the analysis. The discussion focuses on the implications of these findings for more inclusive and effective public policy formulation and suggests pathways for future research that integrate both qualitative and quantitative approaches in the evaluation of well-being and quality of life.

Keywords: quality of life, well-being, topic modeling, natural language processing, Medellín.

Resumo

O presente artigo aborda a complexidade de medir a qualidade de vida e o bem-estar, reconhecendo as limitações das abordagens tradicionais que muitas vezes não conseguem capturar as dimensões subjetivas e contextuais que realmente importam para as pessoas. Através de uma metodologia de modelagem de temas correlacionados, foram identificadas as dimensões-chave da qualidade de vida a partir de respostas textuais de questionários realizados em Medellín, Colômbia. Os resultados revelam maiores frequências em termos como "saúde", "pobre de espírito" e "viver bem". Além disso, a modelagem de temas sugere que a qualidade de vida está relacionada tanto a aspectos objetivos quanto subjetivos do bem-estar. Emergiram dimensões de vulnerabilidade material, necessidades básicas, bem-estar espiritual, superação pessoal e saúde integral. A discussão foca nas implicações desses achados para a formulação de políticas públicas mais inclusivas e eficazes, e sugere caminhos para futuras investigações que integrem abordagens qualitativas e quantitativas na avaliação do bem-estar e da qualidade de vida.

Palavras-chave: qualidade de vida, bem-estar, modelagem de temas, processamento de linguagem natural, Medellín.

Introducción

La medición de la calidad de vida y del bienestar ha sido un eje central en las ciencias sociales y en la formulación de políticas públicas, dado que estas métricas resultan fundamentales

para comprender y mejorar las condiciones de vida en diversos contextos. A pesar de los esfuerzos significativos por desarrollar indicadores precisos y representativos, persiste un desafío crucial: las mediciones actuales son incapaces de capturar de manera exhaustiva y contextualizada las dimensiones que realmente importan a las personas, en consideración de su contexto territorial específico y condiciones cotidianas. Aunque se han desarrollado indicadores como el índice de desarrollo humano (IDH), este ha sido cuestionado por no captar adecuadamente las desigualdades internas y las diferencias culturales en la percepción del bienestar (Klugman et ál., 2011). De manera similar, el índice de progreso social (IPS), que incluye dimensiones como las necesidades humanas básicas, los fundamentos del bienestar y las oportunidades, ha sido valorado por su enfoque multidimensional; sin embargo, enfrenta desafíos en cuanto a la comparabilidad entre países de diferentes niveles de desarrollo (Porter et ál., 2017).

En América Latina, la medición del bienestar ha evolucionado para reflejar las particularidades de la región, caracterizada por altas tasas de desigualdad y diversidad cultural. Un ejemplo notable es el índice de pobreza multidimensional (IPM), adaptado por varios países latinoamericanos siguiendo la metodología del Oxford Poverty and Human Development Initiative (OPHI), que busca capturar de manera más precisa las múltiples facetas de la pobreza y el bienestar (Alkire et ál., 2015; Alkire y Santos, 2014). En la metodología que propone OPHI, siguiendo el enfoque de las capacidades, se insta a realizar ejercicios participativos para la discusión de las mediciones de bienestar, principalmente, para elegir los elementos considerados importantes para la sociedad en particular donde se hará la medición (Alkire et ál., 2015). Lo anterior es un llamado que tiene sustento en investigaciones como el estudio “Voces de los Pobres” del Banco Mundial, que recogió testimonios de más de 60 000 personas en 60 países y reveló que los pobres valoran aspectos que trascienden lo material, tales como la dignidad, el respeto y la libertad para tomar decisiones sobre sus propias vidas (Narayan et ál., 2000).

Estas voces subrayan la necesidad de identificar las dimensiones del bienestar a partir de las vivencias cotidianas de las personas, reconociendo que ellas mismas están en la mejor posición para valorar cómo les va en la vida y cuáles son los objetos de valor que consideran esenciales para su bienestar. Investigaciones más recientes sobre estas cuestiones abogan por la integración de la voz de las personas en los procesos de comprensión de las dimensiones y sus relaciones (ATD Fourth world y University of Oxford, 2019). La omisión de estas perspectivas en las mediciones tradicionales restringe la capacidad de las políticas públicas para atender de manera efectiva las necesidades reales de las poblaciones vulnerables.

Al respecto, son varios los indicadores de pobreza, bienestar y calidad de vida que, por decisión práctica, por adscripción teórica, por rapidez en la implementación o por desconocimiento de los argumentos, deciden evitar los procesos de escuchar la voz de las personas. Por ejemplo, el índice de pobreza multidimensional de Colombia (DNP, 2017) elige

cinco dimensiones de análisis con el argumento de haber revisado los estudios de las voces de los pobres, pero alineando su medición a las necesidades del plan de desarrollo de turno (Angulo et ál., 2016), no se hace ejercicios de participación ciudadana. Otro caso más local es la medición del índice multidimensional de calidad de vida del distrito de Medellín, cuya metodología cambia en el 2024, pero define las dimensiones de análisis retomando propuestas de expertos internacionales para contextos diferentes al local, además sin una perspectiva teórica definida en el reporte de resultados.

En este sentido, este artículo propone el uso de modelamiento de temas para la identificación de dimensiones clave de la calidad de vida y bienestar a partir de respuestas textuales de encuestas realizadas en Medellín, Colombia. Esta metodología tiene el potencial de automatizar la identificación de dimensiones relevantes, permitiendo que las mediciones se adapten al contexto territorial específico, condiciones cotidianas, experiencias y valores particulares de las personas, y, de esta manera, superen las limitaciones de los enfoques tradicionales. Al analizar las respuestas directas de los individuos, se puede deducir con mayor precisión cuáles son los aspectos de la vida que consideran valiosos. De este modo, se aporta en la comprensión de la calidad de vida y el bienestar y se proporcionan herramientas de diagnóstico situado para las políticas públicas y la generación de acciones más contundentes, con el fin de que respondan de manera más precisa a las necesidades y aspiraciones de las personas en distintos contextos.

Estado del arte

El concepto de bienestar ha evolucionado significativamente a lo largo del tiempo, pasando de un enfoque predominantemente económico a una perspectiva más integral que abarca tanto dimensiones objetivas como subjetivas. Durante el siglo XX, el bienestar se medía principalmente a través del crecimiento económico, utilizando indicadores financieros como el producto interno bruto (PIB). Esta visión asumía que la prosperidad económica era la principal vía para mejorar la calidad de vida de las personas. Sin embargo, esta perspectiva ha sido cuestionada, debido a su incapacidad para capturar la complejidad del bienestar humano (Stiglitz et ál., 2009; Sempere et ál., 2010; Guillén y Phelán, 2016). Como respuesta a ello, surgió la necesidad de desarrollar marcos conceptuales que integren diferentes dimensiones del bienestar, lo que ha permitido la aparición de diversas teorías y metodologías.

El enfoque de las capacidades, propuesto por Amartya Sen y Martha Nussbaum, ha transformado profundamente la noción de bienestar, al alejarse del utilitarismo clásico que la medía según la maximización de la utilidad (Sen, 1999). Este enfoque redefine el bienestar en términos de la expansión de las libertades y capacidades que permiten a las personas llevar una vida que valoran (Sen, 1985, 1999, 2009). Nussbaum (2000, 2011) añade que las “capacidades

centrales” deben ser vistas como derechos humanos fundamentales, con lo cual refuerza el marco normativo del desarrollo humano. Esta perspectiva ha influido en la creación de indicadores como el IDH, que mide el bienestar considerando la salud, la educación y el nivel de vida (PNUD, 2010). El enfoque de capacidades también ha impactado la agenda internacional, especialmente en los objetivos de desarrollo sostenible (ODS) y otras políticas de desarrollo humano que buscan erradicar la pobreza y asegurar vidas dignas (ONU, 2015; Stiglitz et ál., 2009). Además, según Robeyns (2017), este enfoque reconoce la subjetividad en la evaluación del bienestar, destacando la relevancia de los valores y preferencias individuales en la toma de decisiones.

El bienestar y la calidad de vida son conceptos que no cuentan con definiciones únicas o universales, debido a la diversidad de enfoques teóricos y contextuales que los moldean. A lo largo del tiempo, diferentes corrientes filosóficas, como el hedonismo o teorías basadas en listas de valores, han intentado capturar lo que constituye el bienestar, pero ninguna de ellas ha logrado abarcar la totalidad de las experiencias humanas. El enfoque de las capacidades, por su parte, sugiere que el bienestar no debe limitarse a la felicidad o satisfacción, sino a lo que una persona es capaz de hacer y ser en su vida (Sen, 1993). Guy Fletcher (2016) apoya esta idea, al señalar que el bienestar no puede reducirse a una única teoría o concepto, sino que depende de una interacción compleja de logros personales, satisfacción de necesidades y el contexto socioeconómico en el que se desenvuelven las personas.

