

MÁSTER EN TRADUCCIÓN ESPECIALIZADA

UNIVERSIDAD DE VIC – UNIVERSIDAD CENTRAL DE CATALUÑA

**EL LENGUAJE CONTROLADO Y LA
TRADUCCIÓN AUTOMÁTICA COMO
VEHÍCULOS DE TRANSMISIÓN DEL
CONOCIMIENTO CIENTÍFICO**

Autora: Marta Acosta Plata

Tutores: Ruben Giró Anglada, Carla Parra Escartín

Fecha de depósito: 5 de septiembre de 202

**EL LENGUAJE CONTROLADO Y LA
TRADUCCIÓN AUTOMÁTICA COMO
VEHÍCULOS DE TRANSMISIÓN DEL
CONOCIMIENTO CIENTÍFICO**

Autora: Marta Acosta Plata

Fecha de depósito: 5 de septiembre de 2021

AGRADECIMIENTOS Y DEDICATORIA

En primer lugar, quisiera agradecerles a Carla y a Ruben, mis tutores, sus valiosos consejos y aportaciones, los ánimos cuando más los necesitaba y, sobre todo, el *atarme en corto* cuando mi mente echa a volar.

Este trabajo va dedicado a mis padres, por siempre apoyarme de manera incondicional en todas mis disparatadas ideas, como la de comenzar una nueva carrera profesional en el mundo de la traducción a estas alturas de la vida.

A mi hermana Irene que, aun estando a miles de kilómetros, tanta compañía me hizo a horas intempestivas durante todos estos meses cuando necesitaba despejarme y pensar en cualquier otra cosa que no fuera el lenguaje controlado.

A mi querida Ruth, por enseñarme a transitar nuevos caminos y devolverme la ilusión por estudiar.

Y, finalmente, a Jeremy, por ayudarme simplemente a *volver*.

GRACIAS.

RESUMEN

La transmisión eficaz del conocimiento científico resulta crucial en tiempos de crisis como el que estamos viviendo debido a la pandemia de SARS-CoV-2. Aunque el uso del inglés como lengua franca de la ciencia permite una comunicación global, su hegemonía podría estar causando una brecha lingüística. El objetivo del presente estudio fue realizar una primera aproximación a la aplicación conjunta del lenguaje controlado, la traducción automática y la posesición a la traducción de artículos científicos en el par de idiomas ES>EN para mejorar la traducibilidad automática sin comprometer la calidad del texto de partida. Para ello se seleccionó un único artículo médico al que se le aplicó control a nivel léxico-terminológico y a nivel gramatical a través de diez reglas recopiladas de la literatura. Tanto el texto bruto como el controlado fueron traducidos a través de Google Translate. Los productos de la traducción automática fueron poseídos y se realizó una anotación exhaustiva de la preedición y de la posesición con el fin de evaluar el esfuerzo técnico de ambos procesos, evaluar la efectividad del lenguaje controlado y estudiar el error. La preedición en conjunto tuvo un efecto positivo al reducir el esfuerzo de posesición en el texto controlado, aunque este fue inferior al esperado. El control gramatical y el control léxico-terminológico tuvieron respuestas dispares. Las reglas de control gramatical fueron menos eficaces de lo previsto, especialmente al tener en cuenta el elevado esfuerzo de preedición que requirieron y que fueron responsables de introducir un número considerable de errores en el texto controlado. El control léxico-terminológico tuvo una eficiencia elevada, ya que consiguió reducir los errores de imprecisión léxica a la mitad. Ambos tipos de control del lenguaje podrían incrementar su eficacia con el desarrollo de herramientas informáticas específicas para el español, no disponibles hasta la fecha.

Palabras clave: lenguaje controlado, traducción automática, traducibilidad, esfuerzo de posesición, artículo científico

ABSTRACT

The effective transmission of scientific knowledge is paramount in times of crisis such as the one we are experiencing due to the SARS-CoV-2 pandemic. Although the use of English as lingua franca of science enables global communication, its hegemonic presence could also be creating a language gap. The goal of the present study was to carry out the first attempt to a joint application of controlled language, machine translation, and post-editing to the translation of scientific articles from Spanish to English and improve automatic translatability without compromising the quality of the source text. A single medical article was selected and subject to restrictions at a lexical-terminological level and grammatical level, which was applied through ten rules as compiled from the literature. Both the raw text and the controlled version were translated using Google Translate. The machine translation products were post-edited, and the pre-edited and post-edited resulting texts were exhaustively annotated to estimate the technical effort required during both processes, evaluate the effectiveness of the controlled language, and study the error. Pre-editing had a positive effect and reduced the post-editing effort in the controlled text, although its impact was lower than expected. Grammatical control and lexical-terminological control had dissimilar outcomes. Grammar control rules were less effective than foreseen, especially considering the high pre-editing effort required, while they were also responsible for introducing a considerable number of errors in the controlled text. The lexical-terminological control was highly efficient and reduced lexical accuracy errors by half. The effectiveness of both types of language control would benefit from the development of computer tools specifically designed for the Spanish language, which are not available to date.

Keywords: controlled language, machine translation, translatability, post-editing effort, scientific article

ÍNDICE

| | |
|----------------------------------------------------------------------|----|
| RESUMEN..... | 4 |
| ABSTRACT | 5 |
| 1. INTRODUCCIÓN | 8 |
| 1.1. JUSTIFICACIÓN DEL ESTUDIO Y ESTADO DE LA CUESTIÓN | 8 |
| 1.2. OBJETO, HIPÓTESIS Y OBJETIVOS CONCRETOS DEL ESTUDIO | 13 |
| <i>Objeto del estudio</i> | 13 |
| <i>Hipótesis de estudio</i> | 13 |
| <i>Objetivos concretos del estudio</i> | 14 |
| 2. ANÁLISIS DEL DISCURSO | 15 |
| 2.1. EL LENGUAJE Y EL ESTILO DE LA CIENCIA..... | 15 |
| <i>La importancia de la comunicación en la ciencia</i> | 15 |
| <i>El lenguaje de la ciencia</i> | 16 |
| 2.2. PROBLEMAS Y ERRORES DEL LENGUAJE MÉDICO EN ESPAÑOL | 18 |
| 2.3. EL ARTÍCULO CIENTÍFICO COMO GÉNERO TEXTUAL | 20 |
| <i>Definición</i> | 20 |
| <i>Función</i> | 20 |
| <i>Estructura</i> | 20 |
| 3. EL LENGUAJE CONTROLADO | 22 |
| <i>Definición</i> | 22 |
| <i>Clasificación</i> | 23 |
| <i>Diseño y aplicación</i> | 24 |
| <i>Lenguaje controlado y traducción</i> | 25 |
| <i>Situación de los lenguajes controlados en español</i> | 27 |
| 4. MATERIALES Y MÉTODOS | 28 |
| <i>Selección del texto</i> | 28 |
| <i>Preedición y aplicación del lenguaje controlado</i> | 29 |
| <i>Esfuerzo técnico de preedición</i> | 33 |
| <i>Traducción automática</i> | 34 |
| <i>Posedición</i> | 34 |
| <i>Esfuerzo técnico de posedición</i> | 35 |
| <i>Eficacia del lenguaje controlado y prevención del error</i> | 36 |

| | |
|---------------------------------------------------------------------------------------|----|
| 5. RESULTADOS | 36 |
| <i>Características generales del texto objeto de estudio</i> | 36 |
| <i>Esfuerzo técnico de preedición: lenguaje controlado y corrección</i> | 37 |
| <i>Esfuerzo técnico de posesición: clasificación y cuantificación del error</i> | 41 |
| <i>Eficacia del lenguaje controlado y prevención del error</i> | 47 |
| <i>Léxico problemático: ambigüedad, variedades dialectales y anglicismos</i> | 52 |
| 6. DISCUSIÓN..... | 56 |
| 7. CONCLUSIONES | 65 |
| REFERENCIAS | 67 |
| ANEXOS..... | 73 |

1. INTRODUCCIÓN

1.1. JUSTIFICACIÓN DEL ESTUDIO Y ESTADO DE LA CUESTIÓN

Vivimos en un mundo globalizado en el que el devenir de una zona concreta puede tener implicaciones a escala global. Esto es especialmente cierto en lo que se refiere a cuestiones medioambientales y de salud pública. La Organización Mundial para la Salud, en el marco del lanzamiento de su Plan de Acción Estratégico para 2019-2023, identificaba los 10 principales riesgos para la salud global (WHO, 2019). Entre ellos destacaba la contaminación del aire y el cambio climático como las mayores amenazas, seguidas de las enfermedades no transmisibles (diabetes, cáncer, enfermedades cardiovasculares) y una amalgama de riesgos relacionados con las enfermedades transmisibles (pandemias, patógenos, resistencia a antibióticos y renuencia a la vacunación) que, en combinación con otros factores, como la debilidad de los sistemas sanitarios en gran parte del globo, constituirían un caldo de cultivo para el resurgimiento de enfermedades altamente infecciosas con consecuencias devastadoras. Prueba de ello es la situación que estamos viviendo en la actualidad con la pandemia de SARS-CoV-2 (COVID-19). La vulnerabilidad de la salud humana en un escenario en el que el cambio climático está alterando las reglas del juego, en el que la salud está más comprometida que nunca a nivel global y en el que un proceso infeccioso se extiende vertiginosamente, la transmisión de información de forma eficaz, especialmente del conocimiento científico relacionado con el medio ambiente y la salud, es crucial para poder actuar en tiempos de crisis.

Por otro lado, también vivimos en lo que conocemos como la sociedad de la información y del conocimiento, en la que existe una tendencia creciente a salvar las brechas digitales y a hacer accesible todo conocimiento humano de una forma democrática (UNGIS, 2019). Dado que el acceso a la información de carácter científico es fundamental para salvaguardar el medio ambiente y la salud de las personas y que, gracias a las nuevas tecnologías, el conocimiento es en teoría mucho más asequible para todos, sería de esperar que la brecha lingüística y, por tanto, las desigualdades con respecto al acceso al conocimiento y al mundo académico, se hubieran reducido enormemente (Gibbs, 1995).

Las barreras geográficas y lingüísticas han constituido un reto a la transmisión del conocimiento a lo largo de la historia. La principal estrategia para solventar dichas barreras comunicativas es el uso de una lengua común, también conocida como lengua vehicular, lengua franca o *lingua franca* (Drubin y Kellogg, 2012). El continente europeo ha sido especialmente prolífico en el uso de lenguas vehiculares, incluidas el griego clásico, la lengua tudesca, el latín clásico (de gran relevancia en la biología, el derecho, la filosofía, la literatura y la religión), el sabir, el alemán, el francés y el inglés hoy en día. El inglés se ha consolidado como lengua vehicular en la mayoría de las esferas de la sociedad a nivel global (por ejemplo, en el ámbito diplomático), con un impacto aún mayor sobre las disciplinas científicas, donde se emplea prácticamente en exclusividad para compartir el conocimiento generado a través de la publicación de artículos científicos en revistas especializadas, revistas de divulgación, redes sociales, etc. Cabe destacar que las lenguas vehiculares no son habitualmente elegidas de forma activa por sus características, tal como se le atribuye al inglés por su supuesta sencillez, claridad y precisión, cualidades más que discutibles, sino que acaban por asentarse debido a razones extralingüísticas o hegemónicas. Tal es el caso del inglés como lengua franca, que comienza a ganar importancia al término de la Segunda Guerra Mundial (Navarro, 2001), lo que ha traído como consecuencia un claro predominio del monolingüismo anglófono en la ciencia y la medicina (Aleixandre Benavent y Amador Iscla, 2001a). Al igual que a lo largo de la historia han sido diversas las lenguas francas empleadas, no sería descabellado pensar que un cambio en la situación geopolítica a nivel mundial en el futuro pudiera dar como resultado el empleo de una lengua vehicular diferente del inglés (Englander, 2014). Dado que el inglés se ha convertido en la lengua franca indiscutible de la ciencia (Navarro, 2001), solo cabe imaginar que un cambio en el idioma empleado para la comunicación global, aunque improbable, tendría un efecto catastrófico para la transmisión del conocimiento.

Aunque el uso del inglés como *lingua franca* ha mejorado (en teoría) la comunicación entre la comunidad científica, su uso no está exento de detractores y algunas investigaciones recientes han revelado que su uso generalizado no solo no ha solventado la barrera lingüística, sino que podría estar agravándola. Existen estudios que ponen de manifiesto que el uso del inglés como vehículo fundamental de transmisión del conocimiento científico en materia medioambiental conlleva que un gran volumen de información científica escrita en

otros idiomas sea obviado por parte de la comunidad científica y que este sesgo podría tener graves implicaciones en un mundo globalizado (Amano *et al.*, 2016). Tal como sintetiza Navarro (2001), la consolidación del inglés como lengua franca en la ciencia presenta grandes desventajas, como (1) la influencia del inglés sobre otros idiomas, (2) la exclusión de las aportaciones realizadas en otras lenguas, (3) la dependencia científica e intelectual, (4) el inglés como sinónimo de calidad, (5) dificultades para la participación activa en la comunidad científica internacional y (6) la discriminación lingüística.

En situaciones como la actual, el conocimiento científico adquiere una especial importancia para la toma de decisiones, pero cuando la ciencia se comunica exclusivamente en inglés, se compromete su tercera función, la de informar al público. Este efecto nunca se había intensificado tanto como durante la pandemia de la COVID-19 (Taşkın *et al.*, 2020). La escasa disponibilidad de producción científica en idiomas locales afecta a la implementación del conocimiento por parte de diversos grupos de actores, incluidos, aparte de la comunidad científica, las instituciones públicas de carácter regulatorio y las entidades o personas que realizan trabajo de campo. En su estudio, Amano *et al.* (2016) ponen de manifiesto que el 54 % de los responsables de la gestión de espacios protegidos en España no puede acceder al conocimiento científico publicado en inglés debido a una barrera lingüística. Este fenómeno también se da presumiblemente en el campo de la producción científica médica y, sin embargo, según la OMS, desde la declaración de la pandemia y hasta junio de 2020, se habían publicado 10 728 artículos científicos y otros documentos relacionados con la COVID-19, de los cuales el 85 % estaban redactados en inglés (Taşkın *et al.*, 2020).

Las desventajas del empleo del inglés como lengua vehicular parecen extenderse de manera bidireccional, ya que publicar en inglés constituye un reto para los autores académicos no nativos, cuyos textos deben someterse a procesos de edición y revisión más costosos (Benfield y Feak, 2006). Las grandes compañías editoriales controlan el ámbito de la publicación científica a nivel internacional y sus publicaciones son altamente valoradas según el sistema actual de evaluación de la calidad científica, que está fuertemente sesgado hacia las publicaciones realizadas en inglés (Giménez Toledo *et al.*, 2019). Debido a ello, el trabajo multilingüe está perdiendo relevancia y muchos investigadores consideran que su visibilidad profesional, el impacto de sus resultados y el progreso dependen de publicar en

revistas científicas en inglés (Kulczycki *et al.*, 2019). Un estudio llevado a cabo en la Universidad de Santiago de Compostela reveló que, aunque existe una actitud positiva entre la comunidad científica hacia el inglés como lengua franca, muchos autores no pueden contribuir al conocimiento científico a nivel internacional y remediar esto a menudo recae directamente sobre los autores, que deben afrontar mayores costes económicos, formarse en idiomas por su cuenta y realizar un mayor esfuerzo individual en comparación con los autores nativos (Fernández Polo y Cal Varela, 2009). Amano *et al.* (2016) proponen que la comunidad científica comience a reconocer esta barrera lingüística, que se valore la producción científica en otros idiomas y que se mejore la accesibilidad a la producción escrita en inglés mediante su traducción para que todo posible usuario de este conocimiento pueda beneficiarse. Aunque existen recomendaciones acerca de comenzar a evaluar de forma equitativa las publicaciones realizadas en idiomas diferentes al inglés, no parece haber directrices claras sobre cómo poner estas medidas en práctica (Kancewicz-Hoffman y Pölönen, 2020).

A raíz de esta situación han surgido diversas iniciativas internacionales en defensa del multilingüismo en la ciencia, como ERIH, INTERCO-SSH, OPERAS y la Iniciativa Helsinki, que afrontan la problemática desde diferentes enfoques y en ámbitos variados, sobre todo en el de las ciencias sociales y las humanidades (Kancewicz-Hoffman y Pölönen, 2020). La propuesta más global es la de la Iniciativa Helsinki, que promueve un modelo de diseminación científica que protege las infraestructuras nacionales, favorece la diversidad lingüística y enfatiza que los sistemas de evaluación actuales llevan a una infrarrepresentación de las publicaciones realizadas en idiomas distintos del inglés y no fomentan que los autores publiquen en su propio idioma, con las consecuencias que esto acarrea al poner en riesgo la difusión científica a nivel local (Taşkın *et al.*, 2020). La bibliodiversidad o diversidad del contenido académico es necesaria para preservar la investigación y el conocimiento sobre temas relevantes a distintas escalas, así como la diversidad epistemológica y metodológica y el lenguaje científico en distintos idiomas, a la par que permite garantizar la supervivencia de la pluralidad en la investigación (Giménez Toledo *et al.*, 2019). Algunos autores como Kancewicz-Hoffman y Pölönen (2020) proponen la traducción y el uso y estudio de los motores de traducción automática como una solución clave a la hora de mejorar los sistemas de evaluación de la calidad de la

investigación para que sean más equitativos (lo que tendría un fuerte impacto en las métricas de evaluación), para reducir las barreras a la diseminación de resultados y a la revisión de trabajos académicos (el multilingüismo también puede constituir una barrera para alcanzar audiencias más amplias) y para incrementar la accesibilidad a trabajos publicados en otros idiomas.