En este sentido, el bienestar se interrelaciona pero no es equivalente a la calidad de vida. Ambos conceptos implican múltiples dimensiones, pero el bienestar, desde el enfoque de capacidades, se enfoca más en las libertades y oportunidades que tienen las personas para llevar una vida que valoran, mientras que la calidad de vida considera más los factores materiales y sociales que permiten o limitan ese bienestar. Sen y Nussbaum (1996) argumentan que la calidad de vida no solo se basa en lo que las personas poseen, sino en lo que pueden hacer con esos recursos. Así mismo, la calidad de vida abarca tanto condiciones objetivas como subjetivas, incluyendo la salud, la educación y la satisfacción personal (Diener et ál., 2009; Robeyns, 2017). Además, el reporte de Stiglitz et ál. (2009) refuerza esta visión, al criticar la dependencia excesiva en indicadores económicos como el PIB para medir el bienestar. El reporte subraya que el bienestar es un concepto más amplio, que debe integrar tanto aspectos objetivos como subjetivos. En línea con el enfoque de Sen, el reporte insiste en que el bienestar debe medirse en función de lo que las personas pueden hacer y ser, y no solo por los recursos que poseen (Stiglitz et ál., 2009).

Un aspecto clave en la medición del bienestar es el bienestar subjetivo, que ha ganado relevancia en las últimas décadas, especialmente en relación con cómo las personas perciben y evalúan su propia vida. Este enfoque se centra en las evaluaciones cognitivas y emocionales de las personas sobre su vida, como la satisfacción con la vida y la experiencia de emociones

positivas y negativas (Diener, 1984; Kahneman y Deaton, 2010). El bienestar subjetivo ha sido ampliamente estudiado en psicología y ha influido en el diseño de políticas públicas orientadas a mejorar la felicidad y la satisfacción personal (Diener et ál., 2018; Kahneman y Riis, 2005; NEF, 2012; De Neve y Layard, 2023). Según Sen (1993), estar feliz puede considerarse como un funcionamiento complejo, es decir, como una parte importante del bienestar, pero no debe reducirse a la perspectiva hedonista o utilitarista.

El reporte de Stiglitz et ál. (2009) introduce la idea de que el bienestar es multidimensional y que las políticas públicas deben enfocarse en promoverlo en todas sus facetas, no solo en el crecimiento económico. Según el reporte, el bienestar de una sociedad debe medirse tanto por la riqueza material como por el acceso a oportunidades, libertades y condiciones que permitan llevar una vida digna. En este sentido, tanto el bienestar como la calidad de vida son conceptos complementarios para entender el bienestar humano. No obstante, el reporte destaca que el bienestar es dinámico y depende no solo de las oportunidades reales de las personas, sino también de sus valoraciones subjetivas sobre su propia calidad de vida.

El reconocimiento de la multiplicidad de dimensiones en la medición del bienestar y la calidad de vida ha sido ampliamente aceptado en los círculos académicos y de política pública. Sin embargo, uno de los principales desafíos radica en la selección de las dimensiones que representen adecuadamente a un grupo social diverso, lo que plantea un problema de elección social. Este asunto fue explorado profundamente por Amartya Sen (2017), quien desarrolló una respuesta a la famosa “imposibilidad” de Arrow en la elección social. El teorema de imposibilidad de Arrow (1951) establece que no existe un sistema de votación que pueda convertir las preferencias individuales en una preferencia colectiva que cumpla con todos los criterios democráticos deseados. Ante esta dificultad, Sen argumenta que la solución no radica en simplemente calcular agregados matemáticos, sino en un proceso deliberativo e informado.

La respuesta que Sen ofrece a este problema se basa en la deliberación pública. Este enfoque propone activar la participación ciudadana y la discusión pública para definir qué dimensiones del bienestar son relevantes en un contexto específico. La deliberación democrática, para Sen, es un proceso necesario con vistas a lograr que la elección de dimensiones sea inclusiva y reflexiva. Esto se alinea con su enfoque de capacidades, que enfatiza la importancia de que las personas tengan la libertad de elegir y participar en las decisiones que afectan su bienestar (Sen, 1999, 2009). Autores como Robeyns (2005) y Nussbaum (2011) también apoyan este enfoque participativo y sugieren que un proceso de elección social que involucre la deliberación pública puede superar las limitaciones impuestas por modelos puramente cuantitativos.

La perspectiva participativa en la elección de dimensiones ha ganado terreno en el campo académico y en el diseño de políticas públicas, como un enfoque que valora la voz de las

personas directamente afectadas. Esta perspectiva se distingue de otras metodologías que proponen listas universales de capacidades, como las desarrolladas por Martha Nussbaum (2001, 2011), o de enfoques más estructurados y constitucionales (Burchi et ál., 2014, 2018, 2021). También, se aleja de los métodos que usan el criterio de experto como única fuente de información. En contraste, el enfoque participativo reconoce que las personas, especialmente aquellas en situación de pobreza, son expertas en sus propias realidades y, por lo tanto, están en la mejor posición para identificar las dimensiones que son relevantes para mejorar su calidad de vida (Frediani, 2019). Un ejemplo clave de esta práctica es el estudio del Banco Mundial (Narayan-Parker et ál., 2000) donde se consultó a más de 20 000 personas pobres de 23 países. El resultado de dicho estudio fue la identificación de diez dimensiones fundamentales, lo que demuestra el poder de la participación en la construcción de políticas públicas.

El enfoque participativo tiene ventajas sobre el método basado en de expertos. Voinov et ál. (2018) señalan que, en muchos procesos de modelado participativo, los expertos tienden a dominar las decisiones metodológicas, lo que puede restringir la diversidad de perspectivas. De manera similar, Mendoza y Prabhu (2000) subrayan que el uso exclusivo del criterio de expertos puede limitar la representatividad de los resultados, por lo cual se sugiere la necesidad de métodos participativos que complementen las decisiones técnicas. Jonsson et ál. (2007) advierten que esta dominación de los expertos puede reducir la influencia real de las partes interesadas locales y comprometer el objetivo participativo. En línea con esta crítica, William Easterly (2014), en su obra *La tiranía de los expertos*, expone cómo el enfoque tecnocrático, al centrarse exclusivamente en el conocimiento experto, puede ignorar las necesidades y deseos de las personas afectadas, con lo cual socava las verdaderas soluciones participativas y sostenibles.

Casos concretos del criterio de experto se pueden observar tanto en mediciones para ciudades como para grupos de países. Por ejemplo, el distrito de Medellín, Colombia, lanzó en el 2024 el nuevo índice multidimensional de calidad de vida sin la inclusión de procesos participativos que involucren a los ciudadanos. En este caso, la selección de las dimensiones de análisis ha sido una adaptación de recomendaciones de expertos de estudios europeos. Con relación al enfoque de expertos, Pacáková (2012) describe cómo en la Unión Europea un conjunto de variables predefinidas fue reducido y validado por expertos europeos para seleccionar indicadores de calidad de vida. Por su parte, Boarini (2015) examina el trabajo de la OCDE, donde comités de expertos desarrollaron indicadores que integran dimensiones económicas, sociales y ambientales, ampliando el análisis del bienestar más allá del PIB. Estos enfoques, aunque basados en el criterio de experto, buscan aportar para que las mediciones de sus contextos sean más robustas.

Cabe señalar que la importancia del enfoque participativo también ha sido respaldada por estudios que validan su eficacia en contextos específicos. Por ejemplo, Biggeri et ál. (2006)

identificaron capacidades clave en el bienestar infantil, a través de estudios participativos, con lo cual revelaron que las dimensiones pueden variar según la edad o el contexto. De manera similar, el trabajo de Bessell (2020) evidenció una disparidad de género significativa en las dimensiones valoradas por las comunidades pobres, como los alimentos, la salud y la seguridad, al involucrar a los propios individuos en el proceso de identificación.

En este sentido, la investigación cualitativa ha desempeñado un papel fundamental en los estudios sobre la pobreza, el bienestar y la calidad de vida, a través del uso de entrevistas y grupos focales para capturar aspectos subjetivos y contextuales que los métodos cuantitativos no pueden abordar completamente (Camfield et ál., 2008; Ruiz, 2015). Estos enfoques cualitativos proporcionan una interpretación más profunda y detallada de las experiencias humanas, a la par que ofrecen una visión rica que es difícil de lograr con enfoques puramente cuantitativos. Sin embargo, la llegada del *big data* y el procesamiento del lenguaje natural (NLP, por sus siglas en inglés) ha transformado las ciencias sociales y ha permitido el análisis de grandes volúmenes de datos que antes resultaban imposibles de manejar con métodos tradicionales (Grim et al., 2006).

Por ejemplo, Mahalakshmi et al. (2017) destacan cómo el *big data* ha mejorado tanto el tiempo como la precisión en los análisis, al superar las limitaciones de los enfoques tradicionales. De manera similar, Hassani et ál. (2020) abordan el uso de técnicas de minería de textos y extracción de entidades mediante NLP para organizar grandes volúmenes de datos digitales no estructurados y, de esta manera, muestran cómo estas herramientas superan las limitaciones de los métodos convencionales. Además, autores como Caicedo-Beltrán y Rubiano (2023) destacan que el uso de tecnologías de Big Data permite monitorizar y predecir tendencias sociales, facilitando la integración de múltiples perspectivas en el análisis de datos. Asimismo, Taboada Villamarín (2024) señala que la automatización del análisis de datos de texto mediante procesamiento de lenguaje natural y aprendizaje automático puede complementar los enfoques participativos al manejar grandes volúmenes de información de manera eficiente.

Así mismo, técnicas avanzadas como el modelado de temas han facilitado el análisis de grandes volúmenes de datos no estructurados, como los generados en redes sociales o encuestas masivas en línea. Investigaciones como las de Rodríguez y Storer (2020) y Chidi et ál. (2043) han demostrado que el modelado de temas mediante algoritmos como Latent Dirichlet Allocation (LDA) facilita la extracción de temas relevantes y guía un análisis cualitativo más profundo. Esto permite integrar el análisis cuantitativo del lenguaje natural con un enfoque interpretativo cualitativo, lo que enriquece el análisis de fenómenos complejos. Además, los avances en el uso de redes neuronales recurrentes, como los presentados por Gupta et ál. (2017), han permitido analizar la evolución de temas a lo largo del tiempo en grandes colecciones de documentos, mostrando su relevancia para el estudio de tendencias en ciencias sociales. Igualmente, Nikolenko et ál. (2017) subrayan que el modelado de temas complementa las metodologías

cualitativas tradicionales, lo que facilita el análisis de grandes cantidades de texto no estructurado. Esto evidencia que la combinación de métodos cualitativos y cuantitativos amplía el alcance del análisis y transforma la capacidad de abordar fenómenos multidimensionales.

Esta integración entre enfoques cualitativos y cuantitativos ha enriquecido los análisis en ciencias sociales, al mismo tiempo que ha ofrecido una visión más holística de los problemas estudiados. Busetto et ál. (2020) destacan que este enfoque mixto incrementa la transparencia y la coherencia de los resultados, lo que permite que las ciencias sociales avancen significativamente en la comprensión y medición de fenómenos complejos. Al combinar la profundidad cualitativa con el alcance cuantitativo que proporciona el *big data*, se ha logrado un desarrollo notable de la investigación sobre fenómenos como la pobreza y la calidad de vida.

Metodología

Datos

Los datos utilizados en esta investigación provienen de la encuesta anual sobre calidad de vida realizada en 2021 por la administración pública de Medellín, Colombia. Esta encuesta abarcó 18 áreas temáticas relacionadas con la calidad de vida y su principal objetivo fue orientar las políticas públicas locales. Los participantes fueron seleccionados por su capacidad para proporcionar información relevante sobre sus hogares, lo que los calificó como informantes cualificados.

La encuesta incluyó dos preguntas abiertas opcionales sobre la definición de calidad de vida y el bienestar, que se ubicaron al final del cuestionario. Estas preguntas buscaban recopilar las percepciones personales de los participantes sobre estos conceptos. De los 9400 participantes en la encuesta, 849 respondieron estas preguntas abiertas. Las respuestas fueron recolectadas a través de un centro de llamadas, registradas y transcritas por asistentes de investigación.

En cuanto a la composición demográfica de los participantes que respondieron las preguntas abiertas, el 77 % eran mujeres y el 23 % hombres, lo cual es consistente con las expectativas en encuestas de hogares, donde las mujeres suelen asumir un papel más destacado en la respuesta a este tipo de cuestionarios. La edad promedio de los encuestados fue de 55 años, con una desviación estándar de 16 años. La distribución por estratos socioeconómicos mostró que el 53.9 % de los participantes eran los niveles más bajos, el 31.9 % en el nivel tres y el 14% en los niveles superiores, lo que refleja una composición socioeconómica que se alinea con la población de la ciudad de Medellín.

Preprocesamiento de los datos

El preprocesamiento de los datos textuales asegura que los textos tengan una forma adecuada para el análisis. El proceso incluyó los siguientes pasos:

- **Tokenización:** este proceso implicó dividir cada documento en unidades básicas llamadas *tokens*, que generalmente corresponden a palabras individuales. Este paso es fundamental para facilitar el análisis posterior, ya que permite contar y analizar cada palabra de manera independiente (Bird et ál., 2009). Por ejemplo, en un documento donde se menciona “El acceso a la educación es fundamental”, la tokenización dividiría esta frase en palabras individuales como “El”, “acceso”, “a”, “la”, “educación” y “es”, lo que facilita su análisis.
- **Minimización:** para asegurar uniformidad en el análisis, todas las palabras se convirtieron a minúsculas. Esto evita que la misma palabra sea tratada como diferentes *tokens* debido a la variación en mayúsculas o minúsculas. Por ejemplo, “Salud” y “salud” se considerarían el mismo *token* después de la minimización (Jurafsky y Martin, 2019).
- **Eliminación de palabras vacías:** se eliminaron palabras que no aportan información significativa al análisis, conocidas como “*stop words*”, como preposiciones, artículos y algunas conjunciones. Este paso es crucial para reducir el ruido en los datos y centrarse en los términos más significativos (Manning et ál., 2008). Por ejemplo, palabras como “y”, “de” y “en”, aunque comunes, no contribuyen directamente al significado en este contexto.
- **Eliminación de dígitos y símbolos:** se eliminaron dígitos numéricos, signos de puntuación y palabras que contenían tres o más letras repetidas, ya que estas no aportan valor semántico al análisis del texto. Esto asegura que el análisis se centre en palabras y frases que realmente importan para entender la calidad de vida.
- **Construcción de bigramas:** además del análisis de palabras individuales, se generaron bigramas, que son pares de palabras consecutivas. Se realiza el proceso para capturar relaciones importantes entre términos que suelen aparecer juntos (Jurafsky y Martin, 2019) y, así, mejorar la interpretación del contexto. Por ejemplo, en el análisis de calidad de vida, bigramas como “calidad de vida” o “seguridad alimentaria” son fundamentales para entender conceptos clave que podrían perderse si se consideraran solo las palabras individuales. El uso de bigramas permite que el análisis sea más profundo y que se identifiquen combinaciones de palabras que, juntas, tienen un significado particular en el contexto del estudio de la calidad de vida. Por ejemplo, “calidad de vida” como bigrama capta un concepto que es crucial en las entrevistas y que no sería completamente identificado si solo se consideraran “calidad” y “vida” por separado.

Métodos

Conteo de términos

Este método básico se fundamenta en el conteo de la frecuencia de aparición de palabras y frases en el corpus textual. Las palabras más frecuentes proporcionan una indicación inicial de los temas más importantes para los participantes. Por ejemplo, si en una serie de entrevistas se menciona con frecuencia “educación” y “salud”, esto indicaría que se trata de temas de gran importancia para los entrevistados (Manning et ál., 2008). El análisis se basa en representaciones gráficas de los conteos y los bigramas.

Modelado correlacionado de temas

El modelado correlacionado de temas (CTM, por sus siglas en inglés) es una técnica de modelado estadístico que agrupa términos que suelen aparecer juntos en el mismo contexto. Este enfoque es útil para identificar temas latentes en los textos que no son evidentes solo con el conteo de palabras. Por ejemplo, si las palabras “empleo”, “ingresos” y “estabilidad” aparecen frecuentemente juntas, el CTM podría identificar un tema relacionado con la seguridad económica. Esta técnica es particularmente útil en ciencias sociales para descubrir relaciones complejas entre conceptos que están implícitos en las conversaciones, con lo cual ofrece una comprensión más profunda de las preocupaciones de los participantes (Blei y Lafferty, 2007).

La formulación matemática se presenta como sigue:

Sea z_d un vector latente que sigue una distribución normal multivariada:

$$z_d \sim N(\mu, \Sigma)$$

Donde μ es el vector de medias y Σ es la matriz de covarianzas.

El vector de proporciones de temas θ_d en el documento d se obtiene aplicando una transformación logística a z_d :

$$\theta_d = \text{soft max}(z_d)$$

Donde

$$\text{soft max}(z_{di}) = \frac{e^{z_{di}}}{\sum_{j=1}^K e^{z_{dj}}}$$

Aquí, K es el número de temas y z_{di} es el valor del i -ésimo elemento del vector z_d .

En términos de la generación de palabras, para cada palabra w_{dn} en el documento d se selecciona un tema k basado en la distribución θ_d y se genera la palabra a partir de la distribución de palabras β_k :

$$w_{dn} \sim \text{Categorical}(\beta_k)$$

Donde β_k es la distribución de probabilidad sobre el vocabulario para el tema k .

Resultados

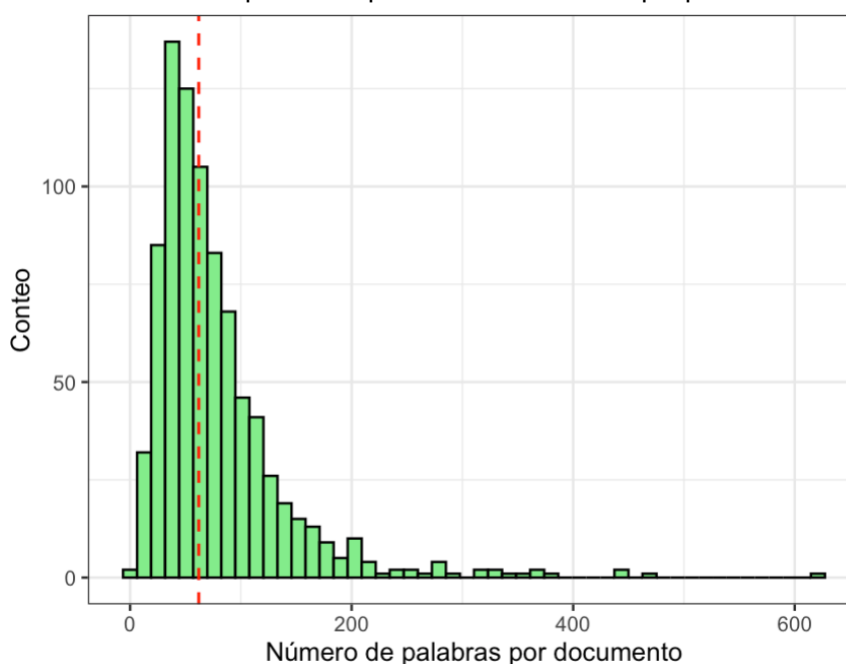
El análisis de los datos obtenidos en el experimento de calidad de vida se llevó a cabo utilizando el método descriptivo de conteo de términos y el CTM. A continuación se describen en detalle los resultados obtenidos con cada uno de estos métodos.