El rápido avance de la inteligencia artificial y de los motores de traducción automática basados en redes neuronales, así como la disponibilidad actual de motores libres y gratuitos, los convierten en una herramienta útil para todos los públicos para su uso, no solo profesional por parte de los traductores, sino también personal (Parra Escartín *et al.*, 2017). No obstante, se considera que los textos resultantes de la aplicación de la traducción automática requieren de cierto grado de posesición para alcanzar una calidad aceptable (O'Brien, 2004). A pesar del gran potencial de esta herramienta para subsanar los problemas derivados del uso del inglés como lengua franca en la ciencia, son escasos los estudios que se han centrado en explorar si la traducción automática puede contribuir a facilitar los procesos de redacción y publicación académica para los autores no nativos del inglés. En su estudio, Parra Escartín *et al.* (2017) recalcan que los profesionales del ámbito médico ya se sirven de la traducción automática para ayudarse durante el proceso de redacción científica, ya que escriben primero en español para después traducir al inglés, pero el texto final poseditado por los propios autores no consigue alcanzar un estándar de calidad publicable.

Existe un gran volumen de trabajos centrados en la calidad de los textos producidos a través de motores de traducción automática y sus métodos de evaluación. Sin embargo, el texto de partida suele recibir menos atención. ¿Es posible redactar un artículo científico de tal manera que se alcance un texto de salida traducido automáticamente con una calidad aceptable y unas necesidades mínimas de posesición? Algunos autores han explorado el impacto de la redacción del texto de partida en la calidad del texto resultante de la traducción automática. Estos estudios se han centrado principalmente en la aplicación de lenguajes controlados o simplificación del lenguaje (traducción intralingüística) en los textos de partida mediante restricciones léxicas, gramaticales y estilísticas con el fin de mejorar la traducibilidad, la comprensibilidad, la legibilidad o la usabilidad de un texto. La mayoría de los trabajos se han ocupado del estudio de textos técnicos traducidos desde el inglés y, aunque existen

resultados contradictorios, parece ser que la aplicación de lenguajes controlados mejora la legibilidad y la traducibilidad automática (reduce el esfuerzo de posesición), especialmente en textos complejos (O'Brien, 2010).

En la actualidad contamos con herramientas tanto lingüísticas como tecnológicas a nuestra disposición que podrían contribuir a derribar las barreras que impiden un acceso democrático a la ciencia y al mundo académico, tales como el control del lenguaje de redacción y la traducción automática. El presente trabajo pretende realizar una primera aproximación a la combinación de ambas técnicas en un nuevo flujo de trabajo para la redacción y publicación de artículos científicos.

1.2. OBJETO, HIPÓTESIS Y OBJETIVOS CONCRETOS DEL ESTUDIO

Objeto del estudio

El objeto del presente estudio es realizar una primera aproximación a la aplicación conjunta de la preedición con lenguaje controlado, la traducción automática y la posesición a la traducción de artículos científicos en el campo de la medicina (en concreto, de la epidemiología) en el par de idiomas ES>EN. En definitiva, se pretende evaluar si resulta factible y útil aplicar restricciones a la redacción en los textos de partida en español para mejorar su traducibilidad automática al inglés sin comprometer su calidad en origen y obtener resultados publicables con menor esfuerzo de posesición, como una herramienta para democratizar el acceso a la ciencia y favorecer comunidades científicas multilingües.

Hipótesis de estudio

La hipótesis inicial de este trabajo es que la aplicación de reglas de lenguaje controlado a los artículos científicos en el ámbito de la medicina redactados en español facilita su traducción al inglés por parte de motores de traducción automática basados en redes neuronales y reduce el esfuerzo de posesición.

Objetivos concretos del estudio

Con el fin de evaluar si el control del lenguaje resulta útil para mejorar la traducibilidad automática de artículos científicos del ámbito de la medicina del español al inglés, se proponen los siguientes objetivos concretos:

- Realizar un análisis del discurso del artículo científico-médico en español para identificar los errores de redacción y los problemas de traducción más comunes.
- Realizar una revisión de las aplicaciones del lenguaje controlado.
- Contrastar las características del discurso del artículo científico-médico en español con las reglas de lenguaje controlado más frecuentes y seleccionar una lista tentativa de restricciones a aplicar (traducción intralingüística).
- Cuantificar y evaluar la aplicación del lenguaje controlado al texto estudiado (ES) con el fin de determinar el esfuerzo de preedición.
- Cuantificar y evaluar el error en la traducción (ES>EN) realizada por un motor de traducción automática basado en redes neuronales, tanto en el texto controlado, como en su versión no controlada o bruta y realizar un análisis comparativo.
- Cuantificar y evaluar el esfuerzo de posesición requerido en las traducciones automáticas de la versión controlada y la versión bruta del texto (ES>EN) y realizar un análisis comparativo.
- Determinar la eficacia del lenguaje controlado e identificar qué reglas resultan útiles para mejorar la traducibilidad (ES>EN) y reducir el esfuerzo de posesición.
- Determinar qué características del discurso y secciones del artículo científico en el ámbito de la medicina pueden ser problemáticos para los motores de traducción automática basados en redes neuronales en el par de idiomas ES>EN.

2. ANÁLISIS DEL DISCURSO

Para llevar a cabo este estudio resulta necesario, en primer lugar, caracterizar el lenguaje y el estilo de la ciencia, los problemas y errores comunes en el lenguaje de la medicina en español y el artículo científico como género textual.

2.1. EL LENGUAJE Y EL ESTILO DE LA CIENCIA

La importancia de la comunicación en la ciencia

El objetivo de la investigación científica es la publicación, ya que permite verificar los nuevos conocimientos e incorporarlos al saber colectivo (principio de reproducibilidad). La comunicación o publicación científica, tal como la conocemos, es relativamente nueva y se remonta a las primeras revistas científicas especializadas nacidas en 1665. En la actualidad, este formato es el medio de comunicación principal en todas las disciplinas (Day, 2005).

Los textos científicos no deben tener valor solo por su contenido (validez científica), sino que además deben cuidar los aspectos formales (De Biase *et al.*, 2009). Los más acríticos opinan que lo importante es que se entienda lo que se quiere decir (Aleixandre Benavent y Amador Iscla, 2001a). Esta visión radica en que la destreza para escribir y la importancia de la calidad de redacción estén sujetas a un cierto menosprecio en el ámbito de la ciencia y la medicina. Además, el aprendizaje de las habilidades de redacción en español se complica porque la mayoría de los artículos se publican en inglés. La baja calidad en la redacción científica es un problema transversal, con denominadores comunes tales como la monotonía, la ampulosidad, el abuso de la jerga, la ambigüedad, la imprecisión y la falta de claridad. Hasta una cuarta parte de las objeciones de los revisores de los manuscritos enviados a revistas científicas se refiere a la comprensibilidad del texto (Pulido, 2010). Para redactar textos de calidad e incrementar las probabilidades de publicación es necesario tener en cuenta al lector (De Biase *et al.*, 2009), organizar bien las ideas (Day, 2005; Pulido, 2010) y corregir el texto o trabajar con mediadores lingüísticos (Gil-Alberdi González, 2009; Pulido, 2010), ya que, además, existe una creciente complejidad en los formatos requeridos por las normativas y los editores de las revistas (Rosales, 2009).

El lenguaje de la ciencia

El lenguaje científico no persigue fines estéticos, creativos ni recreativos, sino informativos, didácticos y comunicativos (Navarro, 2009) y sus formas de elocución más frecuentes son la exposición y la descripción (Domínguez García, 2009), aunque la argumentación es imprescindible en géneros como el artículo científico (Carmona Sandoval, 2013). Se caracteriza, además, por emplear un registro formal tanto a nivel léxico como sintáctico (Giba, 2014). Sin embargo, no todos los idiomas tratan el lenguaje científico por igual: el inglés tiende a disminuir los cultismos y el registro es solo un poco más formal que el habla corriente, al contrario que sucede en español. Esta asimetría es un problema frecuente durante la traducción de textos científicos (Claros Díaz, 2016). El lenguaje científico puede considerarse como una parcela del lenguaje natural (Escobar, 2007) y no existe un único lenguaje científico, sino tantos lenguajes de especialidad como ciencias (Artigas Guillamón, 2000). Esta indivisibilidad representa una clara desventaja dada la ambigüedad e imprecisión de las lenguas naturales y ocasiona un cierto menosprecio hacia el lenguaje en la ciencia, que se considera como un elemento de distorsión (Escobar, 2007).

Al lenguaje científico se le atribuyen frecuentemente características de la ciencia por extensión, lo que lleva a algunos autores a considerar que la objetividad y la neutralidad son sus rasgos definitorios (Carmona Sandoval, 2013; De Biase *et al.*, 2009). Estas cualidades son raramente alcanzables, tanto en la ciencia como en el lenguaje, y sin embargo existen estrategias que buscan proyectarlas, como el uso de referencias bibliográficas y citas, la presentación de datos en tablas o gráficos (De Biase *et al.*, 2009), o la impersonalidad en el discurso mediante estrategias sintácticas como la ausencia de la segunda persona, el plural de modestia y el uso abusivo de verbos impersonales y de la voz pasiva (Carmona Sandoval, 2013). Por el contrario, existe una serie de principios fundamentales altamente interrelacionados a observar para producir manuscritos de alta calidad (Tabla 1) y que deberán aplicarse junto con una gran organización (Day, 2005) y un empleo impecable de la mecánica del lenguaje natural (gramática, ortografía y puntuación) (Giba, 2014).

| Atributo | Definición y estrategias |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Claridad | Presenta un razonamiento o argumento de fácil comprensión (Pulido, 2010). Característica fundamental, ya que la comunicación científica es un proceso bidireccional: un artículo científico publicado resulta inútil si no es recibido y entendido por el público al que se destina (Day, 2005). |
| Fluidez | No muestra dificultad a la lectura (Pulido, 2010). Cohesión y coherencia textual, orden de presentación de la información, construcciones paralelas, terminología consistente, uso racional de las abreviaturas (Giba, 2014). |
| Precisión | Concisión y exactitud rigurosa. No es ambiguo, incierto, dudoso, ni admite interpretaciones distintas según el lector (Pulido, 2010). Atributo más relevante (Carmona Sandoval, 2013; Navarro, 2009). Vocablos específicos, correctos y pertenecientes a un registro adecuado. Atención a la gramática, la puntuación, la ortografía y las convenciones (Giba, 2014). |
| Sencillez | Descripciones llanas, sin artificio, uso de palabras simples y ausencia de tecnicismos (Pulido, 2010). Evitar palabras demasiado complejas, no abusar de la sustantivación, evitar la doble negación, no construir frases demasiado largas, reducir el uso de la pasiva (también en inglés) y reducir el número de modificadores yuxtapuestos (muy frecuente en inglés) (Giba, 2014). |
| Concisión | Sobriedad, brevedad y economía de palabras (Pulido, 2010). Se recomienda eliminar las redundancias, las palabras y expresiones vacías de significado y elegir expresiones más cortas (Giba, 2014). |

Tabla 1. Atributos, definición y estrategias de implementación en el lenguaje científico. Elaboración propia.

La comunicación científica se ayuda del lenguaje simbólico o formal, basado en elementos gráficos que se combinan para expresar nuevas ideas y sujeto a normas universales de cifrado. El lenguaje científico también emplea las nomenclaturas, una combinación del lenguaje natural con el lenguaje formal basada en convenciones internacionales, como la taxonomía o la formulación química (Claros Díaz, 2016). El uso del lenguaje simbólico, de la lógica formal y de las convenciones contribuye a la brevedad y a la exactitud (Carmona Sandoval, 2013), aunque Escobar (2007) recuerda que no todo puede expresarse a través de la lógica formal, por lo que siempre hay que recurrir al lenguaje natural.

Quizá la dimensión más relevante para la caracterización de los lenguajes de especialidad sea el uso de la terminología. En todo lenguaje de especialidad se emplea terminología específica a ese campo concreto, además de terminología común a todas las ciencias y vocablos del lenguaje natural que adquieren significados diferentes en el ámbito en cuestión. La especificidad y densidad terminológica es un rasgo que permite determinar el grado de especialización de un texto científico y su nivel de abstracción (Rodríguez-Tapia, 2020).

2.2. PROBLEMAS Y ERRORES DEL LENGUAJE MÉDICO EN ESPAÑOL

Dado que existen lenguajes específicos a cada una de las disciplinas, o lenguajes de especialidad, solo se comentarán a continuación algunas características generales del lenguaje médico en español que resultan problemáticas o inducen a errores frecuentes.

La interferencia del inglés

El factor más señalado como causante de este problema es la actual hegemonía del inglés en todos los campos científicos (Aleixandre Benavent y Amador Iscla, 2001a). La interferencia lingüística del inglés da lugar a errores en la redacción que afectan a la precisión del texto médico y que se originan frecuentemente por la traducción literal a nivel léxico-semántico y por el calco de estructuras sintácticas. Estos errores se producen normalmente por comodidad, prisas o ignorancia del autor y no solo afectan a los traductores, sino también, y quizás en mayor medida, a los médicos que escriben directamente en español pero están acostumbrados a leer las publicaciones en inglés (De Biase *et al.*, 2009; Navarro, 2009). Los ejemplos más representativos de estos errores son los extranjerismos, los falsos amigos y los calcos, así como el abuso de la voz pasiva y del gerundio en estructuras anglicadas.

Uso impreciso del lenguaje

Las imprecisiones en el lenguaje médico no siempre son atribuibles a la influencia de otras lenguas extranjeras y a menudo emanan de la redacción en nuestro propio idioma. Cuando se producen en términos tan usados que no nos detenemos a pensar en si son correctos, pueden incluso pasar inadvertidos (Navarro, 2009). Algunos ejemplos serían los errores en la terminología específica, los problemas asociados a las abreviaciones, el uso de símbolos, de epónimos, o los pleonasmos.

La Tabla 2 recoge un listado de las principales características problemáticas y errores en el lenguaje médico en español, así como al área a la que afectan, sus posibles causas y algunos ejemplos.

| Error | Nivel | Causas lingüísticas | Ejemplos (EN-ES) |
|-----------------------------------|--------------|---------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| Polisemia | L-S | Interferencia: (1) lingüística (2) entre registros (3) diatópica (4) diacrónica | (1) <i>ignore/ignorar</i> (2) <i>pierna/muslo</i> (3) <i>gotero</i> (España/LATAM) (4) diversidad terminológica |
| Términos sin traducir | L-S | Interferencia Traducción | <i>Test, shock, screening, rash</i> |
| Monosemia (EN) vs. polisemia (ES) | L-S | Interferencia Traducción | <i>Mitad/half, medium, mean y average</i> |
| Préstamo homófono/falso amigo | L-S | Interferencia Traducción | <i>Severe illness/enfermedad severa</i> |
| Yuxtaposición de sustantivos | L-S | Interferencia Traducción | <i>Estudio caso-control/estudio de casos y controles</i> |
| Anglolanismos | L-S | Interferencia Traducción | <i>Córtex/corteza</i> |
| Confusión entre vocablos afines | L-S | Desconocimiento | <i>Molaridad/Molalidad</i> |
| Confusión entre conceptos afines | L-S | Desconocimiento | <i>Chlamydia/clamidia (microbio-enfermedad)</i> |
| Metáforas desafortunadas | L-S | Desconocimiento | <i>Epidemia de autismo</i> |
| Metonimia | L-S | Desconocimiento | <i>Anatomía/cuerpo</i> (disciplina-objeto de estudio) |
| Abreviaciones | L-S | Desconocimiento | Interpretación, polisemia, falta de unanimidad, invención, problemas derivativos |
| Símbolos | L-S | Desconocimiento | Vulneración de las normas del lenguaje o de las convenciones |
| Epónimos | L-S | Desconocimiento | Denominaciones vacías, falta de unanimidad, varios referentes, incompletos, creación difícil |
| Errores ortográficos | O-M | Interferencia Desconocimiento | Omisiones (<i>radioresistencia</i>), acentuación, uso inapropiado de la mayúscula |
| Género gramatical | O-M | Desconocimiento | <i>Enzima</i> (el/la) |
| Barbarismos | O-M | Interferencia Desconocimiento | Conjunción <i>y/o</i> |
| Abuso del gerundio | S-E | Interferencia | Abuso como conector, <i>calco</i> |
| Abuso de la pasiva | S-E | Interferencia | Abuso de la pasiva perifrástica, <i>calco</i> |
| Solecismos | S-E | Interferencia Desconocimiento | Concordancia, de régimen (preposiciones), de construcción |
| Pleonasmos | S-E | Desconocimiento | <i>Caries dental</i> |
| Repeticiones | S-E | Desconocimiento | Muletillas y latiguillos |
| Sinécdote | S-E | Desconocimiento | <i>Tratar una enfermedad</i> |
| Deshumanización | S-E | Desconocimiento | <i>Caso/paciente</i> |
| Sexismo | S-E | Desconocimiento | <i>Hombres vs. seres humanos</i> |

Tabla 2. Listado de problemas y errores del lenguaje médico en español según su nivel: léxico-semántico (L-S), ortográfico-morfológico (O-M) y sintáctico-estilístico (S-E). Elaboración propia a partir de: Aleixandre Benavent y Amador Iscla, 2001a, 2001b; Amador Iscla y Aleixandre Benavent, 2002; Navarro, 2009.

2.3. EL ARTÍCULO CIENTÍFICO COMO GÉNERO TEXTUAL

Definición

Según Day (2005), un artículo científico es un informe escrito y publicado (de una forma concreta) que describe los resultados originales de una investigación. Carmona Sandoval (2013) lo define como una modalidad de discurso escrito empleada por los investigadores para informar de sus hallazgos a sus pares a través de una revista especializada y persuadirlos de la calidad científica de su trabajo.