Conteo de términos

El corpus analizado cuenta con 849 documentos, los cuales representan respuestas individuales que varían en extensión. La longitud promedio de estos documentos es de 62 palabras, con una mediana idéntica, lo que sugiere una distribución relativamente homogénea en términos de extensión. Sin embargo, la desviación estándar de 61.7 palabras indica una cierta variabilidad, con documentos que van desde un mínimo de 4 hasta un máximo de 625 palabras. Esta variabilidad se refleja en la distribución asimétrica observada, donde la mayoría de los documentos son breves, concentrándose entre 10 y 100 palabras (figura 1). Después del preprocesamiento, que incluyó la eliminación de palabras vacías y otros elementos irrelevantes, la mediana se redujo a 21 palabras por respuesta, con un rango final entre 1 y 251 palabras. Es importante destacar que en este recuento no se incluyeron las palabras pronunciadas por los entrevistadores y, además, se eliminaron cinco muestras que contenían menos de cuatro palabras del conjunto de datos.

Figura 1

Nota: histograma del conteo de palabras por documentos sin preprocesar

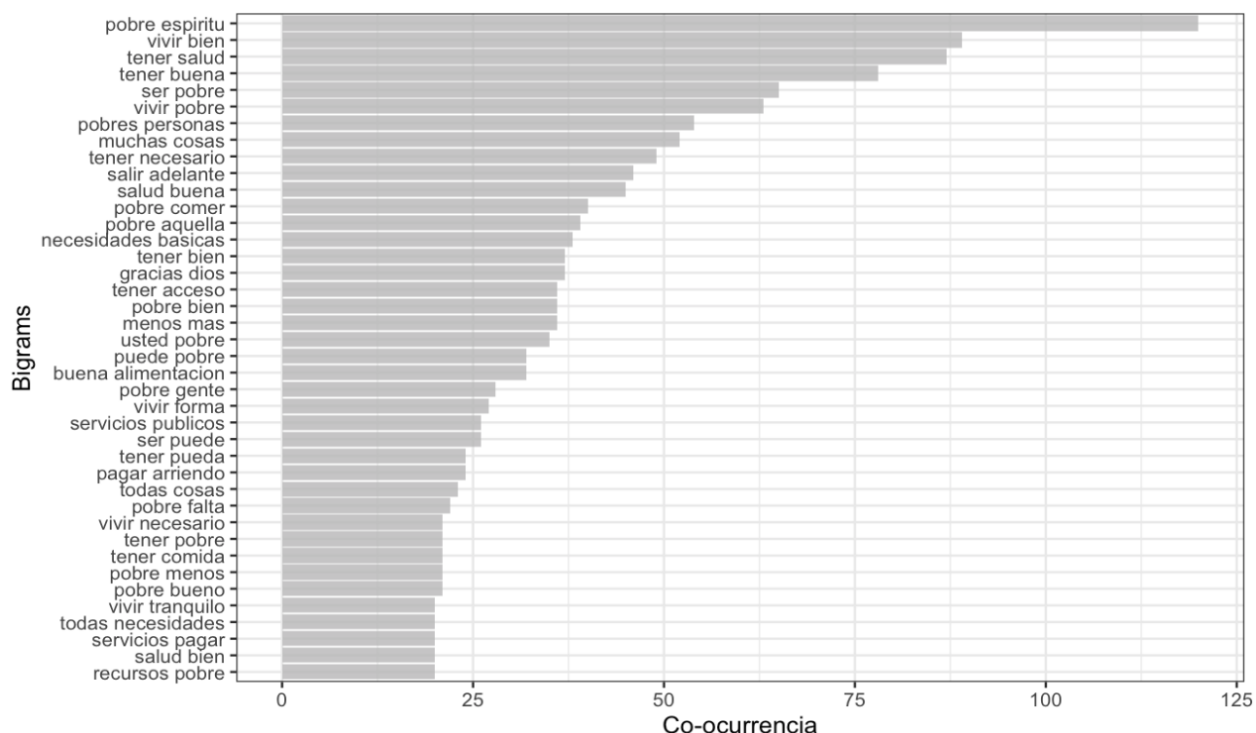


Fuente: elaboración propia.

El gráfico de coocurrencia (figura 2) de bigramas proporciona una representación visual de las combinaciones de palabras que aparecen con mayor frecuencia en el corpus. En el eje y se listan los bigramas más comunes, mientras que el eje x muestra la frecuencia con la que estos aparecen en el corpus. Entre los bigramas más destacados se encuentran “pobre espíritu”, “vivir bien” y “tener salud”, todos con altas frecuencias de coocurrencia, lo que indica que son combinaciones de palabras recurrentes en las respuestas. Estos bigramas sugieren temas centrales como la espiritualidad en la pobreza, la calidad de vida y la salud, que son fundamentales en el corpus. Además, bigramas como “ser pobre”, “pobres personas” y “salir adelante” también muestran frecuencias elevadas, lo que destaca las preocupaciones sobre la pobreza y la superación personal entre los encuestados. La figura confirma que los conceptos relacionados con la pobreza, el bienestar y las necesidades básicas son prominentes en el conjunto de datos.

Figura 2

Nota: términos o bigramas más frecuentes en el corpus.³



Fuente: elaboración propia

El gráfico tipo cadena de Markov (figura 3) muestra las relaciones entre los términos más frecuentes en el corpus, cuya conexión se destaca a través de los bigramas. En este gráfico, los nodos representan palabras individuales y las flechas indican la dirección de la relación dentro de

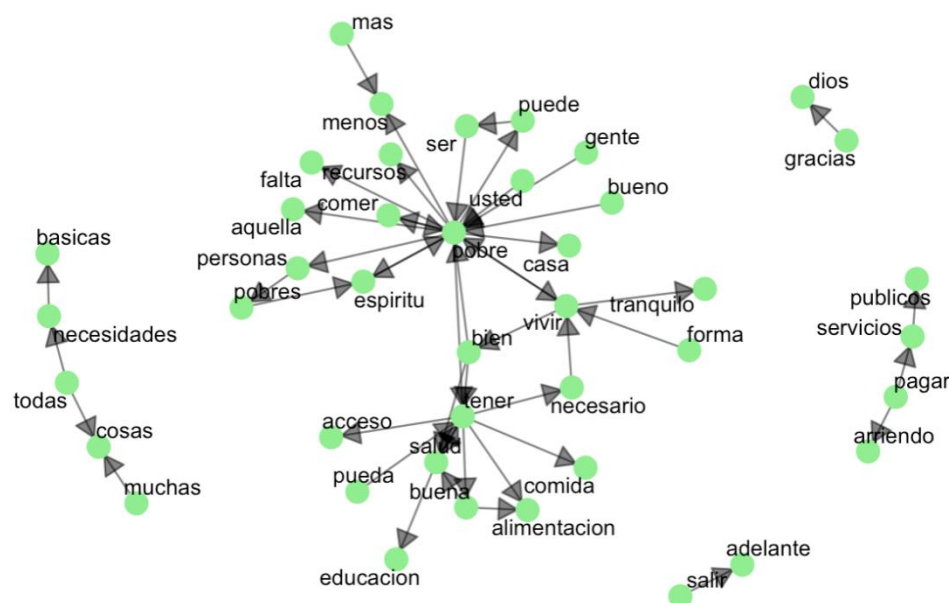
³ El preprocesamiento de texto implica la estandarización de los datos, quitando los marcadores ortográficos, por eso es normal que el gráfico no tenga tildes.

los bigramas. El tamaño de los nodos refleja la frecuencia relativa de las palabras en el corpus, mientras que las conexiones indican la coocurrencia entre palabras.

En el centro de la figura, el nodo “pobre” actúa como un punto de convergencia, lo que muestra múltiples conexiones hacia y desde otros términos clave como “espíritu”, “tener”, “bien” y “personas”. Esto sugiere que “pobre” es un término central que se combina frecuentemente con otros conceptos, formando bigramas como “pobre espíritu” y “pobre bien”. Otros términos como “salud”, “vivienda” y “necesidades” también aparecen conectados, lo que sugiere que las discusiones sobre la pobreza se relacionan con preocupaciones sobre la salud, el acceso a bienes básicos y la estabilidad diaria. La figura también destaca subgrupos de términos que forman cadenas más aisladas, como “servicios públicos”, “pagar arriendo” o “gracias a Dios”. Estas conexiones reflejan temas específicos que, aunque relacionados con la cuestión central de la pobreza, se presentan como conceptos autónomos dentro del discurso del corpus. En conjunto, este gráfico ilustra cómo los términos se relacionan en el corpus, destacando la centralidad de ciertos conceptos y la estructura subyacente de las discusiones sobre la pobreza y el bienestar.

Figura 3

Nota: gráfico de red tipo cadena de Markov de los términos más comunes.⁴



Fuente: elaboración propia

⁴ El preprocesamiento de texto implica la estandarización de los datos, quitando los marcadores ortográficos, por eso es normal que el gráfico no tenga tildes.

El conteo de términos reveló los bigramas más frecuentes en las respuestas de los participantes. En este sentido, es una primera aproximación a las dimensiones más relevantes en su percepción de calidad de vida. Entre los bigramas más comunes se encontraron “vivir bien”, “salud tener”, “estar alimentado” y “necesidades básicas”. Estos resultados sugieren que los participantes priorizan la satisfacción de necesidades básicas y el bienestar general como componentes cruciales de su calidad de vida. Además, se identificó la coocurrencia de términos específicos, lo que permitió descubrir patrones importantes en la percepción de los encuestados. Por ejemplo, los bigramas relacionados con “vivir bien” y “salud” mostraron una alta frecuencia de coocurrencia, lo que sugiere una fuerte asociación entre una buena calidad de vida y un estado de salud favorable. De manera similar, los términos “empleo” y “vivienda” se mencionaron juntos en varias respuestas, lo que refleja la importancia de la estabilidad económica y la seguridad habitacional en la percepción de calidad de vida de los participantes.

Modelado correlacionado de temas (CTM)

Selección del número de temas (k)

Para determinar el número óptimo de temas (k) en el modelo de temas correlacionados (CTM), se evaluaron varias opciones utilizando métricas de perplejidad y coherencia. En este análisis se exploraron valores de k que variaban de 3 a 15, con el fin de identificar el número de temas que mejor captura las estructuras latentes en el corpus.

La perplejidad, una métrica ampliamente utilizada para evaluar modelos de lenguaje, mostró una disminución consistente a medida que aumentaba el número de temas. Aquella métrica, que computa la incertidumbre en la predicción de una palabra en un texto dado el modelo, se redujo de 4.640 478 con $k = 3$ a 1.495 558 con $k = 15$. Esta disminución sugiere que el modelo se ajusta mejor al corpus a medida que se incrementa el número de temas, ya que la perplejidad más baja indica un modelo más predecible y, por lo tanto, más eficaz. Este comportamiento es común en el análisis de temas, donde un mayor número generalmente mejora el ajuste del modelo al capturar más matices dentro del corpus. Sin embargo, también es importante notar que, aunque la reducción de la perplejidad sugiere un mejor ajuste, un número excesivo de temas puede llevar a una fragmentación temática, lo que hace que los temas sean menos significativos y más difíciles de interpretar en un contexto global.

La coherencia de los temas en el modelo se evaluó utilizando dos métricas: Griffiths2004 y Deveaud2014. La métrica Griffiths2004, desarrollada por Griffiths y Steyvers (2004), mide la calidad de la estructura temática de un modelo de temas a través de la probabilidad conjunta de los datos observados dado el modelo. Los valores obtenidos con esta métrica aumentaron, es decir, se volvieron menos negativos a medida que se incrementaba el número de temas (k). Este

patrón sugiere que la estructura temática del modelo mejora con un mayor número de temas, lo que indica una mayor coherencia interna en la representación de los temas a medida que el modelo incorpora más dimensiones.

Por su parte, la métrica Deveaud2014, propuesta por Deveaud et ál. (2014), evalúa la coherencia temática basada en la proximidad semántica entre los términos más probables de cada tema. En este caso, los valores de coherencia también mostraron una tendencia de mejora conforme aumentaba k . Específicamente, los valores de Deveaud2014 fueron más bajos (indicando mejor coherencia) cuando había menos temas y aumentaron gradualmente al incrementarse k . Esta tendencia sugiere que, aunque un número menor de temas puede capturar relaciones más claras entre términos, la adición de más temas puede revelar matices adicionales en los datos, lo que mejora la interpretación general del modelo.

Ahora bien, dado el objetivo de capturar aspectos comunes en un corpus que carece de variabilidad temática significativa, se seleccionó un valor de $k = 4$. Este valor se consideró óptimo porque permite identificar temas diferenciados para capturar aspectos relevantes del corpus, sin fragmentar los datos en temas específicos, lo que podría ser contraproducente por la homogeneidad del contenido. A pesar de que los valores más altos de k muestran una mejoría en las métricas de perplejidad y coherencia, la elección de $k = 4$ ofrece un equilibrio adecuado. Con estos temas, se mantiene la interpretabilidad del modelo, destacando los patrones más significativos en el corpus, lo cual es crucial en contextos donde se debe resaltar temas prevalentes en un entorno de baja variabilidad temática. Con k bajo, el modelo se enfoca en identificar los temas más amplios y presentes en el corpus, sin forzar la creación de muchos temas que podrían no ser claramente distinguibles en un corpus con alta similitud.

Resultados de modelo CTM

El análisis realizado mediante el modelo CTM con $k = 4$ revela dimensiones que estructuran las percepciones de los encuestados sobre la pobreza y la calidad de vida. Los temas identificados destacan preocupaciones específicas, al tiempo que reflejan aspectos transversales que impactan de manera integral el bienestar de las personas. En la tabla 1, se detallan los temas principales, las proporciones esperadas de cada uno en el corpus, los términos más representativos y los *aspectos más destacados* derivados de ellos. Este análisis permite identificar dimensiones críticas como las necesidades básicas, la vulnerabilidad material y el bienestar espiritual, elementos esenciales para comprender las experiencias de pobreza y para informar el desarrollo de políticas públicas más efectivas.

Tabla 1*Resultados resumidos por tema identificados con CTM*

| Tema | Proporción esperada del tema | Siete términos con mayor probabilidad de pertenecer al tema | Aspectos destacados |
|---------------|------------------------------|---|---|
| Tema 1 | 0.269 | pobre_ser, buena_tener, salud_tener, comer_pobre, bien_pobre, necesario_tener, más_menos | Enfocado en las necesidades básicas: este tema refleja preocupaciones sobre la satisfacción de las necesidades básicas, como la salud y la alimentación en el contexto de la pobreza. Los términos sugieren una visión de la pobreza en lo que respecta a carencias materiales. |
| Tema 2 | 0.256 | espíritu_pobre, bien_vivir, personas_pobres, adelante_salir, necesario_tener, cosas_muchas, necesario_pobre | Dimensión espiritual y bienestar: aquí se destaca la importancia de la espiritualidad y el esfuerzo por mejorar las condiciones de vida. Los términos apuntan a una narrativa de superación personal y comunitaria en la que el bienestar no es solo material, sino también espiritual. |
| Tema 3 | 0.271 | buena_tener, bien_vivir, salud_tener, pobre_espíritu, personas_pobres, pobre_vivir, alimentación_buena | Calidad de vida y salud: este tema se centra en la relación entre tener lo necesario y vivir bien. Incluye una combinación de aspectos materiales y espirituales, subrayando que la calidad de vida está intrínsecamente ligada a la salud y el bienestar integral. |
| Tema 4 | 0.204 | pobre_vivir, salud_tener, bien_vivir, pobre_usted, básicas_necesidades, buena_tener, pobre_espíritu | Lucha cotidiana y necesidades básicas: este tema subraya la constante lucha por satisfacer las necesidades básicas, reflejando las dificultades de vivir en condiciones de pobreza. Los términos sugieren un enfoque en la vida diaria de las personas, destacando la importancia de acceder a lo esencial para mantener una vida digna. |

Fuente: elaboración propia.

El Tema 1, con un gamma promedio de 0.269, es uno de los más predominantes en el corpus. Los términos más frecuentes, como “pobre_ser”, “buena_tener”, “salud_tener” y “comer_pobre”, reflejan una fuerte preocupación por las necesidades básicas. La alta frecuencia de estos términos dentro del tema, que acumulan un total de 0.0315, sugiere que los encuestados ven la pobreza principalmente a través de la lente de la carencia material y la incapacidad de satisfacer necesidades fundamentales como la alimentación y la salud. Este tema destaca la vulnerabilidad material como una dimensión central en la percepción de la pobreza, donde la falta de recursos esenciales expone a las personas a riesgos significativos en su bienestar.

El Tema 2 tiene un gamma promedio de 0.256, ligeramente inferior al Tema 1, pero aun así relevante. Este tema resalta la importancia del bienestar espiritual y la superación personal en

la lucha contra la pobreza. Términos como “pobre_espíritu”, “bien_vivir” y “adelante_salir” indican que la espiritualidad y el esfuerzo por mejorar las condiciones de vida son dimensiones cruciales para los encuestados. La frecuencia total de los términos en este tema es de 0.0367, lo que indica un peso significativo de estas dimensiones en las percepciones de los participantes. Además, la aparición de términos como “pobre_ser” y “necesario_pobre” en múltiples temas refuerza la transversalidad de estas preocupaciones en el corpus.