Función

La finalidad de cualquier discurso o texto científico es fundamentalmente la de informar al lector, por lo que la función predominante es la función informativa, representativa o referencial (Domínguez García, 2009), aunque pueden aparecer otras funciones, como la metalingüística cuando se definen términos. Sin embargo, como menciona Day (2005), a los científicos se los juzga y se los conoce por sus publicaciones y es importante recordar que el éxito profesional del investigador depende de su currículo académico y del impacto de sus publicaciones científicas (*publish or perish*). Por ello, algunos autores consideran que la definición del artículo científico como género discursivo centrado exclusivamente en torno a la objetividad y a la imparcialidad es una perspectiva ingenua, ya que el hecho de publicar un artículo científico es inherentemente una acción estratégico-persuasiva (Meza, 2017).

Estructura

Day (2005) realiza un repaso por la evolución de la estructura del artículo científico y explica que, si bien las primeras revistas científicas publicaban artículos eminentemente descriptivos en los que las observaciones solían guardar un orden cronológico, este modelo comienza a quedar obsoleto con el desarrollo de la ciencia y de su metodología a mediados del siglo XIX. El gran impulso a la ciencia en Estados Unidos tras la Segunda Guerra Mundial motivó la proliferación de las publicaciones científicas. La creciente presión sobre las revistas propició que estas comenzaran a exigir un mayor nivel de redacción y concisión, así como la estructuración de los manuscritos, lo que desembocó en la utilización generalizada del

formato IMRyD (Introducción, Métodos, Resultados y Discusión), que el American National Standards Institute estableció como norma en 1972. La popularidad de este sistema radica en que representa una progresión lógica desde el problema hasta su solución y en que es un sistema de comunicación uniforme, conciso y fácilmente comprensible que le facilita al científico medio la actualización de conocimientos, ya que este tiene que enfrentarse a la lectura de un número muy elevado de artículos durante su carrera. La Tabla 3 resume la función y características principales de cada sección en la estructura IMRyD.

| Sección | Función | Características |
|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Introducción (I) | Justifica la necesidad del trabajo (Gil-Alberdi González, 2009) y motiva a su lectura completa (Villagrán y Harris, 2009). Suministra los antecedentes necesarios (estado de la cuestión) para que el lector no necesite consultar otras publicaciones (Day, 2005). | Deben seleccionarse cuidadosamente las referencias para ofrecer un contexto relevante (Day, 2005). Esquema clásico de tres subsecciones: (1) estado de la cuestión, (2) identificación del problema no resuelto que se aborda e (3) hipótesis y objetivos de la investigación. Debe ser breve, concisa y estar redactada en tiempos verbales en presente (Villagrán y Harris, 2009). |
| Materiales y Métodos (M) | Proporciona toda la información necesaria como para que cualquier otro investigador competente pueda repetir el experimento y cumplir con el principio de reproducibilidad de la ciencia (Day, 2005). | Sección más explícita y detallada. Debe ser muy clara (Day, 2005). Solo describe los métodos novedosos (Villagrán y Harris, 2009). Uso de citas bibliográficas que suelen presentarse en orden cronológico. Es recomendable hacer coincidir los subapartados de los métodos con los epígrafes de los resultados (cohesión). Predomina la redacción en pasado (Day, 2005). |
| Resultados (R) | Incluye los nuevos conocimientos adquiridos a través del estudio. Función expositiva. Presenta los resultados de forma clara y objetiva, pero sin comentarlos (Gil-Alberdi González, 2009) o interpretarlos (Villagrán y Harris, 2009). | Debe redactarse en pretérito (Day, 2005). Debe estar bien estructurada para facilitar su lectura. Suele ser imprescindible incluir elementos gráficos, tales como tablas y figuras, para facilitar la comprensión de los datos, pero el texto solo debe enfatizar los hallazgos más relevantes (Gil-Alberdi González, 2009). Estos elementos deben ir siempre referenciados en el texto (Villagrán y Harris, 2009). Deben incluirse los resultados negativos, inesperados o contradictorios (Day, 2005). |
| Discusión (D) | Muestra las relaciones existentes entre los hechos observados (Day, 2005). Explica el valor de los resultados (Gil-Alberdi González, 2009) y los relaciona con los objetivos, la hipótesis inicial y el estado de la cuestión. Describe las fortalezas y debilidades del estudio, establece las preguntas que permanecen sin contestación y propone futuras líneas de investigación. Incluye las conclusiones. (Villagrán y Harris, 2009). | Un cierto grado de subjetividad y de especulación por parte del autor no solo es admisible, sino deseable (Gil-Alberdi González, 2009). Se redacta en tiempo presente (Villagrán y Harris, 2009) y no debe ser excesivamente verbosa (Day, 2005). |

Tabla 3. Función y características de cada una de las secciones del artículo científico en la estructura IMRyD.

Elaboración propia.

En el artículo científico además aparecen otros elementos fundamentales previos al núcleo del trabajo (IMRyD) cuya función es contribuir a la búsqueda e indización de este (Day, 2005). Estos elementos son: (1) el título, (2) el resumen o *abstract* y (3) las palabras clave o *keywords* (Tabla 4).

| Sección | Función | Características |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Título | Es el menor número de palabras que describen el contenido de un artículo científico. Función informativa. Facilita la indización, búsqueda y recuperación del artículo (Day, 2005). | Debe ser lo más conciso posible (Gil-Alberdi González, 2009). Debe cuidarse la sintaxis y el orden de las palabras y no ser demasiado corto ni general (Day, 2005). Debe ser atractivo y usar una de las siguientes estrategias: (1) tema de la tesis, (2) pregunta central, (3) destacar la controversia o (4) destacar la conclusión (Villagrán y Harris, 2009). No debe incluir abreviaturas, comillas, palabras subrayadas (Gil-Alberdi González, 2009), palabras vacías o superfluas (Villagrán y Harris, 2009), ni jerga, palabras patentadas o términos obsoletos (Day, 2005). |
| Resumen | Sumario breve y claro de cada una de las secciones del artículo. Permite identificar rápidamente el contenido y su pertinencia. Facilita la indización, búsqueda y recuperación del artículo (Day, 2005). Suele ser la única sección de libre acceso (Gil-Alberdi González, 2009). | Debe ser tratado como una entidad textual en sí misma que se puede leer sin consultar otras secciones (Gil-Alberdi González, 2009). Extensión típica de un máximo de 250 palabras. Predominan los tiempos verbales en pasado (Villagrán y Harris, 2009). No debe incluir referencias bibliográficas. Se deben evitar las abreviaturas (Gil-Alberdi González, 2009). |
| Palabras clave | Facilita la indización, búsqueda y recuperación del artículo (Day, 2005). | Es recomendable elegir entre 3 y 10 términos representativos que ya hayan sido descritos (Gil-Alberdi González, 2009). |

Tabla 4. Función y características de los elementos iniciales del artículo científico. Elaboración propia.

3. EL LENGUAJE CONTROLADO

Definición

Tal y como apunta Kuhn (2014), no existe una definición clara y consensuada de lenguaje natural controlado ni de sus propiedades características. Esto se debe, en gran medida, a que este tipo de lenguajes emergen en diferentes ámbitos (industrial, académico y gubernamental), en múltiples disciplinas (ciencia computacional, filosofía, lingüística e ingeniería) y a que han experimentado una evolución a lo largo de décadas (desde 1930), dando como resultado una gran diversidad. De la misma manera, tampoco existe una denominación consistente y los términos *lenguaje natural controlado*, *lenguaje controlado*, *lenguaje natural restringido* y *lenguaje simplificado* a menudo se utilizan indistintamente. Por razones de brevedad, en el presente trabajo se utilizará *lenguaje controlado* (LC).

Huijsen (1998, citado O'Brien, 2019) define el lenguaje controlado como la restricción explícita del lenguaje natural sobre el léxico, la gramática y el estilo. Kuhn (2014) considera que el lenguaje controlado es un lenguaje construido basado en un lenguaje natural, aunque es más restrictivo en cuanto al léxico, la sintaxis o la semántica, a la par que preserva gran parte de sus propiedades naturales y señala que, en general, los lenguajes controlados son más formales que los lenguajes naturales, pero más naturales que los lenguajes formales.

En el contexto de la traducción, Ramírez Polo (2012) define el lenguaje controlado como un conjunto de elementos del lenguaje utilizado con fines o en dominios específicos, dirigido a un grupo de destinatarios concreto, que se caracteriza por presentar restricciones en el vocabulario, la gramática y el estilo. Recalca, asimismo, que el lenguaje controlado se utiliza fundamentalmente en la creación de documentación técnica debido tanto a las características inherentes de estos tipos de texto, como al contexto empresarial globalizado en el que se producen, que a menudo requiere de la traducción de grandes volúmenes de texto.

Clasificación

Como se ha mencionado anteriormente, existe una gran diversidad de lenguajes controlados, lo que dificulta enormemente la creación de una taxonomía. Como consecuencia, la clasificación suele centrarse en cuestiones extralingüísticas relacionadas con el ámbito de aplicación, con la excepción de la clasificación PENS, introducida por Kuhn (2014), que se basa en cuatro propiedades del lenguaje: (1) precisión, (2) expresividad, (3) naturalidad y (4) simplicidad. En la línea de clasificación pragmática, los lenguajes controlados se dividen siguiendo la propuesta de Huijsen (1998, citado en O'Brien, 2019) en lenguajes orientados a personas o «Human-Oriented Controlled Languages» (HOCL), que tienen como fin mejorar la legibilidad, la comprensibilidad y la traducción humana, y los lenguajes orientados a máquinas o «Machine-Oriented Controlled Languages» (MOCL), que pretenden facilitar el procesamiento del lenguaje natural, por ejemplo, para mejorar la traducibilidad automática. El autor también señala que resulta difícil discernir entre HOCL y MOCL, ya que la simplificación del lenguaje tiene efectos en ambas direcciones (Huijsen, 1998, citado en Kuhn, 2014). Los lenguajes destinados al doble fin serían considerados como lenguajes de orientación dual o «Dual-Oriented Controlled Languages» (DOCL) (O'Brien, 2019).

Diseño y aplicación

Los lenguajes controlados están compuestos por reglas o directrices que los redactores deben seguir (O'Brien, 2019) con el fin de mejorar la traducibilidad, la comprensibilidad, la legibilidad o la usabilidad de un texto (O'Brien, 2010). A la hora de seleccionar las reglas, resulta fundamental tener en cuenta el tipo de lenguaje controlado a aplicar. Las reglas destinadas a los MOCL deben ser precisas y procesables por la computadora, mientras que las reglas HOCL pueden ser más vagas (Huijsen, 1998, citado en O'Brien, 2019).

El diseño de lenguajes controlados no es sencillo, en primer lugar, porque pueden existir interferencias entre las cualidades que se busca mejorar en un texto. Las reglas que mejoran la traducibilidad automática no siempre producen una mayor comprensibilidad y legibilidad. Tampoco es tarea fácil determinar qué reglas pueden tener el efecto deseado (Ramírez Polo, 2012), por lo que la experimentación es un paso imprescindible.

Mitamura y Nyberg (1995) definen tres niveles de aplicación de las reglas de lenguaje controlado: (1) léxico, (2) gramatical y (3) textual, mientras que Adriaens (1995) considera que existe un nivel adicional: (4) puntuación y control de caracteres. El área de control del lenguaje más común es el control léxico, que afecta a dos áreas: (1) el control del vocabulario del lenguaje natural, para evitar la sinonimia, la homonimia y demás fenómenos causantes de ambigüedad e imprecisión, y (2) el control de la terminología especializada. El control gramatical es más diverso y se extiende a la sintaxis, la morfología, la ortografía y el estilo (Ramírez Polo, 2012). El nivel textual se centra en la estructura del texto y en la dimensión pragmática, pero su aplicación parece ser residual (O'Brien, 2003).

O'Brien (2003) estudió en profundidad 8 lenguajes controlados basados en el inglés con el fin de identificar las reglas más comunes entre ellos e intentar diseñar un listado básico que pudiera ser de utilidad para cualquier interesado en su aplicación. Sorprendentemente, y a pesar de que los lenguajes controlados a menudo abordan problemas similares, los resultados revelaron que no parece haber una forma única de solventarlos, ya que la diversidad de reglas identificadas fue muy elevada (61 reglas en 3 niveles). De todas ellas, solo se detectó un denominador común a todos los casos de estudio: *frases de menos de 20-25 palabras*. Esta

regla refleja el consenso en cuanto a que las frases muy largas son más difíciles de entender por parte de los humanos y más difíciles de procesar para las máquinas. Los resultados sugieren que la creación de lenguajes controlados es, en gran medida, un proceso individual. La autora achaca esta enorme diversidad a: (1) los objetivos (mejora de la legibilidad, la comprensibilidad, la traducibilidad, etc.), (2) el tipo de sistema de traducción automática o el par de idiomas, (3) la influencia de las reglas de redacción corporativas y de los autores, o (4) la mera subjetividad. Ante la imposibilidad de producir una lista de reglas comunes a los lenguajes controlados en inglés, la autora proporciona un listado de reglas de utilidad seleccionadas en base a su experiencia.

Trabajar con un número reducido de reglas no es un problema para el redactor, pero a medida que estas aumentan en número y complejidad, no resulta razonable esperar que el redactor pueda recordarlas todas. Algunos lenguajes controlados son bastante flexibles, mientras que otros presentan un control muy estricto, pudiendo alcanzar las 60 reglas, lo que dificulta enormemente el trabajo de redacción (O'Brien, 2019). Independientemente del tipo y características del lenguaje controlado, del destinatario o de su finalidad, su aplicación práctica es muy exigente a nivel cognitivo, por lo que suele requerirse la automatización del control léxico-terminológico (Ramírez Polo, 2012) y el uso de herramientas denominadas «Controlled Language Checkers» (verificadores o correctores de lenguaje controlado) o «Assisted Authoring Tools» (asistentes para la creación de contenidos). Algunos ejemplos de estas herramientas son MAXit Checker, Acrolinx o HyperSTE, todos ellos centrados prácticamente en exclusividad en el inglés (O'Brien, 2019).

Lenguaje controlado y traducción

Si ya resulta difícil redactar teniendo en cuenta al lector (De Biase *et al.*, 2009), la tarea se complica si además hay que considerar a una audiencia global (O'Brien, 2019).

En el ámbito de la traducción se asume ampliamente que para que la traducción automática produzca resultados de alta calidad, o de calidad publicable, es necesario un proceso de posesición realizado por parte de un profesional. Medir el esfuerzo de posesición es crucial a la hora de evaluar la utilidad de los sistemas de traducción automática, además de constituir

una estimación indirecta de la traducibilidad automática de un texto, mucho más difícil de cuantificar (O'Brien, 2004), ya que la traducibilidad no es un concepto absoluto y varía según el tipo de texto y el par de lenguas, entre otros factores (Ramírez Polo, 2012). La cuantificación del esfuerzo de posesición puede realizarse a tres niveles: (1) temporal o tiempo dedicado, (2) técnico o número de cambios realizados en el texto y (3) cognitivo o tipo de procesos cognitivos involucrados (Klings, 2001, en O'Brien, 2004). Algunas estrategias permiten reducir el esfuerzo de posesición, como el uso de memorias de traducción, de diccionarios y bases de datos terminológicas, o de lenguajes controlados (O'Brien, 2004), aplicados en origen mediante un proceso de preedición.

El estudio del esfuerzo de posesición permite asimismo evaluar el impacto de las reglas de lenguaje controlado. Si el esfuerzo de posesición se reduce, se considera que el lenguaje controlado ha tenido un impacto positivo. Frecuentemente, las reglas de lenguaje controlado simplemente reflejan buenas prácticas de redacción, ya que un texto bien redactado presenta una mejor traducibilidad automática (O'Brien, 2019).

O'Brien (2004) apunta que uno de los principales objetivos del lenguaje controlado es mejorar la traducibilidad y que la mayoría de los estudios centrados en medir la traducibilidad automática se basan en la identificación de las características lingüísticas problemáticas de un texto o indicadores de traducibilidad. Como ejemplos del inglés menciona: la voz pasiva, el gerundio, las frases nominales largas y la elipsis. También recalca que algunos indicadores de traducibilidad parecen ser más problemáticos en determinados pares de idiomas o motores de traducción automática. En su estudio, cuyo objetivo fue identificar la relación entre la traducibilidad y el esfuerzo de posesición en textos traducidos del inglés al alemán, los resultados sugieren que la aplicación del lenguaje controlado mejora la traducibilidad y reduce el esfuerzo de posesición, aunque este proceso siempre sigue siendo necesario.

En otro trabajo, O'Brien (2010) estudió el efecto de la aplicación del lenguaje controlado sobre la legibilidad y la traducibilidad automática del inglés al español, al francés y al chino simplificado de textos técnicos con una calidad de redacción alta de partida. El estudio de la legibilidad es un aspecto crucial para el desarrollo de lenguajes controlados de tipo HOCL y DOCL y parte de la base de que, si los lenguajes controlados reducen efectivamente la

ambigüedad de un texto, la legibilidad debería mejorar, así como su traducibilidad automática. Sin embargo, según este estudio, los textos no controlados no resultaron ser más difíciles de leer que los controlados. Además, el estudio sugiere que la variable que más afecta a la mejora de la legibilidad es la complejidad del texto. No se observó ninguna mejora en la calidad del producto de la traducción automática en los textos controlados.

Rossetti y O'Brien (2019) evaluaron la utilidad de un verificador de lenguaje controlado (Acrolinx) a la hora de redactar resúmenes de temática médica simplificados en cuanto al contenido y al lenguaje empleado (traducción intralingüística), así como el impacto de esta herramienta sobre la traducibilidad automática de los textos del inglés al español. El estudio constató que el empleo del verificador de lenguaje controlado no tuvo gran impacto sobre la traducibilidad automática y que Google Translate produjo un resultado de gran calidad. Sin embargo, destacan que el uso conjunto de verificadores de lenguaje controlado y de motores de traducción automática puede reducir la carga de trabajo y aumentar la productividad.