El Tema 3, con un gamma promedio de 0.271, es el más predominante en el corpus. La calidad de vida y la salud integral son las dimensiones clave aquí, con términos como “buena_tener”, “bien_vivir”, “salud_tener” y “alimentacion_buena”, que aparecen con alta frecuencia. La suma total de las frecuencias de los términos más representativos es de 0.0381, lo que subraya la importancia de tener acceso a los recursos necesarios para asegurar un bienestar físico y emocional. La presencia reiterada de términos relacionados con la salud y el bienestar en todos los temas también sugiere que estos son aspectos fundamentales en la definición de calidad de vida para los encuestados.

Por su parte, el Tema 4 tiene un gamma promedio de 0.204, lo que lo convierte en el menos prevalente de los cuatro, pero aun así es importante. Este tema se enfoca en la lucha cotidiana para satisfacer las necesidades básicas en condiciones de pobreza, con términos como “pobre_vivir”, “salud_tener” y “básicas_necesidades” que dominan el tema. La frecuencia total de los términos principales es de 0.0495, la más alta entre los cuatro temas, lo que indica una fuerte preocupación por la supervivencia diaria y el acceso a recursos esenciales. La presencia de términos como “necesidades_todas” y “pobre_usted” refuerza la idea de que la lucha por sobrevivir es una experiencia común y compartida entre los encuestados.

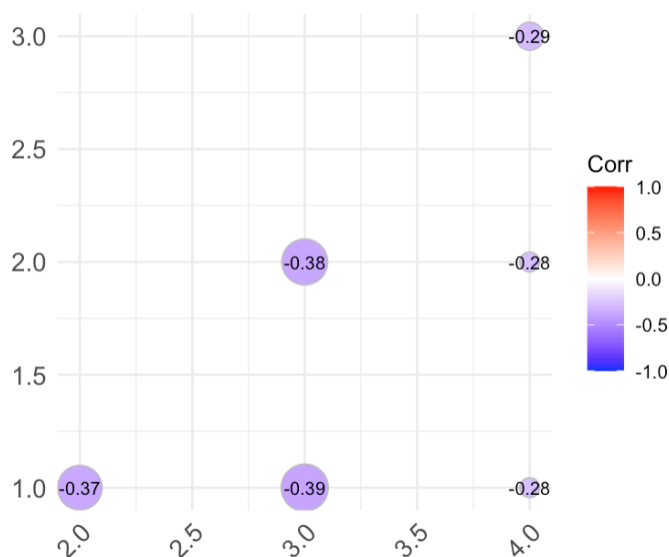
El análisis de los términos repetidos en múltiples temas revela la centralidad de ciertos conceptos en las percepciones de los encuestados. Términos como “bien_vivir” (frecuencia total de 0.0193) y “pobre_espíritu” (frecuencia total de 0.0170) aparecen en los cuatro temas, lo que sugiere que la calidad de vida y la dimensión espiritual son preocupaciones transversales que atraviesan todas las percepciones de la pobreza. Esto refuerza la idea de que la pobreza no es vista solo como una carencia material, sino también como una condición que afecta el bienestar integral y la espiritualidad de las personas.

Para identificar temas relacionados, se utiliza un gráfico de correlación. La figura 4 muestra un mapa de calor que visualiza las correlaciones entre los temas identificados mediante el modelo CTM. Las correlaciones representadas en este gráfico son en su mayoría negativas, lo que sugiere que los temas tienden a estar distanciados o en oposición dentro del espacio discursivo del corpus. La correlación más negativa se observa entre el Tema 1 y el Tema 3 (-0,39), lo cual sugiere que estos temas raramente coexisten en los documentos y podrían estar

vinculados a conceptos o dimensiones distintas en las percepciones de los encuestados. Otras correlaciones, como la existente entre los Temas 2 y 3 (-0,38), también son negativas, lo que refuerza la idea de que estos temas reflejan aspectos discutidos de forma separada por los encuestados. Este patrón de correlaciones negativas indica que los temas están claramente diferenciados en el corpus, un resultado esperado cuando los temas representan diferentes facetas de un fenómeno complejo como la pobreza.

Figura 4

Correlación entre temas identificados con CTM



Fuente: elaboración propia.

De acuerdo con el análisis de las relaciones entre los temas, si se consideran tanto las correlaciones positivas como negativas, las conexiones más fuertes se observan entre el Tema 1 y el Tema 3, así como entre el Tema 1 y el Tema 2, con correlaciones cercanas a -0,36. Esto sugiere que estos temas están en mayor oposición en cuanto a cómo los encuestados perciben la pobreza y la calidad de vida. El Tema 4 también presenta una correlación moderada con los otros temas, aunque su fuerza de correlación es menos prominente en comparación con los demás. Las correlaciones negativas observadas entre los temas sugieren que cada uno de ellos podría estar capturando dimensiones distintas de la experiencia de pobreza y calidad de vida, lo que puede resultar útil para comprender mejor las diferentes perspectivas presentes en el corpus.

Discusión

El experimento sobre la calidad de vida, basado en el análisis de corpus textuales derivados de entrevistas, puso de relieve aspectos clave tanto en los logros como en las limitaciones del enfoque metodológico empleado. Uno de los hallazgos más significativos fue la identificación de las dimensiones más valoradas por los participantes, asociadas a bigramas centrales como “vivir bien”, “salud”, “pobre espíritu” y “necesidades básicas”. El uso de técnicas avanzadas, como el CTM, resultó fundamental para desentrañar patrones conceptuales complejos en los discursos de los encuestados, lo que proporcionó una visión más matizada de los desafíos percibidos en torno a la pobreza y el bienestar (Alkire et ál., 2015; Griffiths y Steyvers, 2004). A través del CTM se identificaron temas latentes clave sobre la percepción de la pobreza, el bienestar y la calidad de vida, lo que fue particularmente útil para descubrir patrones subyacentes que no eran evidentes mediante el simple análisis de palabras o bigramas, tal como destacan Blei y Lafferty (2007) en sus estudios científicos. Ello permite captar matices contextuales que los indicadores tradicionales podrían pasar por alto (Hatef et ál., 2022). A través del CTM se descubren patrones asociados a las discusiones sobre las dimensiones ocultas de la pobreza y el bienestar (Alkire, 2007; Andreoni et ál., 2021; Mills, 2014).

El análisis identificó dimensiones relacionadas con las necesidades básicas, la vulnerabilidad material, el bienestar espiritual, la superación personal, la calidad de vida, la salud integral y la lucha cotidiana. Dichas dimensiones emergieron de los datos en lenguaje natural proporcionados por los encuestados. Estas dimensiones subrayan la importancia de la percepción subjetiva en la calidad de vida de la población estudiada. Los resultados revelan temas que van desde la falta de recursos esenciales hasta el bienestar espiritual, lo que refleja la complejidad multidimensional del bienestar (Diener et ál., 2018; Narayan et ál., 2000; Acosta, 2013).

La correlación negativa entre los temas identificados mediante CTM sugiere que las percepciones de los encuestados sobre la pobreza y la calidad de vida están diferenciadas. Esta separación temática podría indicar la necesidad de políticas diferenciadas para abordar cada dimensión de la pobreza. Por ejemplo, la correlación negativa entre el Tema 1, centrado en la carencia de recursos materiales, y el Tema 3, enfocado en el bienestar integral, sugiere que ambos aspectos son percibidos como experiencias separadas. Esto refuerza la idea de que las políticas públicas deben atender tanto las necesidades materiales como las subjetivas. En este sentido, los resultados siguen el objetivo de incluir dimensiones no materiales en la evaluación del bienestar, invocando una perspectiva más holística (Stiglitz et ál., 2009; Diener et ál., 2018, 2009; OECD, 2013).

El uso de métodos basados en procesamiento de lenguaje natural (NLP) es crucial para captar las percepciones y voces de las personas en contextos de pobreza y bienestar, ya que

permite representar con mayor precisión sus experiencias en las políticas públicas. Como señalan Narayan et ál. (2000), escuchar directamente a los afectados revela dimensiones que van más allá de los indicadores tradicionales y que abordan aspectos subjetivos y contextuales que los encuestados consideran vitales para su calidad de vida (Creswell y Poth, 2017). Este enfoque permite identificar necesidades y aspiraciones mediante términos y conceptos espontáneos que reflejan tanto preocupaciones materiales, como la salud o la alimentación, como aspectos más intangibles, como el bienestar espiritual, en línea con el llamado a incorporar el bienestar subjetivo (Acosta, 2013; Diener et ál., 2018). Al integrar estas voces en la formulación de políticas, como promueve el índice de pobreza multidimensional (IPM) de Alkire y Foster (2011), se garantiza una aproximación más inclusiva y adaptada a las realidades locales (Alkire et ál., 2015).

También se subraya la importancia de considerar las particularidades contextuales y regionales al diseñar políticas públicas. Como argumentan Fleurbaey y Blanchet (2013), las intervenciones políticas que se ajustan a las necesidades locales tienden a ser más eficaces. En el caso de Medellín, la centralidad de temas como “vivir bien” y “salud” en las percepciones de calidad de vida sugiere que las políticas públicas deben centrarse en mejorar la seguridad y el acceso a servicios de salud de calidad, tal como lo recomienda el Programa de las Naciones Unidas para el Desarrollo (PNUD, 2010). En este sentido, los resultados se alinean con las posturas teóricas y políticas que consideran el bienestar, la calidad de vida y la pobreza como fenómenos multidimensionales (Sen, 1999; Narayan-Parker et ál., 2000; Wisor et ál., 2014; Stiglitz et ál., 2009; Whelan et ál., 2014; Atkinson, 2019; Ferreira, 2011; Beck et ál., 2020; PNUD, 2010; Ravallion, 2011; Alkire et ál., 2015; Nolan y Whelan, 2011; World Bank, 2017, 2022; UNDP y OPHI, 2021). Todo lo anterior en consonancia con los ODS, donde la perspectiva multidimensional es fundamental.