Situación de los lenguajes controlados en español

Dada la clara hegemonía del inglés a nivel global, no es de extrañar que la mayoría de los lenguajes controlados empleados en la actualidad estén basados en este (Kuhn, 2014) y que, por consiguiente, las herramientas desarrolladas para su implementación también lo estén (O'Brien, 2019). En su exhaustiva revisión sobre el estado del arte del lenguaje controlado, González Sigüero (2017) solo pudo identificar dos aplicaciones relevantes en español: (1) la propuesta de Ruiz Cascales y Sutcliffe (2003) de un lenguaje controlado para la simplificación del español técnico empleado en los manuales de mantenimiento de aeronaves (equivalente al inglés simplificado AECMA) y su correspondiente *parser* (analizador sintáctico), así como (2) el español técnico simplificado desarrollado por Gobbi (2015). Asimismo, en cuanto a las herramientas informáticas, el autor solo destaca otra aplicación relacionada con la simplificación del español, desarrollada por Saggion *et al.*, (2015).

Durante esta revisión no ha sido posible hallar estudios sobre aplicaciones del lenguaje controlado orientadas a la mejora de la traducibilidad automática del español al inglés. Tampoco se han identificado herramientas de ayuda a la implementación de lenguajes

controlados en español que permitieran la modificación de las reglas y que pudieran ser de utilidad práctica para el desarrollo de este estudio. Debido a la carencia de investigaciones centradas en el español y relacionadas con el ámbito que atañe al presente trabajo, así como a los objetivos de este, no resulta pertinente realizar una extensa revisión bibliográfica sobre los lenguajes controlados basados en el inglés. Los trabajos mencionados en los apartados anteriores de esta sección se han incluido por la utilidad de sus conclusiones para el diseño de la metodología del presente trabajo y el análisis de sus resultados.

4. MATERIALES Y MÉTODOS

Selección del texto

A partir de una búsqueda a través del Centro de Información para Profesionales y Pacientes sobre el Coronavirus de Elsevier, un punto de información que incluye artículos de libre acceso (www.elsevier.es/es-coronavirus), se seleccionó un único manuscrito desde el listado general de artículos científicos del ámbito de la medicina redactados en español y publicados en revistas especializadas, lo que constituye una garantía de que el texto cumple con los estándares básicos de redacción y estilo típicos del género (estructura y lenguaje de especialidad). La búsqueda se realizó sin filtros por especialidad o temática. Se descartaron todos aquellos textos que correspondían a cartas al editor, editoriales, casos clínicos, artículos breves, comunicaciones en foros, actualizaciones, correspondencia, análisis de imágenes y casos de estudio concretos hasta encontrar un artículo científico original en el listado ordenado por fecha de publicación. Durante la selección, la autora no leyó el contenido del artículo para evitar familiarizarse con el texto. Solo se prestó atención al título para verificar que coincidía con la temática de interés (medicina, epidemiología, COVID-19) y se comprobó que el artículo seguía la estructura IMRyD o una variación de esta, ya que pueden existir ligeras variaciones según los requerimientos de cada revista. El artículo se dividió por secciones en diferentes documentos en el procesador de textos Microsoft Word. Asimismo, se extrajo todo el texto contenido de tablas y figuras y se eliminaron los hipervínculos a las referencias bibliográficas y las fórmulas matemáticas.

Preedición y aplicación del lenguaje controlado

Los textos correspondientes a cada una de las secciones del artículo o textos brutos (TB) fueron sometidos a un pretratamiento o preedición consistente en la aplicación de una serie de reglas de lenguaje controlado. Se elaboró un listado tentativo de reglas teniendo en cuenta las características del discurso del artículo científico (Sección 2), en concreto, los errores típicos en el lenguaje de especialidad (Tabla 2), las características de las secciones de este tipo de textos (Tabla 3) y las aplicaciones del lenguaje controlado recopiladas (Sección 3). Asimismo, se seleccionaron algunas reglas en base a problemas concretos de la traducción en el par de idiomas bajo estudio que, si bien no quedan explícitamente recogidos en las secciones anteriores, según la experiencia de la autora en el campo de la traducción automática y de la posesición de artículos científicos, podrían ser de utilidad.

Este proceso de análisis y reflexión sobre el discurso, el lenguaje controlado y la traducibilidad automática se realizó teniendo en cuenta tanto el par de idiomas, como el motor de traducción automática a emplear y los objetivos de la aplicación (O'Brien, 2019; Ramírez Polo, 2012). Esto fue fundamental a la hora de seleccionar el conjunto de reglas, ya que se pretendía desarrollar una propuesta de lenguaje controlado para mejorar la traducibilidad ES>EN de artículos científicos del ámbito médico a través de motores de traducción automática basados en redes neuronales, reducir el esfuerzo de posesición y mejorar la calidad del texto resultante para alcanzar una calidad publicable sin comprometer la calidad del texto original (lenguaje controlado tipo DOCL) (O'Brien, 2019). La propuesta inicial de reglas de lenguaje controlado incluye tanto el control léxico-terminológico, como el gramatical.

Control léxico-terminológico (CLT)

El control léxico-terminológico aplicado tiene como fin desambiguar el léxico del lenguaje natural y solventar los problemas relacionados con la polisemia, la homonimia o la sinonimia (O'Brien, 2003; Ramírez Polo, 2012), verificar la terminología especializada (Ramírez Polo, 2012) para mejorar la traducibilidad y eliminar los anglicismos innecesarios y los falsos amigos (Aleixandre Benavent y Amador Iscla, 2001a).

Para llevar a cabo el control léxico-terminológico, en lugar de establecer reglas que aumentasen la complejidad del lenguaje controlado y dificultasen su aplicación (O'Brien, 2019), se optó por verificar el léxico y la terminología con la ayuda de diccionarios y recursos especializados, tales como la base terminológica de Cosnautas (www.cosnautas.com/es), el glosario específico sobre COVID-19 de Tremédica (www.tremedica.org/tremediteca/glosarios/diccionario-de-covid-19-en-es/) y las convenciones y nomenclaturas recogidas en los cuadernos de redacción de la Fundación Dr. Antonio Esteve (Giba, 2014; Navarro, 2009). Ante la ausencia de herramientas para detectar el léxico problemático y prever todas sus posibles traducciones, este proceso se realizó de forma manual y ciertamente subjetiva. Cabe asumir, por tanto, que un cierto volumen de léxico problemático podría haberse pasado por alto durante la preedición.

Control gramatical (CG)

El control gramatical se realizó a través de un conjunto de reglas de fácil aplicación seleccionadas para mejorar, tanto la calidad del texto de partida, como su traducibilidad. Dado que la complejidad del lenguaje controlado afecta a su implementación (O'Brien, 2019; Rossetti y O'Brien, 2019), se optó por una selección tentativa de diez reglas (Tabla 5). Este formato reducido tiene como objeto facilitar su aplicación por parte de la autora, que realizó la preedición de los textos de forma manual ante la ausencia de herramientas específicas (González Sigüero, 2017).

| Regla | Explicación |
|--------------|---------------------------------------------------------------------------------------------|
| CG1 | Acortar las frases (aprox. 30 palabras) |
| CG2 | Respetar orden natural de elementos sintácticos, acortar distancias sujeto-verbo |
| CG3 | Solo una subordinada por frase |
| CG4 | Evitar elipsis (especialmente cuando el referente se encuentra en frases diferentes) |
| CG5 | Evitar anáfora o desarrollarla de forma explícita |
| CG6 | Uso de verbos simples (evitar perífrasis verbales) |
| CG7 | Uso de tiempos verbales simples |
| CG8 | Evitar la doble negación |
| CG9 | Evitar gerundios en estructuras anglicadas |
| CG10 | Evitar la voz pasiva (perifrástica-anglicismo en todo caso, y refleja cuando existe sujeto) |

Tabla 5. Reglas de lenguaje controlado seleccionadas.

Tal y como puso de manifiesto la revisión llevada a cabo por O'Brien (2003), la redacción en frases cortas (por debajo de 20-25 palabras) parece ser la única regla común y consensuada en los lenguajes controlados en inglés, ya que las frases muy largas resultan problemáticas tanto para los humanos, como para las máquinas. Dado que la economía del lenguaje también es una característica más que deseable en la redacción de textos científicos (Giba, 2014; Pulido, 2010) se optó por incluir esta regla de control gramatical (CG1). Debido a que el español es un idioma más verboso, se extendió el límite de palabras a un máximo de 30, lo que resulta más que razonable teniendo en cuenta las diferencias en el discurso científico entre ambos idiomas (Claros Díaz, 2016).

En su manual para la escritura científica, Giba (2014) señala la importancia del orden de las palabras en una frase y la distancia que existe entre el sujeto y el verbo en una oración: a menor sea esta, más fácil será de leer y comprender. Aunque sus recomendaciones están centradas en el inglés científico, la autora ha considerado pertinente formularlas en la regla CG2, ya que, presumiblemente, una frase bien ordenada en español no solo es más comprensible, sino que además se traduce mejor.

La longitud de una frase y su complejidad son variables estrechamente relacionadas. La literatura parece ofrecer consenso acerca de la necesidad de evitar construcciones extremadamente complejas, tanto en español (Pulido, 2010), como en inglés (Giba, 2014; O'Brien, 2003). Con el objetivo de mejorar la redacción del texto bruto y el producto de la traducción automática, se optó por limitar el número de subordinadas por frase a una (CG3).

Si bien la redacción científica debe buscar la cohesión textual (Giba, 2014), algunos recursos que contribuyen a ella no siempre son deseables en los textos científicos y pueden afectar negativamente a su traducibilidad automática (O'Brien, 2004). Tanto Ramírez Polo (2012) como O'Brien (2003) recomiendan en sus propuestas de reglas de lenguaje controlado evitar la elipsis (CG4) y la anáfora (CG5).

Para cada par de idiomas existen indicadores de traducibilidad automática concretos (O'Brien, 2004). La autora estimó necesario incluir una limitación al uso de las perífrasis verbales (CG6), de elevada complejidad y diversidad en español, ya que presumiblemente

pueden ser altamente problemáticas para la traducibilidad automática entre el español y el inglés (Martínez Llenas, 2009). En aras de evitar la excesiva complejidad del texto de partida (Giba, 2014; Pulido, 2010), la propuesta de lenguaje controlado incluye reglas que afectan a la selección de los verbos empleados de acuerdo con las recomendaciones de O'Brien (2003), que sugiere limitar el uso de tiempos verbales complejos (CG7).

Giba (2014) recomienda evitar el uso de la doble negación en inglés, ya que añade complejidad innecesaria al texto. Según el caso, una doble negación en español puede ser traducida como una afirmación en inglés, lo que puede constituir un problema de traducción. Por tanto, se optó por limitar su uso en el texto a través de la regla CG8.

El español científico se aqueja de un problema de interferencia del inglés, que se manifiesta, con frecuencia, en el calco de construcciones que no son naturales en español, entre las que cabe destacar el abuso del gerundio en estructuras anglicadas (Amador Iscla y Aleixandre Benavent, 2002). Su uso también aparece como desaconsejado en gran parte de los lenguajes controlados en inglés analizados por O'Brien (2003), ya que podría reducir la traducibilidad (O'Brien, 2004). Parece razonable, por tanto, limitar su uso en español en el texto de partida (CG9), no solo con el objetivo de mejorar la traducibilidad automática, sino con el fin de eliminar errores típicos en el discurso médico en español.

Por último, la regla CG10 aborda la cuestión del uso de la pasiva, limitando su uso. Si bien la voz pasiva es un recurso frecuente en el discurso científico para transmitir impersonalidad y objetividad, su empleo generalizado resulta abusivo (Carmona Sandoval, 2013). Amador Iscla y Aleixandre Benavent (2002) señalan que su abuso en el lenguaje médico escrito puede atribuirse a la influencia del inglés y, tanto en la forma perifrástica, como en la refleja (más común), es extraño al español. También recuerdan que la voz pasiva no es ni más científica ni objetiva que la voz activa, sino más imprecisa. Por su parte, Giba (2014) aclara que la voz pasiva cada vez se usa menos en el lenguaje científico en inglés por la misma razón y que la mayoría de las publicaciones consideran que el uso de la voz activa es preferible, ya que es un discurso más directo y fácil de procesar cognitivamente.

Corrección de errores (CE) y corrección ortotipográfica (CO)

Asimismo, durante la preedición se realizó una corrección de errores de ortografía, adición de palabras, errores gramaticales, etc., y una corrección ortotipográfica para eliminar espacios adicionales y errores de puntuación y darle consistencia ortotipográfica al texto.

Cambios de estilo esenciales (EE)

Para poder aplicar las reglas de lenguaje controlado es necesario realizar cambios adicionales que afectan al estilo. Por ejemplo, para acortar frases es posible que haya que dividir las y usar conectores entre las frases resultantes, o reescribirlas por completo. Todo cambio de estilo con este fin se consideró como un cambio esencial en la preedición. Una misma frase puede contener varios cambios de este tipo si se han aplicado diversas reglas, o un mismo cambio puede servir varios propósitos. Para evitar la sobrerepresentación de los cambios de estilo esenciales, solo se anotó una vez por frase modificada en lugar de por palabras (Parra Escartín y Goulet, 2020), aunque la frase estuviese afectada por varios cambios de este tipo complementarios a la aplicación del lenguaje controlado.

La preedición se realizó en el procesador de textos Microsoft Word con el control de cambios activado. Cada cambio realizado en el texto controlado (TC) fue anotado con el tipo de regla que se había aplicado o el tipo de corrección realizada (CE, CO) con el fin de cuantificar posteriormente el esfuerzo técnico de preedición y determinar la frecuencia de aplicación de cada una de las reglas, así como su utilidad. No se introdujeron cambios no esenciales durante la preedición y, si durante el proceso de anotación se estimó algún cambio de estilo como no esencial para aplicar las reglas de lenguaje controlado, corregir errores o mantener el estilo del texto bruto en la medida de lo posible, el cambio fue revertido.

Esfuerzo técnico de preedición

Dado que no existen herramientas de aplicación del lenguaje controlado de utilidad para el presente trabajo, resulta necesario evaluar el esfuerzo técnico de preedición de manera similar a como se evalúa el esfuerzo de posesición para así obtener una perspectiva más completa sobre la eficiencia del flujo de trabajo propuesto.

Tras el proceso de anotación, se cuantificó el número de ocasiones en las que se aplicó cada regla de lenguaje controlado (CLT, CG) a los textos brutos correspondientes a las secciones del artículo bajo estudio, así como todos los cambios de estilo esenciales implementados para poder aplicarlas (EE) y todas aquellas modificaciones relacionadas con la corrección de errores (CE) y la corrección ortotipográfica (CO). Este procedimiento permite determinar qué reglas fueron aplicadas de forma más amplia, si existen diferencias entre secciones y en qué medida fueron útiles la corrección de errores y la corrección ortotipográfica.

Traducción automática

La traducción automática ES>EN de los textos brutos (artículo original por secciones) y controlados (artículo preeditado por secciones) se llevó a cabo a través de Google Translate, un motor de traducción automática basado en redes neuronales de libre acceso, gratuito, ampliamente conocido y ya empleado con frecuencia por profesionales del ámbito científico-médico para redactar y leer artículos científicos (Parra Escartín *et al.*, 2017). Es posible realizar la traducción en Google Translate de tres maneras: (1) pegando el texto en la interfaz en el buscador de Google, (2) pegando el texto en la interfaz desde el enlace o (3) subir cada documento en el propio enlace. Se optó por la segunda opción tras realizar una prueba con un fragmento del texto controlado y observar que las otras opciones introducían un número muy elevado de espacios adicionales.

Posedición

La posedición de los textos traducidos mediante Google Translate se realizó reproduciendo, en la medida de lo posible, las condiciones de trabajo habituales del sujeto (la propia autora) y empleando el procesador de textos Microsoft Word con control de cambios activado. Como ventaja de emplear un único sujeto para el experimento destaca la consistencia en términos de capacidades, formación, conocimiento y experiencia a la hora de abordar la posedición, que sería difícil de alcanzar empleando a varios sujetos.

La hipótesis de estudio presume que el texto controlado se traducirá mejor por parte del motor de traducción automática y contendrá menos errores a subsanar durante la posedición. Por ello, abordar la posedición del texto bruto en primer lugar teóricamente impediría que

soluciones traductológicas mejores motivadas por los resultados de la traducción automática de los textos controlados pudieran afectar positivamente a la posesición del texto bruto debido a la familiarización del sujeto con los textos.

Una vez completada la posesición, se procedió a realizar la anotación de los cambios realizados con el fin de clasificar y cuantificar el error. Son muy numerosas las clasificaciones del error para la evaluación de la traducción automática y no existe un marco consensuado para clasificar los cambios implementados durante el proceso de posesición (de Almeida, 2013), por lo que no resultó pertinente para los propósitos del presente trabajo el realizar una extensa revisión o elaborar una clasificación demasiado compleja. Por ello, se optó por una aproximación más pragmática y orgánica y se empleó un marco de referencia simplificado con 6 categorías de error: ortotipografía (OT), gramática (G), semántica (S), estilo (E), adición de palabras (A) y omisión de palabras (O). Durante la anotación, se describió el tipo de error dentro del marco de referencia para después agruparlos y cuantificarlos según la subcategoría a la que se adscriben. Por razones obvias, esta clasificación no es extrapolable a otros textos en los que pueden aparecer otros tipos de error. La anotación permite cuantificar el error, estudiar su distribución por secciones, determinar el esfuerzo técnico de posesición como indicador indirecto de la traducibilidad automática y realizar una comparativa entre los resultados del texto bruto y del texto controlado.

Esfuerzo técnico de posesición

Debido a que el mismo sujeto (la autora) realizó la posesición de ambos tipos de texto (brutos y controlados), existe un sesgo de memoria positivo por familiarización con el texto que podría afectar al estudio del esfuerzo de posesición si, por ejemplo, solo se estudiase el esfuerzo temporal. Para mitigar esta limitación metodológica se llevó a cabo una evaluación del esfuerzo técnico de posesición. Este procedimiento permite comparar los resultados entre los textos brutos y controlados, determinar qué errores son más frecuentes y si existen diferencias entre secciones.

Eficacia del lenguaje controlado y prevención del error

Con el fin de evaluar la efectividad de los cambios realizados durante la preedición, se realizó un tercer proceso de anotación comparativo mediante el estudio pormenorizado, frase por frase, del texto controlado y las traducciones automáticas del texto controlado y del texto bruto. Se consideró como error prevenido aquel error presente en la traducción del texto bruto pero ausente en la del texto controlado como consecuencia de un cambio realizado y anotado durante la preedición. Este proceso permite determinar qué reglas de control del lenguaje tuvieron un impacto relevante sobre la traducibilidad del texto y el esfuerzo técnico de posesición, qué reglas tuvieron un efecto menor, nulo, o incluso negativo, y detectar los problemas de traducción automática que no pudieron ser solventados mediante el lenguaje controlado propuesto. Asimismo, la anotación comparativa permite indagar en la posible presencia de errores generados, entendidos como errores presentes en la traducción del texto controlado como consecuencia de un cambio realizado y anotado durante la preedición, pero ausentes en la del texto bruto. Por último, este proceso también resulta de utilidad a la hora de estudiar en mayor profundidad el posible léxico problemático.

5. RESULTADOS

Características generales del texto objeto de estudio

El texto seleccionado es un artículo científico de índole teórica, cuyo tema principal es la propuesta de un algoritmo de decisión para prevenir la diseminación del SARS-CoV-2 en Argentina (Anexo I). Este artículo cuenta con una variación de la estructura IMRyD descrita en la Sección 2 y se articula en los siguientes apartados: (1) Título (18 palabras), (2) Resumen (275 palabras), (3) Palabras clave (8 palabras), (4) Introducción (794 palabras), (5) Materiales y Métodos (2545 palabras), (6) Resultados y Discusión (1833 palabras) y (7) Conclusiones (103 palabras), con un número total de 5576 palabras. Dado que el texto fue aceptado para su publicación puede considerarse, por tanto, que cuenta con una calidad publicable. El texto está redactado en español y resulta patente que los autores emplean la variedad dialectal del español en Argentina, que se pone de manifiesto, principalmente, a través de la elección léxica (*insumo* vs. *recurso*; *rubro* vs. *gremio*).

A pesar de que el artículo tiene como principal destinataria a la comunidad científica en español, el estilo de redacción revela como claro objeto que su contenido pueda ser asimilado por una audiencia más amplia con motivo de la pandemia actual. Según Rodríguez-Tapia (2020), el grado de especialización de un texto es un continuo afectado por la relación emisor-receptor. Así, este artículo podría considerarse en esencia como un texto semiespecializado. Los autores emplean un lenguaje llano y claro, tanto a nivel léxico como sintáctico. Abunda el léxico del lenguaje natural, con la ambigüedad añadida que conlleva, aparece léxico natural terminologizado (*aislamiento*), se emplean términos típicos de la epidemiología que se encuentran ampliamente extendidos o banalizados (*cuarentena, sensibilidad, prevalencia*), aparecen imprecisiones semánticas (*agente de contagio vs. vector de contagio*) y anglicismos léxicos de uso común en las variedades dialectales del español en Sudamérica (*testeo*). La densidad terminológica del texto es baja. No obstante, el artículo refleja la mayoría de las características esperables del artículo científico, incluidas una baja variedad léxica para aportar precisión y consistencia (Giba, 2014), reformulaciones (Rodríguez-Tapia, 2020), los anglicismos o calcos de estructuras (Amador Iscla y Aleixandre Benavent, 2002) y la presencia del sujeto impersonal (Carmona Sandoval, 2013).

Esfuerzo técnico de preedición: lenguaje controlado y corrección

La Tabla 6 recoge el número de ocasiones en que las reglas de lenguaje controlado se aplicaron sobre cada una de las secciones del artículo y en el texto en su conjunto.

| Sección | CLT | CG1 | CG2 | CG3 | CG4 | CG5 | CG6 | CG7 | CG8 | CG9 | CG10 | LC |
|------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------|----------|----------|-----------|------------|
| Título | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Resumen | 4 | 4 | 1 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 5 | 18 |
| Palabras clave | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Introducción | 7 | 11 | 0 | 1 | 0 | 10 | 6 | 3 | 0 | 1 | 3 | 42 |
| Materiales y Métodos | 30 | 25 | 1 | 5 | 8 | 16 | 4 | 2 | 0 | 3 | 37 | 131 |
| Resultados y Discusión | 8 | 15 | 10 | 2 | 4 | 2 | 4 | 1 | 0 | 3 | 43 | 92 |
| Conclusiones | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 6 |
| TOTAL | 52 | 57 | 12 | 10 | 12 | 28 | 16 | 6 | 0 | 8 | 88 | 289 |

Tabla 6. Número absoluto de ocasiones en las que se aplicó cada regla de lenguaje controlado (CLT y CG) a cada una de las secciones y al conjunto del texto.

La aplicación de las reglas de lenguaje controlado resultó en un total de 289 cambios realizados durante la preedición sobre el texto bruto (TB) en español para dar lugar al texto controlado (TC), de los cuales 52 correspondieron al control léxico-terminológico (17.99 % de los cambios totales). El texto preeditado resultante puede consultarse en el Anexo II.

Las reglas de lenguaje controlado no parecen haber sido aplicadas con homogeneidad a lo largo de las secciones, lo que podría deberse a que estas presentan una extensión variable y a que el estilo de redacción de cada una de ellas varía conforme a su propósito, tal y como recoge la Sección 2. Con el ánimo de contextualizar los resultados según la extensión de cada sección, en la Tabla 7 se reproducen los resultados anteriores, divididos por el número de palabras de cada sección y totales del texto bruto sobre el que se aplicó el lenguaje controlado, multiplicados por 100 para una mejor visualización numérica, dado que el orden de magnitud de los resultados es muy bajo.

| Sección | CLT | CG1 | CG2 | CG3 | CG4 | CG5 | CG6 | CG7 | CG8 | CG9 | CG10 | LC |
|------------------------|-------------|-------------|------|------|------|------|------|------|-------------|------|-------------|------|
| Título | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Resumen | 1.45 | 1.45 | 0.36 | 0.73 | 0.00 | 0.00 | 0.73 | 0.00 | 0.00 | 0.00 | 1.82 | 6.55 |
| Palabras clave | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Introducción | 0.88 | 1.39 | 0.00 | 0.13 | 0.00 | 1.26 | 0.76 | 0.38 | 0.00 | 0.13 | 0.38 | 5.29 |
| Materiales y Métodos | 1.18 | 0.98 | 0.04 | 0.20 | 0.31 | 0.63 | 0.16 | 0.08 | 0.00 | 0.12 | 1.45 | 5.15 |
| Resultados y Discusión | 0.44 | 0.82 | 0.55 | 0.11 | 0.22 | 0.11 | 0.22 | 0.05 | 0.00 | 0.16 | 2.35 | 5.02 |
| Conclusiones | 2.91 | 1.94 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.97 | 0.00 | 5.83 |
| TOTAL | 0.93 | 1.02 | 0.22 | 0.18 | 0.22 | 0.50 | 0.29 | 0.11 | 0.00 | 0.14 | 1.58 | 5.18 |

Tabla 7. Número relativo de ocasiones en las que se aplicó cada regla de lenguaje controlado (CLT y CG) a cada una de las secciones y al conjunto del texto; calculado al dividir el número de cambios realizados por el número total de palabras de cada sección y multiplicado por 100 para una mejor visualización numérica.

La regla de lenguaje controlado más ampliamente aplicada en el conjunto del texto fue CG10 *evitar la voz pasiva* (1.58), empleada fundamentalmente cuando existe un sujeto activo y claramente identificable de la acción, por ejemplo, los autores de la publicación. Su uso más destacable correspondió a las secciones de Resultados y Discusión (2.35), Resumen (1.82) y Materiales y Métodos (1.45). La siguiente regla más frecuentemente utilizada en el conjunto del texto fue CG1 *acortar las frases* a aproximadamente 30 palabras (1.02). Las

secciones con una mayor proporción de cambios con esta regla fueron las Conclusiones (1.94), el Resumen (1.45) y la Introducción (1.39), las secciones más cortas. Por último, el tercer tipo de control de lenguaje más frecuente en el conjunto del texto fue el control léxico-terminológico (CLT), con un valor relativo de 0.93. Destaca su uso en las secciones de Conclusiones (2.91), Resumen (1.45) y Materiales y Métodos (1.18). Cabe mencionar que, en el caso del Resumen y las Conclusiones, al ser secciones tan breves, muy pocos cambios originan una sobrerrepresentación del uso de las reglas aplicadas. Entre las reglas de control menos empleadas, sobresale especialmente la regla CG8 *evitar la doble negación*, que no se empleó en ninguna ocasión (0.00).

Las frases con cambios de estilo esenciales (EE) también fueron cuantificadas, tal y como recoge la Tabla 8. Como ya se ha mencionado anteriormente (Sección 4), este tipo de marcador se refiere a frases completas que pueden contener uno o varios cambios de estilo esencial y no cuantifica el número de cambios de estilo individuales por frase.

| Sección | CLT | CG | EE |
|------------------------|-----------|------------|-----------|
| Título | 0 | 0 | 0 |
| Resumen | 4 | 14 | 4 |
| Palabras clave | 0 | 0 | 0 |
| Introducción | 7 | 35 | 18 |
| Materiales y Métodos | 30 | 101 | 30 |
| Resultados y Discusión | 8 | 84 | 17 |
| Conclusiones | 3 | 3 | 5 |
| TOTAL | 52 | 237 | 74 |

Tabla 8. Número absoluto de ocasiones en las que se aplicó el lenguaje controlado (CLT y CG) y el número de frases con cambios de estilo esenciales (EE) para facilitar su aplicación, por secciones y en el conjunto del texto.

La predicción para la aplicación de las reglas de lenguaje controlado consistió en la aplicación de un total de 363 cambios de distinta naturaleza: 52 cambios léxico-terminológicos (CLT), 237 cambios gramaticales (CG) y 74 cambios de estilo esenciales asociados a los cambios gramaticales (EE), que podían incluir uno o varios cambios en la misma frase y cuyos ejemplos más significativos están representados por la necesidad de emplear conectores como consecuencia de acortar las frases y dividir las (CG1), reescribir las frases para eliminar subordinadas (CG3) y desarrollar las anáforas (CG5).

El proceso de preedición sirvió también para corregir errores en el texto fuente (CE) y realizar una corrección ortotipográfica (CO). Los errores corregidos (CE) fueron de variada tipología, incluidos los errores de concordancia de género y número, adiciones de palabras, omisiones de artículos y faltas de ortografía, entre otros. La corrección ortotipográfica (CO) se centró en la eliminación de espacios adicionales y en el correcto y consistente uso de los signos de puntuación y de las convenciones. La Tabla 9 recoge el número de cambios asociados a la corrección de errores (CE) y a la corrección ortotipográfica (CO).

| Sección | CE | CO |
|------------------------|-----------|-----------|
| Título | 0 | 0 |
| Resumen | 4 | 1 |
| Palabras clave | 0 | 0 |
| Introducción | 7 | 5 |
| Materiales y Métodos | 24 | 11 |
| Resultados y Discusión | 13 | 12 |
| Conclusiones | 3 | 0 |
| TOTAL | 51 | 29 |

Tabla 9. Número absoluto de ocasiones en las que se realizaron cambios para la corrección del error (CE) y la corrección ortotipográfica (CO), por secciones y en el conjunto del texto.

El número total de correcciones de error (CE) y correcciones ortotipográficas (CO) ascendió a 51 y a 29 para el conjunto del texto, respectivamente. Por tanto, junto con el control léxico-terminológico (CLT, n=52), el control gramatical (CG, n=237) y los cambios de estilo esenciales (EE, n=74) (Tabla 8) en total se aplicaron 443 cambios durante la preedición (esfuerzo técnico de preedición). Como se puede observar tanto en la Tabla 8 como en la Tabla 9, el grueso de los cambios se realizó en las secciones de Materiales y Métodos y Resultados y Discusión, tal y como cabría esperar dada su extensión.

Esfuerzo técnico de posesición: clasificación y cuantificación del error

Los textos resultantes del proceso de posesición de la traducción del texto bruto y del texto controlado pueden consultarse en el Anexo III y el Anexo IV, respectivamente. La Tabla 10 recoge los tipos de error detectados durante la posesición, agrupados según las 6 categorías de referencia (Sección 4).

| Subcategoría de error | Categoría de error | Descripción |
|--------------------------|-------------------------|------------------------------------------------------------------------------------------------------------------------|
| Puntuación | Ortotipografía (OT) | Errores de puntuación |
| Convención | Ortotipografía (OT) | Errores como consecuencia de la aplicación errónea de convenciones traductológicas o científicas |
| Mayúscula | Ortotipografía (OT) | Introducción de mayúsculas de forma arbitraria |
| Adición de espacios | Ortotipografía (OT) | Adición de espacios de forma arbitraria |
| Orden | Ortotipografía (OT) | Orden alterado en símbolos y combinaciones de palabras y símbolos o números |
| Tiempo verbal | Gramática (G) | Tiempo verbal erróneo o inadecuado según el estilo del texto |
| Preposición | Gramática (G) | Uso incorrecto de las preposiciones, con y sin implicaciones semánticas |
| Concordancia | Gramática (G) | Errores de concordancia de número |
| Precisión léxica | Semántica (S) | Imprecisión semántica como consecuencia de la ambigüedad léxica, la polisemia o los requerimientos de estilo del texto |
| Cambio de sentido | Semántica (S) | Cambios de sentido totales |
| Modificación del sentido | Semántica (S) | Modificaciones menores en el sentido |
| Naturalidad | Estilo (E) | Estructuras no naturales en el idioma de destino que no conllevan cambios semánticos ni errores gramaticales |
| Consistencia | Estilo (E) | Consistencia en la traducción de estructuras repetidas |
| Adición de palabras | Adición de palabras (A) | Adición de palabras sin implicaciones semánticas |
| Omisión de palabras | Omisión de palabras (O) | Omisión de palabras sin implicaciones semánticas |

Tabla 10. Listado de los errores observados durante la posesición según las categorías preestablecidas (ortotipografía, OT; gramática, G; semántica, S; estilo, E; adición de palabras, A; omisión de palabras, O) y su descripción general.

La Tabla 11 refleja el número total de cambios realizados en el conjunto del artículo durante la posesición de las traducciones automáticas del texto controlado (TC) y del texto bruto (TB), así como el porcentaje de incremento del error en el segundo (TB) con respecto al primero (TC), anotado según la tipología descrita en el marco general de categorías preestablecidas (Sección 4). Las secciones de Título y Palabras clave se excluyen del análisis en adelante debido a que no fue necesario realizar predicción en estos apartados (Tabla 6).

| Categoría de error | TC | TB | % Incremento |
|---------------------------|------------|------------|---------------------|
| Ortografía | 63 | 128 | 103.17 |
| Gramática | 53 | 35 | -33.96 |
| Semántica | 106 | 139 | 31.13 |
| Estilo | 48 | 84 | 75.00 |
| Adición de palabras | 1 | 5 | 400.00 |
| Omisión de palabras | 25 | 21 | -16.00 |
| TOTAL | 296 | 412 | 39.19 |

Tabla 11. Número total de errores observados durante la posesición por categoría en las traducciones de TC y TB y porcentaje en el incremento de errores en TB con respecto a TC.

La posesición de la traducción del texto bruto requirió de un mayor número de ediciones que la del texto controlado (412 vs. 296), en concreto un 39.19 % más en el conjunto del texto. La traducción del texto bruto necesitó de un número de ediciones ortotipográficas sustancialmente más elevado (63 vs. 128; incremento del 103.17 %). La categoría correspondiente a la ortografía incluye algunas subcategorías del error (Tabla 10) que los motores de traducción automática son susceptibles de incluir de manera indiscriminada y arbitraria, como la adición innecesaria de espacios junto a los símbolos y el uso de mayúsculas que no proceden. Esta cuestión se analiza pormenorizadamente más adelante en esta misma sección (consultar Figura 2). Al optar por excluir los errores ortotipográficos del análisis, el volumen de errores resulta más equilibrado entre el texto bruto y el texto controlado (233 vs. 284) y el incremento porcentual en el número de cambios realizados en el conjunto del texto se reduce al 21.89 % (Tabla 12).

| Categoría de error | TC | TB | % Incremento |
|---------------------------|------------|------------|---------------------|
| Gramática | 53 | 35 | -33.96 |
| Semántica | 106 | 139 | 31.13 |
| Estilo | 48 | 84 | 75.00 |
| Adición de palabras | 1 | 5 | 400.00 |
| Omisión de palabras | 25 | 21 | -16.00 |
| TOTAL | 233 | 284 | 21.89 |

Tabla 12. Número total de errores observados durante la posesición por categoría (a excepción de los errores ortotipográficos) en las traducciones de TC y TB y porcentaje en el incremento de errores en TB con respecto a TC.