Conclusiones y recomendaciones

Los resultados del modelo no supervisado de CTM revelan cuatro temas clave que destacan la complejidad del bienestar en contextos de pobreza, en los que se combinan dimensiones tanto materiales como espirituales. El tema más prominente (27.1 %) subraya la relevancia de la calidad de vida y la salud, lo que pone en evidencia que tanto los recursos básicos como el bienestar integral son esenciales para una vida digna. Otros temas, como las necesidades básicas (26.9 %) y la lucha cotidiana (20.4 %), enfatizan la falta de recursos fundamentales como la alimentación y el acceso a la salud, mientras que el tema relacionado con la superación personal y espiritual (25.6 %) resalta la importancia del bienestar emocional y comunitario en la mejora de las condiciones de vida. Un hallazgo relevante de este estudio es la alta valoración que los participantes otorgan al bienestar emocional, lo que sugiere que las

políticas públicas deben ir más allá de los indicadores económicos tradicionales y adoptar un enfoque más holístico, que incluya tanto indicadores objetivos como subjetivos. Hay que reconocer que la percepción individual del bienestar es un componente crítico para la calidad de vida que resulta esencial en este contexto.

El análisis de la correlación entre temas reveló un patrón interesante: las correlaciones negativas sugieren que los encuestados perciben la pobreza y la calidad de vida a través de dimensiones claramente diferenciadas. En particular, la correlación más negativa, observada entre el Tema 1 y el Tema 3 (-0.39), indica que las preocupaciones relacionadas con las necesidades básicas y la calidad de vida se perciben como aspectos separados. Este hallazgo es crucial para el diseño de intervenciones políticas, ya que sugiere que cada dimensión capturada en los temas requiere un enfoque específico. La diferenciación temática evidenciada por esta separación en las percepciones refuerza la necesidad de políticas diversificadas que aborden de manera efectiva las distintas facetas de la pobreza y promuevan el bienestar integral de las comunidades afectadas.

La presente investigación introduce una metodología semiautomática que combina el uso de inteligencia artificial y procesamiento del lenguaje natural (NLP) para identificar dimensiones valoradas por las personas en relación con la calidad de vida y la pobreza. Este enfoque no solo permite extraer aspectos clave del discurso público, sino que ofrece una comprensión más matizada de las percepciones sobre la calidad de vida, con lo cual supera las limitaciones de los enfoques tradicionales. Como sugieren los estudios previos, tales como los de Bird et ál. (2009) y Manning et ál. (2008), el uso de tecnologías avanzadas facilita la captura de grandes volúmenes de datos textuales, lo que habilita a los responsables de políticas públicas a formular intervenciones más contextualizadas y alineadas con las necesidades territoriales. Además, este enfoque podría complementarse con métodos cualitativos tradicionales, como entrevistas en profundidad o grupos focales, para explorar con mayor detalle las experiencias de los participantes, lo que fortalecería aún más la visión holística del bienestar. La posibilidad de aplicar esta metodología en distintos contextos socioculturales sugiere su flexibilidad, lo que permitiría observar cómo varían las percepciones a lo largo del tiempo frente a eventos socioeconómicos significativos. Al combinar enfoques cualitativos y cuantitativos, este estudio abre el camino para una creación de políticas públicas más inclusivas y adaptadas a las necesidades reales de las comunidades (Fleurbaey y Blanchet, 2013).

Las implicaciones de estos hallazgos para la formulación de políticas públicas son profundas. En primer lugar, destacan la necesidad de diseñar políticas que vayan más allá de los indicadores económicos, integrando las percepciones subjetivas de los ciudadanos. En segundo lugar, subrayan la importancia de adaptar las políticas a las necesidades y prioridades locales, lo que podría mejorar significativamente la efectividad de las intervenciones gubernamentales. Finalmente, la integración de tecnologías avanzadas como el NLP en la evaluación del bienestar

ofrece un camino prometedor para capturar las dimensiones subjetivas y contextuales de manera más precisa, lo que podría llevar a la formulación de políticas más inclusivas y efectivas.

A pesar de los resultados alentadores, este estudio presenta varias limitaciones que deben considerarse. En primer lugar, la interpretación de las respuestas a preguntas abiertas puede variar significativamente en su nivel de detalle, lo que introduce subjetividad en el análisis y dificulta la captura de todas las sutilezas individuales. Además, el tamaño de la muestra, 849 participantes, es relativamente reducido en comparación con la población total de Medellín y la mayoría de los encuestados eran mujeres, lo que podría sesgar los hallazgos. Otro aspecto relevante es la reducción significativa en la longitud de las respuestas tras el preprocesamiento, lo que puede haber limitado la profundidad del análisis. Finalmente, la recolección de datos a través de un centro de llamadas y la transcripción manual introducen un potencial margen de error humano.

Agradecimientos

Se agradece al Fondo de Apoyo a la Investigación del CODI de la Universidad de Antioquia por la financiación del proyecto [código 2024-67391], del cual se deriva este artículo.

Referencias

- Acosta, A. (2013). *El Buen Vivir: Sumak Kawsay, una oportunidad para imaginar otros mundos*. Abya Yala.
- Alkire, S. (2007). The missing dimensions of poverty data: An introduction. *Oxford Development Studies*, 35(4), 347-359.
- Alkire, S. y Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7-8), 476-487.
- Alkire, S., Roche, J. M., Ballon, P., Foster, J., Santos, M. E. y Seth, S. (2015). *Multidimensional poverty measurement and analysis*. Oxford University Press.
- Alkire, S. y Santos, M. E. (2014). Multidimensional poverty index: A new approach. *Social Indicators Research*, 112(2), 109-129.
- Andreoni, A., Chang, H. y Estevez, I. (2021). The missing dimensions of the human capabilities approach: Collective and productive. *The European Journal of Development Research*, 33, 179-205.
- Angulo, R., Díaz, Y., & Rodríguez Pardo, R. (2016). The Colombian multidimensional poverty index: Measuring poverty in a public policy context. *Social Indicators Research*, 127, 1-38.
- Arrow, K. (1951). *Social choice and individual values*. Yale University Press.
- Atkinson, A. B. (2019). *Measuring poverty around the world*. Princeton University Press.
- ATD Fourth World y University of Oxford. (2019). *Hidden dimensions of poverty*. <https://www.atd-quartmonde.org/wp-content/uploads/2019/12/Hidden-Dimensions-of-Poverty-20-11-2019.pdf>
- Beck, V., Hahn, H. y Lepenies, R. (2020). *Dimensions of poverty: Measurement, epistemic injustices, activism* (Vol. 2). Springer Nature.
- Bessell, S. (2020). The individual deprivation measure: A gender-sensitive approach to multidimensional poverty measurement. En M. Sawer, F. Jenkins y K. Downing (Eds.), *How gender can transform the social sciences: Innovation and impact* (pp. 137-145). Springer.

- Biggeri, M., Libanora, R., Mariani, S. y Menchini, L. (2006). Children conceptualizing their capabilities: Results of a survey conducted during the first children's world congress on child labour. *Journal of Human Development*, 7(1), 59-83.
- Bird, S., Klein, E. y Loper, E. (2009). *Natural language processing with Python: Analyzing text with the Natural Language Toolkit*. O'Reilly Media.
- Blei, D. M. y Lafferty, J. D. (2007). A correlated topic model of science. *Annals of Applied Statistics*, 1(1), 17-35.
- Boarini, R. (2015). Wellbeing and quality of life reporting at the OECD. En *Sustainable development and wellbeing: Integrating quantitative and qualitative approaches* (pp. 747-761). Springer.
- Busetto, L., Wick, W. y Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and Practice*, 2(14), 1-10.
- Burchi, F., De Muro, P., & Kollar, E. (2014). Which dimensions should matter for capabilities? A constitutional approach. *Ethics and Social Welfare*, 8(3), 233-247. <https://doi.org/10.1080/17496535.2014.932415>
- Burchi, F., De Muro, P., & Kollar, E. (2018). Constructing well-being and poverty dimensions on political grounds. *Social Indicators Research*, 137(2), 441-462. <https://doi.org/10.1007/s11205-017-1618-0>
- Burchi, F., Espinoza-Delgado, J., Montenegro, C. E., & Rippin, N. (2021). An individual-based index of multidimensional poverty for low- and middle-income countries. *Journal of Human Development and Capabilities*, 22(4), 682-705. <https://doi.org/10.1080/19452829.2021.1964450>
- Caicedo-Beltrán, J. F., & Rubiano, P. J. (2023). Uso de Tecnologías de Big Data para Monitorizar y Predecir Tendencias Sociales. *Revista Vínculos*, 20(1). <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/16062>
- Camfield, L., Crivello, G. y Woodhead, M. (2008). Wellbeing research in developing countries: Reviewing the role of qualitative methods. *Social Indicators Research*, 90(1), 5-31.
- Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, 4(1), 1. <https://doi.org/10.3390/bdcc4010001>
- De Neve, J.-E., & Layard, R. (2023). *Wellbeing: Science and policy*. Cambridge University Press.
- Departamento Nacional de Planeación (DNP). (2012). *Metodologías oficiales y arreglos institucionales para la medición de la pobreza en Colombia* (CONPES 150). Departamento Nacional de Planeación. <https://colaboracion.dnp.gov.co/cdt/conpes/social/150.pdf>
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, 17(1), 61-84. <https://doi.org/10.3166/dn.17.1.61-84>
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95(3), 542-575.
- Diener, E., Lucas, R. E., Schimmack, U. y Helliwell, J. F. (2009). *Well-being for public policy*. Oxford University Press.
- Diener, E., Oishi, S. y Lucas, R. E. (2018). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 134(2), 276-302. <https://doi.org/10.1038/s41562-018-0307-6>
- Easterly, W. (2014). *The tyranny of experts: Economists, dictators, and the forgotten rights of the poor*. Basic Books.
- Ferreira, F. H. G. (2011). Poverty is multidimensional. But what are we going to do about it? *The Journal of Economic Inequality*, 9, 493-495. <https://doi.org/10.1007/s10888-011-9193-3>
- Fletcher, G. (2016). *The philosophy of well-being: An introduction*. Routledge.
- Frediani, A. A. (2019). Participatory research methods and the capability approach: Researching the housing dimensions of squatter upgrading initiatives in Salvador da Bahia, Brazil. En D. A. Clark, M. Biggeri y A. A. Frediani (Eds.), *The capability approach, empowerment, and participation: Concepts, methods, and applications* (pp. 261-288). Palgrave Macmillan.