Existen diferencias notables en la cantidad total de errores pertenecientes a cada categoría general entre la posesición del texto controlado y la del texto bruto. Los errores gramaticales fueron un 33.96 % más numerosos en el texto controlado (53 vs. 35). Este resultado es sorprendente, dado que uno de los objetivos principales de la aplicación del lenguaje controlado fue precisamente simplificar la gramática del texto de partida. Esta cuestión se analiza más adelante, pero entre las posibles causas podrían encontrarse (1) que las frases reescritas durante la posesición incrementen su complejidad y por tanto dificulten la traducción automática y (2) que las reglas de lenguaje controlado hayan provocado errores de forma directa (Tabla 14). Los errores semánticos, estrechamente relacionados con la precisión léxica, son considerablemente menos abundantes en el texto controlado que en el bruto (106 vs. 139), que cuenta con un 31.13 % más de errores. El mismo patrón se observa en los errores de estilo (48 vs. 84), con un incremento del 75 % en el texto bruto. La adición y la omisión de palabras son comúnmente concebidos como fenómenos aleatorios en las traducciones automáticas y no procede analizarlos en detalle ni especular sobre sus causas.

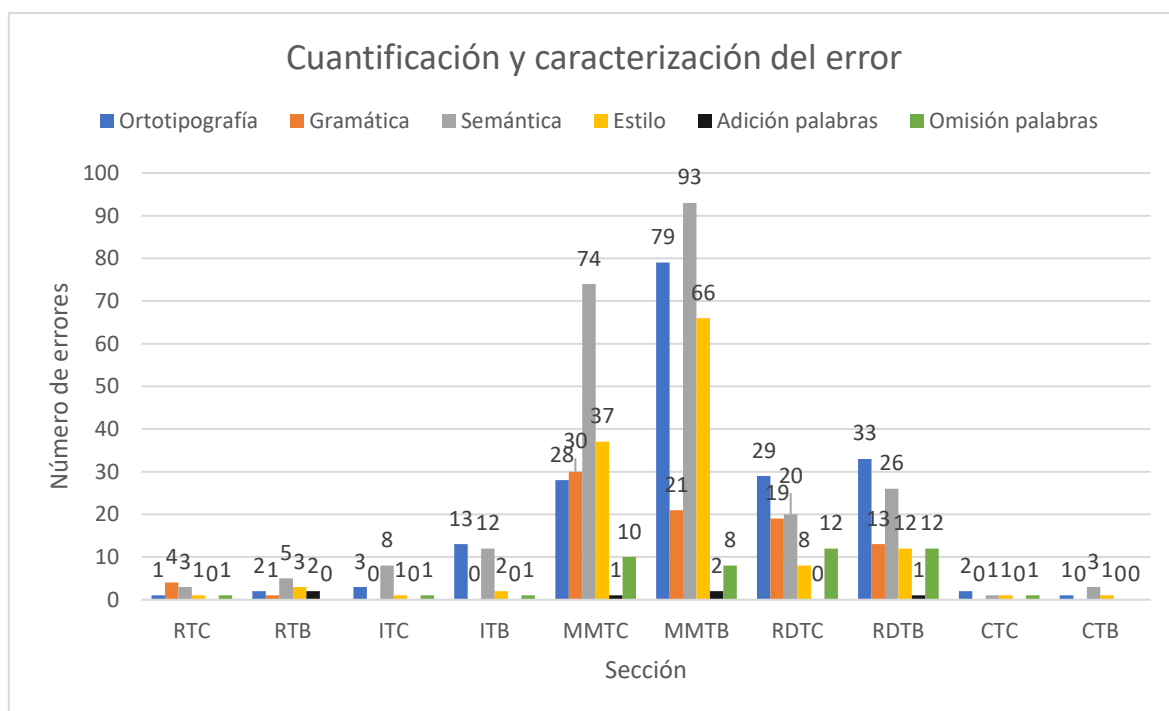


Figura 1. Número total de errores observados durante la posesición por categoría general en las traducciones de TC y de TB para cada sección (Resumen, R; Introducción, I; Materiales y Métodos, MM; Resultados y Discusión, RD, Conclusiones, C).

Con el fin de analizar la distribución del error con mayor detalle, en la Figura 1 se presenta el desglose del error para cada texto por secciones. El patrón de distribución del error en la traducción del texto bruto es equiparable a la del texto controlado en tanto que la mayor parte de los errores se concentran en las secciones más extensas, Materiales y Métodos (MM) y Resultados y Discusión (RD). Resulta notorio el desproporcionado volumen de errores semánticos en la sección de Materiales y Métodos (n=74 en TC, n=93 en TB) con respecto a otras categorías, que parecen tener rangos más proporcionados y equilibrados. Al comparar el número de errores de esta categoría en esta sección entre ambos textos, podemos deducir que el control léxico-terminológico sí parece haber tenido efectos positivos en la reducción del error (74 vs. 93), aunque el volumen sigue siendo muy elevado. Asimismo, el número de errores de estilo (37 vs. 66) y ortotipográficos se disparan en esta sección en el texto bruto con respecto a la misma sección en el texto controlado (28 vs. 79). Para analizar en profundidad esta disparidad, en los siguientes párrafos y figuras se desglosa el error según las subcategorías recogidas en la Tabla 10.

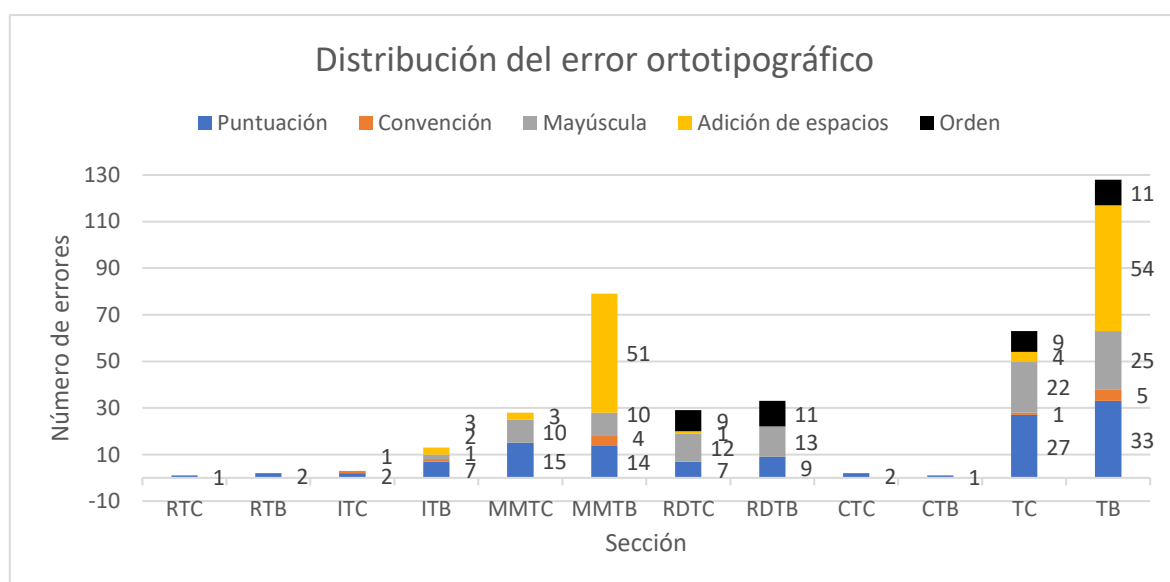


Figura 2. Número total de errores ortotipográficos observados durante la posesición por subcategoría en las traducciones de TC y de TB para cada sección (Resumen, R; Introducción, I; Materiales y Métodos, MM; Resultados y Discusión, RD, Conclusiones, C) y totales.

Los errores ortotipográficos en el texto controlado difirieron en su composición y proporciones con respecto al texto bruto (Figura 2), aunque los errores de puntuación parecen ser una constante a todas las secciones en ambos casos. Quizá la diferencia más notable entre ambos textos es la observada en la sección de Materiales y Métodos con respecto a la adición de espacios (3 vs. 51). El motor de traducción automática introdujo en el texto bruto un elevado número de espacios adicionales junto a los símbolos que describen variables, muy abundantes en este apartado.

Otra cuestión destacable con respecto a esta categoría del error es la presencia de errores de orden solo en el apartado de Resultados y Discusión. Esto se debe a que en esta sección abundan las combinaciones entre palabras y números, que aparecen con una secuencia alterada en ambos textos como resultado de la traducción automática (*médicos del hospital 1* vs. *hospital doctors 1* en lugar de *doctors of hospital 1*). La introducción de mayúsculas mantuvo una proporción casi idéntica en ambos textos en las secciones más extensas (Materiales y Métodos y Resultados y Discusión).

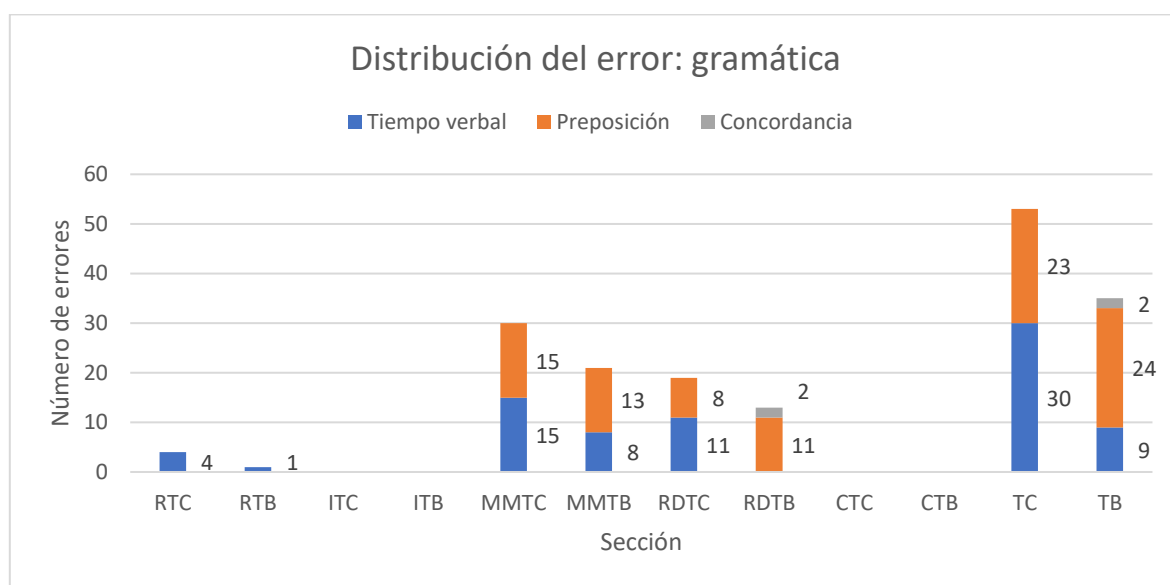


Figura 3. Número total de errores gramaticales observados durante la posesición por subcategoría en las traducciones de TC y de TB para cada sección (Resumen, R; Introducción, I; Materiales y Métodos, MM; Resultados y Discusión, RD, Conclusiones, C) y totales.

Al estudiar los errores gramaticales por subcategoría (Figura 3) se aprecia la distribución del error adelantada en la Tabla 12. Los errores de tiempo verbal fueron más abundantes en el texto controlado en las tres secciones en las que aparecieron: Resumen (4 vs. 1), Materiales y Métodos (15 vs. 8) y Resultados y Discusión (11 vs. 0). Como ya se ha mencionado, este fenómeno se analiza en detalle más adelante en esta misma sección (Tabla 14).

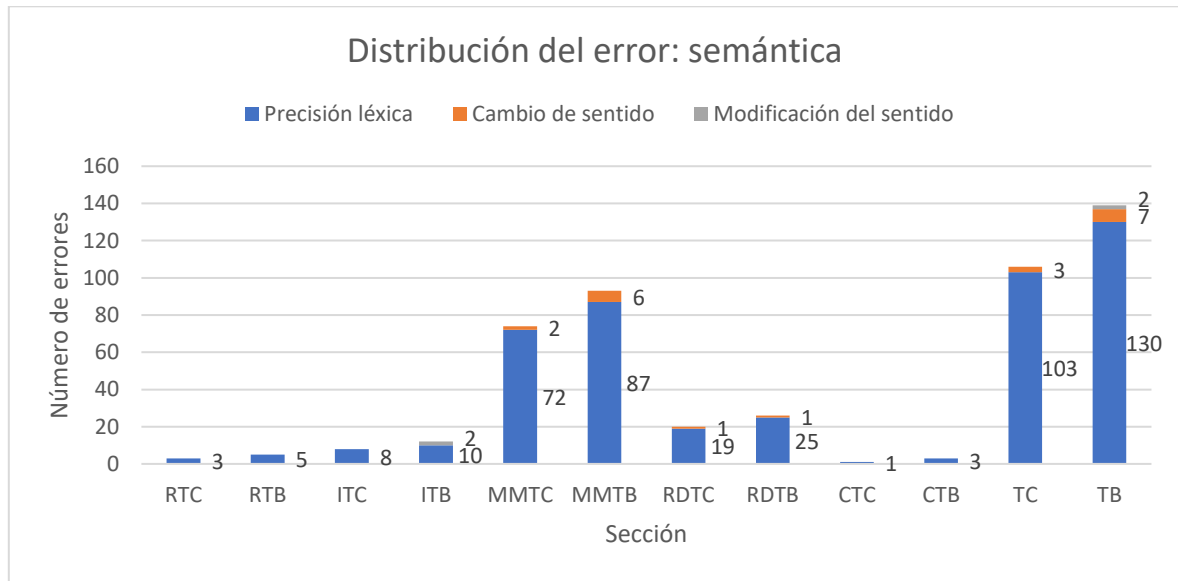


Figura 4. Número total de errores semánticos observados durante la posesición por subcategoría en las traducciones de TC y de TB para cada sección (Resumen, R; Introducción, I; Materiales y Métodos, MM; Resultados y Discusión, RD, Conclusiones, C) y totales.

La Figura 4 muestra una distribución muy similar del error semántico en los textos controlado y bruto, con un claro predominio de los problemas de precisión léxica. Estos errores son más numerosos en el texto bruto, lo que confirma que el control léxico-terminológico (CLT) tuvo un efecto positivo ($n=103$ en TC vs. $n=130$ en TB). Los problemas de cambio de sentido también experimentaron una reducción en el texto controlado en la sección de Materiales y Métodos (2 vs. 6) y los errores de modificación del sentido presentes en la Introducción del texto bruto ($n=2$) fueron prevenidos.

Por último, la Figura 5 revela que los errores de naturalidad presentaron un patrón consistente en ambos textos, con un elevado número en la sección de Materiales y Métodos. La aplicación de lenguaje controlado resultó en una reducción importante en el número de

errores de esta subcategoría, especialmente en la sección de Materiales y Métodos (37 vs. 66), pero no consiguió eliminarlos por completo. El error de consistencia solo fue detectado en una ocasión en la sección de Resultados y Discusión del texto controlado, ya que la traducción automática ofreció dos traducciones diferentes para un segmento repetido. Esto no ocurrió en el texto bruto, fruto de que la traducción automática ofrece en ocasiones traducciones arbitrarias o inconsistentes. No procede realizar un comentario en profundidad acerca de los errores de adición (n=1 en TC, n=5 en TB) y de omisión de palabras (n=25 en TC, n=21 en TB), ya que ambos parecen responder a fenómenos que no se pueden controlar y aparecen en proporciones similares en ambos textos, especialmente en el caso de las omisiones, que en los textos bajo estudio correspondieron principalmente a la omisión arbitraria del artículo definido (*the*).

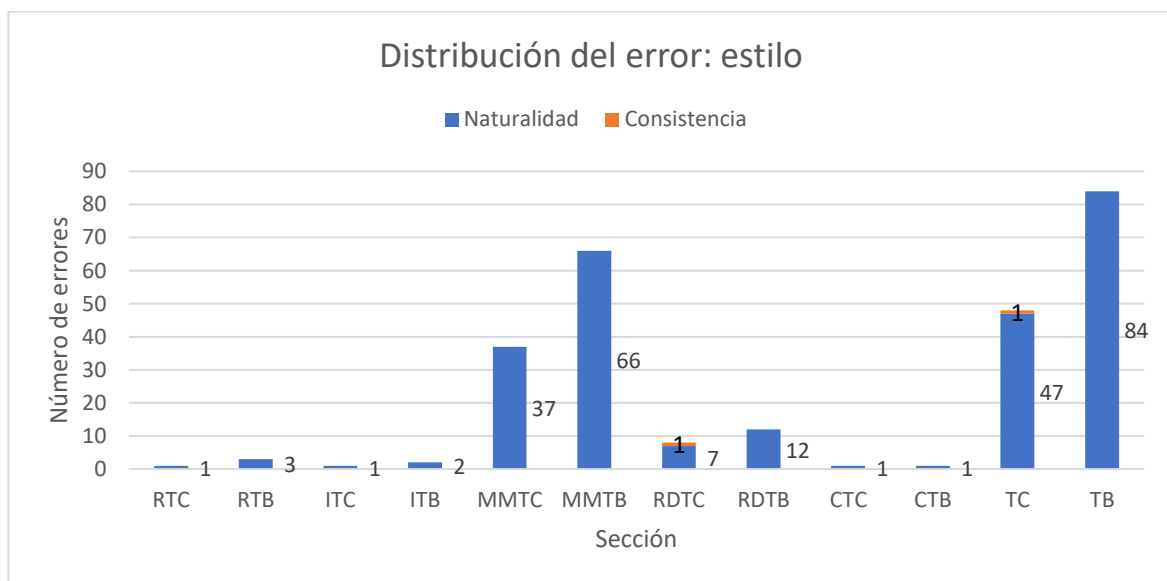


Figura 5. Número total de errores de estilo observados durante la posesición por subcategoría en las traducciones de TC y de TB para cada sección (Resumen, R; Introducción, I; Materiales y Métodos, MM; Resultados y Discusión, RD, Conclusiones, C) y totales.

Eficacia del lenguaje controlado y prevención del error

Con el objetivo de evaluar la efectividad de la preedición y del lenguaje controlado, se realizó un proceso de anotación comparativo mediante el estudio detallado del texto controlado y las traducciones automáticas de los textos controlado y bruto. La Tabla 13

muestra el número de cambios realizados en la preedición con cada regla de lenguaje controlado (CLT y CG), el número de frases con cambios de estilo esenciales (EE) para aplicar dichas reglas y el número de correcciones de error (CE), así como el número de errores prevenidos por cada uno de ellos en el texto controlado con respecto al bruto. Asimismo, incluye el porcentaje de cambios efectivos (% de cambios que lograron prevenir el error) y el porcentaje de errores prevenidos (% de errores prevenidos con respecto al número total de errores detectados en la posesición del texto bruto). Todas las variables fueron calculadas para cada tipo de cambio y para la preedición en conjunto. Se optó por excluir los cambios ortotipográficos (CO) del análisis, ya que la traducción automática no siempre es capaz de trasladar la puntuación, las convenciones y demás normas de forma adecuada al texto e idioma de destino, sino que tiende a reproducirlas de forma fiel al texto de partida y puede introducir errores de forma arbitraria, tal y como se observó y mencionó anteriormente con respecto al uso de la mayúscula o la adición de espacios (Figura 2).

| Tipo de cambio aplicado | Nro. de cambios realizados | Nro. de errores prevenidos | % de cambios efectivos | % errores prevenidos |
|---------------------------|----------------------------|----------------------------|------------------------|----------------------|
| CE | 51 | 6 | 11.76 | 2.11 |
| CLT | 52 | 26 | 50.00 | 9.15 |
| CG1 | 57 | 0 | 0.00 | 0.00 |
| CG2 | 12 | 1 | 8.33 | 0.35 |
| CG3 | 10 | 2 | 20.00 | 0.70 |
| CG4 | 12 | 4 | 33.33 | 1.41 |
| CG5 | 28 | 4 | 14.29 | 1.41 |
| CG6 | 16 | 1 | 6.25 | 0.35 |
| CG7 | 6 | 0 | 0.00 | 0.00 |
| CG8 | 0 | 0 | 0.00 | 0.00 |
| CG9 | 8 | 2 | 25.00 | 0.70 |
| CG10 | 88 | 7 | 7.95 | 2.46 |
| EE | 74 | 2 | 2.70 | 0.70 |
| Preedición (TOTAL) | 414 | 55 | 13.29 | 19.37 |

Tabla 13. Número absoluto de cambios realizados en la preedición, número absoluto de errores prevenidos en la traducción de TC respecto a la de TB, porcentaje de cambios efectivos en la prevención del error y porcentaje de errores prevenidos con respecto al número total de errores en TB.

De la totalidad de los cambios realizados durante la preedición para implantar el lenguaje controlado a nivel gramatical (CG) y léxico-terminológico (CLT), incluidos los cambios de

estilo esenciales (EE) y las correcciones de errores (CE), solo un 13.29 % de ellos resultaron efectivos a la hora de prevenir el error en el texto controlado. Sin embargo, teniendo en cuenta el número total de errores en la traducción del texto bruto (n=284, Tabla 12), con la excepción de los errores ortotipográficos excluidos en esta fase del análisis, la prevención de 55 errores (Tabla 13) supuso un porcentaje total de errores prevenidos del 19.37 %.

Tres reglas de control gramatical presentaron una capacidad nula para la prevención de errores, aunque las causas de estos resultados parecen ser sustancialmente diferentes.

- En el caso de la regla CG1 *acortar las frases*, resulta imposible determinar cuándo este cambio conduce a la prevención de errores o si la mejora de la traducibilidad se debe a los cambios de estilo necesarios (EE) para reducir el número de palabras (reescritura) y no a la longitud de la frase en sí. Esto es especialmente cierto si en la misma frase se han aplicado otras reglas. Por ello, no tiene sentido tratar de cuantificar el número de errores por frase para determinar la eficacia de la regla y se optó por atribuir un error o prevención de error a este cambio cuando claramente la traducción automática no pudo seguir la secuencia de la frase debido exclusivamente a su longitud. Como consecuencia, no se detectó ningún error prevenido.
- La regla CG8 *evitar la doble negación* no fue empleada en ninguna ocasión.
- La regla CG7 *uso de tiempos verbales simples*, aunque se empleó en un total de 6 ocasiones, no logró prevenir ningún error.

La regla que logró evitar un mayor porcentaje de errores fue CG10 *evitar la voz pasiva*, que previno el 2.46 % de los errores presentes en el texto bruto. Sin embargo, teniendo en cuenta el elevado número de veces en que se aplicó (n=88), el porcentaje de cambios útiles fue relativamente bajo (7.95 % de cambios efectivos). Al considerar ambas variables en conjunto, otras reglas resultaron mucho más eficientes, tales como CG4 *evitar elipsis* y CG5 *evitar anáfora*, que lograron prevenir el 1.41 % de los errores cada una, mientras que el número de cambios efectivos ascendió al 33.33 % y al 14.29 %, respectivamente. Les siguen CG9 *evitar gerundios en estructuras anglicadas* con una prevención del error del 0.7 % y un porcentaje de cambios eficaces del 25 % y CG3 *solo una subordinada por frase* con el mismo porcentaje de errores prevenidos y un 20 % de cambios eficaces.

Por su parte, el control léxico-terminológico (CLT) resultó en una prevención del error del 9.15 % y un 50 % de cambios efectivos, por lo que compensa de forma muy positiva la baja eficacia relativa de las reglas de control gramatical (CG) a nivel individual, aunque en conjunto, el control gramatical logró evitar el 7.39 % de los errores (Tabla 13). Por último, cabe mencionar que los cambios de estilo esenciales (EE) y la corrección del error (CE) también contribuyeron positivamente a la preedición, logrando prevenir el 2.11 % y el 0.7 % de los errores, respectivamente.

Durante la anotación comparativa también se detectaron errores en la traducción del texto controlado que no estaban presentes en la traducción del texto bruto. Estos errores surgieron como consecuencia de la introducción del lenguaje controlado, tal y como queda reflejado en la Tabla 14, elaborada forma pareja a la Tabla 13, pero teniendo en cuenta el número de errores ocasionados con respecto al número total de errores presentes en el texto controlado (n=233) en lugar de los errores prevenidos con respecto al texto bruto.

| Tipo de cambio aplicado | Nro. de cambios implementados | Nro. de errores generados | % cambios generadores de error | % errores generados |
|---------------------------|-------------------------------|---------------------------|--------------------------------|---------------------|
| CE | 51 | 0 | 0.00 | 0.00 |
| CLT | 52 | 0 | 0.00 | 0.00 |
| CG1 | 57 | 0 | 0.00 | 0.00 |
| CG2 | 12 | 0 | 0.00 | 0.00 |
| CG3 | 10 | 0 | 0.00 | 0.00 |
| CG4 | 12 | 0 | 0.00 | 0.00 |
| CG5 | 28 | 1 | 3.67 | 0.43 |
| CG6 | 16 | 0 | 0.00 | 0.00 |
| CG7 | 6 | 0 | 0.00 | 0.00 |
| CG8 | 0 | 0 | 0.00 | 0.00 |
| CG9 | 8 | 0 | 0.00 | 0.00 |
| CG10 | 88 | 20 | 22.72 | 8.58 |
| EE | 74 | 1 | 1.35 | 0.43 |
| Preedición (TOTAL) | 414 | 23 | 5.56 | 9.87 |

Tabla 14. Número absoluto de cambios realizados en la preedición, número absoluto de errores generados en la traducción de TC respecto a la de TB, porcentaje de cambios generadores de error y porcentaje de errores generados con respecto al número total de errores en TC.

Aunque la preedición en conjunto (excluida la corrección ortotipográfica) generó el 9.87 % de los errores en el texto controlado, la mayoría de los tipos de cambio no tuvieron este efecto. Dos categorías generaron un número residual de errores (0.43 %): CG5 *evitar anáfora* y los cambios esenciales de estilo (EE). A pesar de ello, la capacidad de la preedición para evitar el error fue superior a la generación de errores (19.37 % vs. 9.87 %, Tabla 13).

Tal como se especulaba anteriormente (Tabla 12 y Figura 3), es especialmente notorio el efecto generador de errores de la regla CG10 *evitar la voz pasiva*, que ocasionó un 8.58 % de los errores presentes en el texto controlado. Al estudiar el efecto de esta regla en su conjunto se observa que su contribución a la prevención del error fue muy inferior a su capacidad generadora de errores (2.46 % vs. 8.58 %). Como consecuencia de reescribir las frases en la voz activa, se introdujeron errores en la traducción automática en aquellas instancias en las que los verbos convergen en la forma verbal de la primera persona del plural del pretérito indefinido y del presente simple (cuando el sujeto son los propios autores). La Tabla 15 recoge el listado exhaustivo de los verbos problemáticos detectados, así como el número de errores introducidos en cada caso. Dado que el número total de errores de tiempo verbal fue de 30 en el texto controlado en conjunto (Figura 3), los errores introducidos como consecuencia de la aplicación de la regla CG10 (n=20) representaron el 66.67 % de los errores de este tipo.

| Verbo | Nro. errores en TC |
|-----------------------------|--------------------|
| Asignar (asignamos) | 9 |
| Considerar (consideramos) | 2 |
| Definir (definimos) | 4 |
| Desarrollar (desarrollamos) | 1 |
| Ilustrar (ilustramos) | 1 |
| Puntuar (puntuamos) | 1 |
| Sugerir (sugerimos) | 1 |
| Verificar (verificamos) | 1 |
| TOTAL | 20 |

Tabla 15. Listado exhaustivo de los verbos problemáticos para la aplicación de la regla CG10 y número de errores introducidos como consecuencia en TC.

La Figura 6 resulta útil para visualizar qué categorías de error lograron ser prevenidas por cada tipo de cambio durante la preedición. Como se puede apreciar, el control léxico-terminológico solo tuvo efecto sobre la precisión léxica (n=26). El resto de las reglas de lenguaje controlado que demostraron cierta eficacia afectaron a errores de máximo dos tipologías de error. La corrección de errores (CE) y los cambios de estilo esenciales (EE) tuvieron un efecto más diverso y lograron prevenir errores en más de dos subcategorías, con los errores de naturalidad, modificación del sentido y precisión léxica como denominadores comunes. CG4 *evitar la elipsis* fue la regla que resultó más eficaz para evitar modificaciones en el sentido (n=3). Las subcategorías de error que lograron ser evitadas en mayor medida fueron los errores de precisión léxica (n=31) y los problemas de naturalidad (n=22).

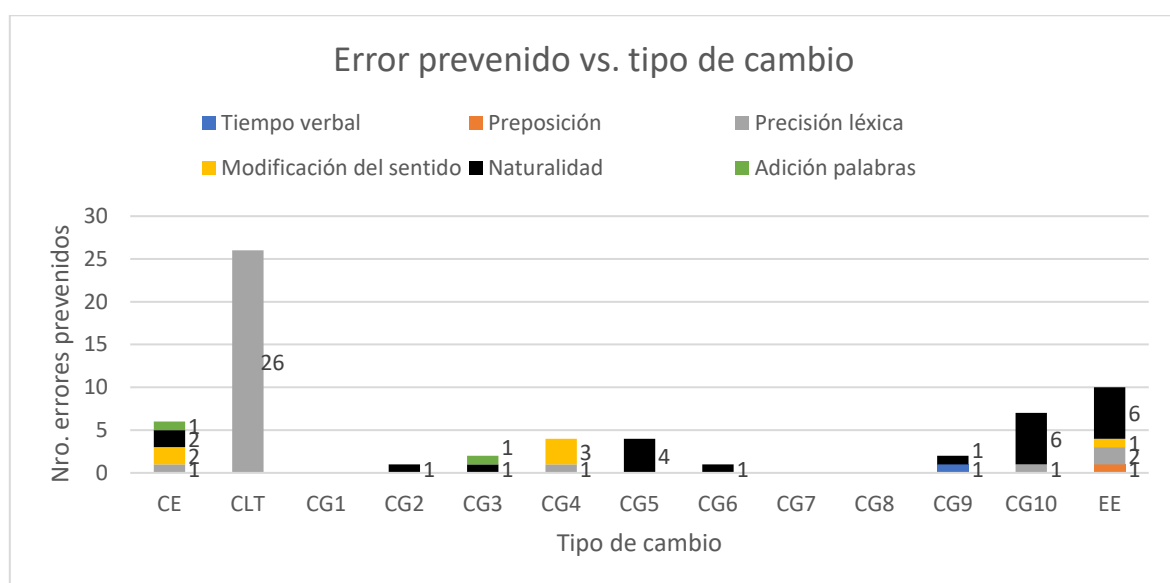


Figura 6. Número de errores prevenidos por subcategoría de error y tipo de cambio aplicado.

Léxico problemático: ambigüedad, variedades dialectales y anglicismos

Con el fin de arrojar algo de luz acerca de los problemas de traducción automática ocasionados por el léxico, se realizó una revisión final de la anotación de los textos preeditados y poseditados para elaborar un listado de palabras causantes de error que no podrían haber sido ni previstos ni solventados en origen en ausencia de herramientas informáticas específicas de control léxico-terminológico. La Tabla 16 recoge un listado de palabras problemáticas que se estimaron como irresolubles en origen, debido a una cuestión

de polisemia que no se puede resolver empleando un sinónimo en el contexto en el que se emplean. También se incluye el número de errores ocasionados tanto en las traducciones del texto bruto y del texto controlado, así como los errores totales causados.

| Palabra | Nro. errores TB | Nro. errores TC | Traducción errónea | Traducción correcta |
|--------------|-----------------|-----------------|-----------------------|---------------------|
| bajo | 6 | 6 | short | low |
| medio/media | 7 | 7 | medium, half, average | medium |
| muestra | 1 | 1 | specimen | sample |
| rango | 3 | 3 | rank | range |
| seguridad | 2 | 2 | safety | security |
| TOTAL | 19 | 19 | | |

Tabla 16. Listado de léxico problemático por ambigüedad irresoluble en origen debido a la polisemia con sus traducciones erróneas y correctas y el número de errores ocasionados en TB y TC por cada palabra y valores totales.

Como se puede observar, el número de errores causados por léxico cuya ambigüedad resultó irresoluble en origen ascendió a 19 errores tanto en el texto bruto como en el controlado. Teniendo en cuenta que el número de errores de precisión léxica en ambos tipos de texto fue de 130 y 103 errores, respectivamente (Figura 4), el número de errores de precisión léxica irresolubles en origen correspondió por tanto al 14.62 % y al 18.45 %.

En la Tabla 17 se incluye un listado de palabras aparentemente no problemáticas en español si se tiene en cuenta el contexto del discurso en el que se producen y que el motor de traducción automática trasladó al inglés de una forma errónea, en ocasiones modificando la categoría gramatical (*analyses* vs. *analyzes*; *nursing home* vs. *geriatric*; *distancing* vs. *detach*), mientras que en otras simplemente eligió una acepción totalmente descontextualizada (*argentine* vs. *Argentinean*; *test* vs. *evidence*). También se incluye el número de errores ocasionados tanto en la traducción del texto bruto como del texto controlado, así como los errores totales.

| Palabra | Nro. errores TB | Nro. errores TC | Traducción errónea | Traducción correcta |
|----------------------|-----------------|-----------------|--------------------|--------------------------|
| análisis (s. pl.) | 3 | 6 | analyzes (v.) | analysis/analyses |
| argentino (adj.) | 2 | 2 | argentine (adj.) | argentinean/of Argentina |
| asignar (v.) | 2 | 0 | award (v.) | assign |
| distanciamiento (s.) | 1 | 1 | detach (v.) | distancing |
| geriátrico (s.) | 1 | 5 | geriatric (adj.) | nursing home |
| prueba (s.) | 1 | 0 | evidence (s.) | test |
| TOTAL | 10 | 14 | | |

Tabla 17. Listado de léxico problemático por ambigüedad imprevisible en origen con sus traducciones erróneas y correctas y el número de errores de precisión léxica ocasionados en TB y TC por cada palabra y valores totales.

El número de errores generados como consecuencia de la elección descontextualizada de palabras fue dispar entre ambos tipos de texto, al contrario que en el caso de los errores debidos a problemas de polisemia irresoluble (Tabla 16). El texto bruto presentó 10 errores y el texto controlado 14 errores, que representan, respectivamente, el 7.69 % y el 13.59 % de los errores de precisión léxica con respecto al total para cada uno de los textos (Tabla 17; 130 vs. 103; Figura 4).

Tal y como se detalla en la Sección 4, el artículo científico objeto de estudio está redactado en la variedad dialectal del español en Argentina, que se procuró respetar durante el proceso de predicción. Sin embargo, solo se detectaron dos palabras que claramente se pueden adscribir a ella: *rubro* e *insumo*. La primera se trasladó al inglés de una forma imprecisa generando 1 error en ambos textos (*area* vs. *sector*) y la segunda dio lugar a 4 errores en el texto bruto. La variedad dialectal puede afectar a cómo se trasladan las palabras al idioma de destino (*insumo* vs. *supplies* o *input* en lugar de *resources*) y a la acepción que se les otorga en origen. En el artículo, *insumo* se utilizó con una acepción errónea (*producto* en lugar de *recurso*) que se corrigió en el texto controlado. Los errores derivados del uso de palabras pertenecientes a la variedad dialectal representaron el 3.85 % de los errores en el texto bruto y el 0.97 % en el controlado (5 vs. 130 y 1 vs. 103, respectivamente; Figura 4).