- Griffiths, T. L. y Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1), 5228-5235. <https://doi.org/10.1073/pnas.0307752101>
- Grim, B., Harmel, R. y Smith, T. (2006). The data revolution: Using large datasets for comparative politics. *Journal of Politics and Society*, 12(3), 76-91.
- Guillén, M. & Phelán, M. (2016). *La medición del bienestar en América Latina*. Editorial Universidad del Rosario.
- Gupta, A., Kumar, P. y Singh, R. (2017). Dynamic topic modeling using recurrent neural networks for trend analysis. *International Journal of Data Science. Volume 1 (Long Papers)*, pages 1079–1089
- Hatef, E., Rouhizadeh, M., Nau, C., Xie, F., Rouillard, C., Abu-Nasser, M. y Roblin, D. (2022). Development and assessment of a natural language processing model to identify residential instability in electronic health records' unstructured data: A comparison of 3 integrated healthcare delivery systems. *JAMIA Open*, 5(1), ooac006, 1 – 10.
- Jonsson, A., Andersson, L., Alkan Olsson, J., & Johansson, M. (2007). Defining goals in participatory water management: merging local visions and expert judgments. *Journal of Environmental Planning and Management*, 50(2), 257–281.
- Jurafsky, D. y Martin, J. H. (2019). *Speech and language processing* (3rd ed.). Pearson.
- Kahneman, D. y Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38), 16489-16493.
- Kahneman, D. y Riis, J. (2005). Living, and thinking about it: Two perspectives on life. En F. A. Huppert, N. Baylis y B. Keverne (Eds.), *The science of well-being* (pp. 285-304). Oxford University Press.
- Klugman, J., Rodríguez, F. y Choi, H. J. (2011). The HDI 2010: New controversies, old critiques. *Journal of Economic Inequality*, 9(2), 249-288.
- Mahalakshmi, G., MuthuLakshmi, S. y Rajesh, S. (2017). Big Data analytics in crime analysis. *International Journal of Advanced Research in Computer Science. Vol (6)*, 1.311 - 1319.
- Manning, C. D., Raghavan, P. y Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Mendoza, G. A., & Prabhu, R. (2000). Development of a methodology for selecting criteria and indicators of sustainable forest management: A case study on participatory assessment. *Environmental Management*, 26(6), 659–673. <https://doi.org/10.1007/s002670010123>
- Mills, C. (2014). Veiled sentiments: Shame and the hidden dimensions of poverty. *Sociology*, 48(2), 290–306.
- Narayan, D., Patel, R., Schafft, K., Rademacher, A. y Koch-Schulte, S. (2000). *Voices of the poor: Can anyone hear us?* Oxford University Press for the World Bank.
- New Economics Foundation (NEF). (2012). *Measuring well-being: A guide for practitioners*. NEF.
- Nikolenko, S. I., Broslavsky, P., & Tishchenko, V. (2017). Topic modeling for qualitative studies. *Journal of Information Science*, 43(6), 864-879.
- Nolan, B., & Whelan, C. T. (2011). *Poverty and deprivation in Europe*. Oxford University Press.
- Nussbaum, M. (2000). *Women and human development: The capabilities approach*. Cambridge University Press.
- Nussbaum, M. (2011). *Creating capabilities: The human development approach*. Harvard University Press.
- OECD. (2013). *How's life? 2013: Measuring well-being*. OECD Publishing.
- Pacáková, Z. (2012). Variables selection for quality of life evaluation in the European Union. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 60(7), 199-206.
- Porter, M. E., Stern, S., & Green, M. (2017). Social Progress Index 2017. Social Progress Imperative. Recuperado de [\[https://www.socialprogress.org/\]](https://www.socialprogress.org/)
- PNUD. (2010). *La verdadera riqueza de las naciones: Caminos al desarrollo humano* (Reporte del desarrollo humano 2010). Programa de las Naciones Unidas para el Desarrollo. http://hdr.undp.org/sites/default/files/hdr_2010_es_complete_reprint.pdf
- Ravallion, M. (2011). On multidimensional indices of poverty. *The Journal of Economic Inequality*, 9, 235–248. <https://doi.org/10.1007/s10888-011-9173-4>

- Robeyns, I. (2005). The capability approach: A theoretical survey. *Journal of Human Development*, 6(1), 93-114.
- Robeyns, I. (2017). *Wellbeing, freedom, and social justice: The capability approach re-examined*. Open Book Publishers.
- Rodriguez, C. y Storer, T. (2020). Using topic modeling to analyze large datasets from social media. *Journal of Data Science*. Vol 18 No. 4.
- Ruiz, F. (2015). Métodos cualitativos y cuantitativos en la investigación social: Un análisis crítico. *Revista Mexicana de Sociología*, 77(4), 661-682.
- Sen, A. (1985). *Commodities and capabilities*. Oxford University Press.
- Sen, A. (1992). *Inequality reexamined*. Harvard University Press.
- Sen, A. (1999). *Development as freedom*. Knopf.
- Sen, A. (2009). *The idea of justice*. Harvard University Press.
- Sen, A. (2017). *Collective choice and social welfare*. Harvard University Press.
- Sen, A. y Nussbaum, M. (Eds.). (1996). *La calidad de vida*. Fondo de Cultura Económica.
- Sempere, J., Acosta, A., Abdallah, S. y Ortí, D. (2010). *Enfoques sobre bienestar y buen vivir*. Flacso.
- Chidi, U. C., Onyesolu, M. O., Asogwa, D. C., & Egwu, C. V. (2024). Exploring Latent Dirichlet Allocation (LDA) in topic modeling: Theory, applications, and future directions. *Newport International Journal of Engineering and Physical Sciences*, 4(1), 9-16. <https://doi.org/10.59298/NIJEP/2024/41916.1.1100>
- Stiglitz, J. E., Sen, A. y Fitoussi, J. P. (2009). *Report by the commission on the measurement of economic performance and social progress*. Paris.
- Taboada Villamarín, A. (2024). Big data en ciencias sociales. Una introducción a la automatización de análisis de datos de texto mediante procesamiento de lenguaje natural y aprendizaje automático. *Revista CENTRA de Ciencias Sociales*, 3(1), 51-75. <https://doi.org/10.54790/rccs.51>
- World Bank. (2017). *Monitoring global poverty: Report of the commission on global poverty*. World Bank Group.
- Wisor, S., Bessell, S., Castillo, F., Crawford, J., Donaghue, K., Hunt, J., Jaggar, A., Liu, A. y Pogge, T. (2014). *The individual deprivation measure: A gender-sensitive approach to poverty measurement*. International Women's Development Agency Inc.
- Whelan, C. T., Nolan, B. y Maître, B. (2014). Multidimensional poverty measurement in Europe: An application of the adjusted headcount approach. *Journal of European Social Policy*, 24(2), 183-197.
- Voinov, A., Jenni, K., Gray, S., Kolagani, N., Glynn, P. D., Bommel, P., Prell, C., Zellner, M., Paolisso, M., Jordan, R., Sterling, E. J., Olabisi, L. S., Giabbanelli, P. J., Sun, Z., Le Page, C., Elsawah, S., BenDor, T. K., Hubacek, K., Laursen, B. K., Jetter, A. J., Basco-Carrera, L., Singer, A., Young, L. G., Brunacini, J., & Smajgl, A. (2018). Tools and methods in participatory modeling: Selecting the right tool for the job. *Environmental Modelling & Software*, 109, 232-255. <https://doi.org/10.1016/j.envsoft.2018.08.02>