Por último, en el texto se detectaron dos anglicismos léxicos frecuentes: *test* (y *testear*) y *pool* que los autores emplearon con acepciones inconsistentes. Por ejemplo, *test* se empleó indistintamente como *prueba*, *análisis*, *analizar* y *muestrear*, mientras que *pool* se utilizó tanto para designar *grupos* de personas o de muestras, como muestras *compuestas* o muestras

grupales. La Tabla 18 recoge el número de ocasiones en las que se emplearon estos anglicismos, el número de ocasiones en que se corrigieron durante la preedición y el número de errores de precisión léxica asociados tanto en el texto bruto como controlado.

| <i>Test</i> | Sección | Nro. ocasiones | Nro. correcciones | Nro. errores TB | Nro. errores TC |
|-------------|--------------|----------------|-------------------|-----------------|-----------------|
| | I | 1 | 1 | 0 | 0 |
| | MM | 7 | 7 | 3 | 2 |
| | RD | 1 | 1 | 0 | 0 |
| | TOTAL | 9 | 9 | 3 | 2 |
| <i>Pool</i> | Sección | Nro. ocasiones | Nro. correcciones | Nro. errores TB | Nro. errores TC |
| | I | 0 | 0 | 0 | 0 |
| | MM | 10 | 10 | 4 | 5 |
| | RD | 1 | 1 | 0 | 0 |
| | TOTAL | 11 | 11 | 4 | 5 |

Tabla 18. Número de ocasiones en que aparecen los anglicismos léxicos, número de correcciones realizadas y número de errores de precisión léxica asociados en TB y TC.

Como se puede apreciar, el motor de traducción automática no parece haber respondido de forma consistente ante los anglicismos. En algunas ocasiones la presencia del anglicismo tuvo el mismo efecto que su corrección y no generó error (*test* en la Introducción y los Resultados y Discusión; *pool* en Resultados y Discusión), mientras que en otras se dieron más errores en el texto bruto (*test*) o en el texto controlado (*pool*). En el caso de *test* (y *testear*), los dos errores registrados correspondieron a la corrección de *test* por *análisis*, que se tradujo automáticamente como *analyzes* (Tabla 17). En el caso de *pool*, la totalidad de los errores generados en el texto controlado correspondieron a instancias en las que se había eliminado el anglicismo y se revirtió a *pool* en la posesición por cuestiones estilísticas. En conjunto, los errores de precisión léxica relacionados con la presencia de estos anglicismos representaron el 5.38 % de los errores de este tipo en el texto bruto y el 6.8 % en el texto controlado (7 vs. 130 y 7 vs. 103, respectivamente; Figura 4).

6. DISCUSIÓN

El propósito del presente estudio fue evaluar la utilidad de la aplicación conjunta de la preedición con lenguaje controlado, la traducción automática y la posesición para la traducción al inglés de artículos científicos en el campo de la medicina redactados en español. En definitiva, el objetivo fue evaluar si este flujo de trabajo permite mejorar la traducibilidad automática al inglés sin comprometer la calidad del texto en origen y obtener resultados publicables con un menor esfuerzo de posesición. Aunque los resultados de este trabajo no se pueden extrapolar ni generalizar dado que el estudio se realizó sobre un único artículo científico publicado, con unas características muy concretas (baja densidad terminológica, lenguaje claro y llano), restringido a un único ámbito de la medicina (epidemiología) y con una variedad dialectal específica (español de Argentina), resulta una buena primera aproximación al estudio de la mejora de la traducibilidad automática al inglés de este tipo de textos.

La propuesta de esta investigación incluyó la aplicación de lenguaje controlado a dos niveles: el control léxico-terminológico (CLT) y el control gramatical (CG), plasmado en un formato de 10 reglas de redacción a modo de decálogo (Tabla 5). Asimismo, se aprovechó el proceso de preedición para corregir errores presentes en el texto fuente, tanto de índole general, como ortotipográficos. Como resultado de la preedición, se realizaron 289 cambios en el conjunto del texto exclusivamente destinados a la aplicación del lenguaje controlado, de los cuales 52 correspondieron al control léxico-terminológico (CLT) y 237 al control gramatical (CG) (Tabla 6). Además, fue necesario reescribir frases en 74 ocasiones mediante cambios de estilo esenciales (EE) y realizar una corrección del error (CE) y una corrección ortotipográfica (CO), que representaron 51 y 29 cambios respectivamente. El resultado final de la preedición ascendió a 443 cambios anotados (Tablas 8 y 9).

La regla más ampliamente aplicada fue CG10 *evitar la voz pasiva*, aunque esta se empleó de forma dispar según la sección del artículo, con una mayor frecuencia en las secciones más extensas, Materiales y Métodos y Resultados y Discusión. Estos son los apartados más explícitos, detallados y expositivos del artículo científico y en ellos se describe paso a paso cómo se ha realizado un estudio y las implicaciones de los resultados obtenidos, por lo que

son susceptibles de incluir numerosas frases en las que el sujeto son los propios autores (Day, 2005). En el artículo estudiado, los autores emplean, prácticamente de forma abusiva, la pasiva refleja para dotar al texto de impersonalidad y ocultar el sujeto. La siguiente regla más ampliamente utilizada fue CG1 *acortar las frases*, cuyo uso fue notable en las secciones de Introducción y Conclusiones. Es posible que esto sea consecuencia asimismo del contenido y estilo típicos de estos apartados, que requieren de cierto grado de argumentación por parte de los autores (Gil-Alberdi González, 2009; Villagrán y Harris, 2009) y tienden por ello a redactarse en frases más elaboradas y largas. El control léxico-terminológico (CLT) fue el tercer tipo de cambio más empleado, con especial relevancia para la sección de Materiales y Métodos, que suele ser la sección con mayor densidad terminológica y más repetitiva a nivel léxico. En contraposición, la regla CG8 *evitar la doble negación* no fue necesaria en ninguna instancia (Tabla 7). La evaluación del esfuerzo técnico de preedición reveló, por tanto, que cada sección del artículo científico requirió un control del lenguaje ligeramente diferente en términos cualitativos y cuantitativos, presumiblemente como consecuencia de las diferencias inherentes a la función y estilo de cada apartado, que podrían considerarse prácticamente como subgéneros dentro del género textual del artículo científico (Tablas 3 y 4).

Tras proceder a la traducción de los textos bruto y controlado mediante Google Translate, un motor de traducción automática basado en redes neuronales libre, gratuito y ampliamente accesible, se realizó la posesición de ambas traducciones y se anotó la tipología del error conforme a un marco general de 6 categorías: ortotipografía (OT), gramática (G), semántica (S), estilo (E), adición de palabras (A) y omisión de palabras (O) (Sección 4). Durante el proceso de anotación, se optó por describir el tipo de error observado para después agruparlos y cuantificarlos (Tabla 10). Aunque esta metodología resultó muy práctica para los fines de este estudio por su exhaustividad y especificidad, por razones obvias, esta clasificación no es extrapolable a otros textos en los que pueden aparecer otros tipos de error diferentes. Aparte de identificar la tipología del error, la anotación de las posesiciones permitió cuantificar cada tipo de error y evaluar su distribución a nivel de sección dentro del artículo estudiado.

El motor de traducción Google Translate produjo resultados de elevada calidad para ambos textos, tal y como ya observaron Rossetti y O'Brien en su estudio (2019). La posesición del texto bruto requirió de un 39.19 % de cambios más para subsanar errores que la posesición del texto controlado, lo que pone de manifiesto que el control del lenguaje y el resto de las correcciones aplicadas durante la preedición tuvieron un efecto ciertamente positivo (Tabla 11). No obstante, al excluir del análisis los errores ortotipográficos, que se introdujeron en gran número y de forma arbitraria en una sección del texto bruto (Materiales y Métodos), estas diferencias en el esfuerzo técnico de posesición se redujeron al 21.89 % (Tabla 12). En términos generales, el texto controlado experimentó una reducción de errores en la mayoría de las categorías del error. Sin embargo, los errores gramaticales se elevaron en un 33.96 % en el texto controlado. Este sorprendente resultado evidencia que el lenguaje controlado también es susceptible de introducir errores, bien de forma directa, o como consecuencia de modificaciones en la complejidad de las frases. Por su parte, los errores semánticos fueron un 31.13 % más elevados en el texto bruto y los errores de estilo, un 75 %. Estos resultados sí eran esperables, ya que la preedición permitió solventar una parte importante de la imprecisión léxica gracias al control léxico-terminológico (CLT) al reducir la ambigüedad y atajar la polisemia. Por otro lado, los motores de traducción automática tienden a producir textos que, si bien son correctos gramaticalmente y precisos a nivel semántico, suenan poco naturales (problema de estilo). La reescritura durante la preedición probablemente afectó de forma positiva a la naturalidad del texto traducido.

El grueso de los errores se concentró en las secciones más extensas (Materiales y Métodos y Resultados y Discusión) para ambos textos, tal y como cabía esperar (Figura 1). El número de errores semánticos en la sección de Materiales y Métodos resultó notorio para ambos textos. La imprecisión léxica es el tipo de error que contribuye en mayor medida a los problemas semánticos y el texto de esta sección, aparte de incluir una mayor densidad terminológica, también experimentó un elevado número de errores de precisión léxica inevitables (Tabla 16). Los errores de estilo y ortotipográficos también presentaron una distribución muy dispar.

Al desglosar el error por subcategorías y por sección se observó que, si bien el volumen de errores de puntuación y de introducción de mayúsculas se mantuvo relativamente parejo en

ambos textos, el motor de traducción automática introdujo un gran volumen de espacios adicionales en la sección de Materiales y Métodos en el texto bruto, especialmente asociados a los símbolos, abundantes en esta sección. Por otro lado, en la sección de Resultados y Discusión se dio un tipo de error que no apareció en ninguna otra sección, el de orden, asociado a la combinación de palabras con números o símbolos (Figura 2).

Como ya se ha comentado, los resultados sobre el error gramatical arrojaron unos resultados inesperados, ya que los errores de tiempo verbal fueron más abundantes en el texto controlado en todas las secciones en las que aparecieron (Figura 3). El volumen y distribución de los errores con respecto al uso de las preposiciones fue similar en ambos textos (Figura 3), lo que sugiere que existen problemas inevitables con respecto a la polisemia preposicional que el motor de traducción no consigue solventar de forma eficaz. Los errores semánticos se encontraron dominados por los problemas de precisión léxica que, si bien fueron más numerosos en el texto bruto, lo que demuestra que el control léxico-terminológico (CLT) tuvo efectos positivos, este resultó insuficiente para eliminarlos por completo (Figura 4). Los errores de estilo, en concreto, de naturalidad, presentaron un patrón consistente en ambos textos y, aunque reducidos en el texto controlado, el control del lenguaje tampoco consiguió prevenirlos suficientemente (Figura 5). Por último, cabe destacar que, aunque la adición y la omisión de palabras puedan responder a fenómenos arbitrarios, sí se observó un problema recurrente con la omisión, aparentemente aleatoria, del artículo definido (*the*).

Un tercer proceso de anotación comparativo frase por frase entre el texto controlado y las posesiones del texto controlado y del texto bruto permitió realizar un estudio detallado de la eficacia del lenguaje controlado en la prevención del error. De todos los cambios para implantar el lenguaje controlado a nivel gramatical (CG y EE) y léxico-terminológico (CLT), así como la corrección del error (CE), tan solo un 13.29 % fueron efectivos para prevenir el error, logrando evitar el 19.37 % de los errores. Tres reglas no resultaron de utilidad en la prevención del error, bien porque su eficacia no se pudo evaluar (CG1 *acortar las frases*), porque no se aplicaron en ninguna instancia (CG8 *evitar la doble negación*) o porque no lograron prevenir ningún error (CG7 *uso de tiempos verbales simples*). Estos resultados contrastan con las recomendaciones realizadas por O'Brien (2003). La regla que logró evitar

más errores fue CG10 *evitar la voz pasiva*, con una prevención del 2.46 % de los errores totales, aunque se aplicó en tantas ocasiones, que el porcentaje de cambios efectivos fue relativamente bajo (7.95 %). En contraposición, al considerar ambas variables en conjunto, otras reglas resultaron más eficientes. Tal es el caso de CG4 *evitar elipsis* y CG5 *evitar anáfora*, que lograron prevenir el 1.41 % de los errores totales cada una, mientras que el número de cambios efectivos fue del 33.33 % y del 14.29 %, respectivamente. CG9 *evitar gerundios en estructuras anglicadas* y CG3 *solo una subordinada por frase* presentaron una prevención del error del 0.7 % y un porcentaje de cambios eficaces del 25 % y de 20 %, respectivamente. Estos resultados concuerdan con las recomendaciones y observaciones de O'Brien (2004, 2003) y de Ramírez Polo (2012). El control léxico-terminológico (CLT) resultó en una prevención del error del 9.15 % y un 50 % de cambios efectivos, por lo que compensó de forma muy positiva la baja eficacia relativa de las reglas de control gramatical (CG) a nivel individual, aunque en conjunto, el control gramatical logró evitar el 7.39 % de los errores. Los cambios de estilo esenciales (EE) y la corrección del error (CE) también contribuyeron positivamente a la eficacia de la preedición, con una prevención de errores del 2.11 % y del 0.7 %, respectivamente (Tabla 13).

Gracias a la anotación comparativa, también se pudo estudiar la introducción de errores como consecuencia de la aplicación del lenguaje controlado. La cuestión más destacable fue la introducción de errores de tiempo verbal debidos a la aplicación de la regla CG10 *evitar la voz pasiva*. Al estudiar el efecto de esta regla en su conjunto se evidenció que su contribución a la prevención del error fue muy inferior a su capacidad generadora de errores (2.46 % vs. 8.58 %) (Tablas 13 y 14). Como consecuencia de reescribir las frases en la voz activa, se introdujeron errores en la traducción automática en aquellas instancias en las que los verbos convergen en la forma verbal de la primera persona del plural del pretérito indefinido y del presente simple (cuando el sujeto son los propios autores). Los errores introducidos a causa de la aplicación de esta regla representaron el 66.67 % de los errores de tiempo verbal observados en el texto controlado. Así, verbos como *asignar*, *considerar* o *definir*, entre otros, resultaron problemáticos para la aplicación de esta regla (Tabla 15). A pesar de esto, la capacidad de la preedición en su conjunto para evitar el error fue muy superior a la generación de errores (19.37 % vs. 9.87 %; Tablas 13 y 14).

En cuanto al tipo de error prevenido por cada tipo de cambio en la preedición, el control léxico-terminológico solo afectó a la precisión léxica, como cabía esperar. El resto de las reglas que tuvieron alguna eficacia en la prevención de errores tuvieron una respuesta inconsistente y afectaron a dos subcategorías variables según la regla aplicada. La corrección de errores (CE) y los cambios de estilo esenciales (EE) tuvieron un efecto más diverso y lograron prevenir errores en más de dos subcategorías diferentes, con los errores de naturalidad, modificación del sentido y precisión léxica como denominadores comunes. Las categorías del error que lograron ser evitadas en mayor medida fueron los errores de precisión léxica (semántica) y los problemas de naturalidad (estilo) (Figura 6).

Durante la anotación comparativa también se estudió en profundidad el léxico, lo que permitió detectar 4 cuestiones de índole léxica que resultaron problemáticas para la traducción automática: (1) problemas de ambigüedad irresoluble previsible en origen; (2) problemas de ambigüedad imprevisible, (3) la variedad dialectal y (4) los anglicismos. Existen palabras cuyo carácter polisémico genera ambigüedad en origen, pero esta no se puede resolver ya que no existen alternativas debido a una asimetría polisemia (ES) – monosemia (EN). Tal sería el caso de *media*, que se traduce en el texto como *medium*, *half* y *average* (Aleixandre Benavent y Amador Iscla, 2001a). Este tipo de léxico problemático representó el 14.62 % y el 18.45 % de los errores de precisión léxica en el texto bruto y en el texto controlado, respectivamente (Tabla 16). Asimismo, se recopiló una serie de palabras aparentemente no problemáticas en español que sí resultaron serlo para la traducción automática de forma imprevisible para la autora durante la preedición. El motor las trasladó al inglés bien modificando la categoría gramatical, como en el caso de *análisis* (s. pl.) vs. *analyzes* (v.), mientras que en otras ocasiones simplemente eligió una acepción totalmente descontextualizada, como *argentino*, que fue traducido como *argentine* (adj.) en lugar de *Argentinean* (adj.). Estos errores representaron el 7.69 % y el 13.59 % de los errores de precisión léxica en el texto bruto y en el texto controlado, respectivamente (Tabla 17). Por su parte, el léxico asociado a la variedad dialectal en que estaba redactado el artículo también resultó problemático para la traducción, a pesar de que solo se detectaron dos palabras claramente adscritas al español de Argentina, *rubro* e *insumo*. Si bien la primera simplemente se trasladó al inglés de una forma imprecisa (*area* vs. *sector*), en la segunda la

variedad dialectal afectó tanto a cómo se trasladó al inglés (*insumo* vs. *supplies* o *input* en lugar de *resources*; problema en destino), como a la acepción de la palabra en el texto de partida (*producto* en lugar de *recurso*; problema en origen). Los errores derivados del uso de palabras pertenecientes a la variedad dialectal representaron el 3.85 % de los errores en el texto bruto y el 0.97 % en el controlado. Por último, en el texto también se detectaron anglicismos léxicos, como *test* (y *testear*) y *pool*, que fueron empleados por los autores con acepciones variadas y que la traducción automática tampoco trasladó al inglés de forma consistente. En conjunto, el 5.38 % de los errores de precisión léxica en el texto bruto y el 6.8 % en el texto controlado estuvieron relacionados con estos anglicismos (Tabla 18).

Los resultados obtenidos confirman la hipótesis inicial del estudio (la aplicación del lenguaje controlado reduce el esfuerzo de posesición), aunque quizá no en la medida esperada por parte de la autora. No obstante, dado que no existen estudios similares en los que el lenguaje controlado haya sido aplicado al español en este ámbito, resulta imposible realizar una valoración comparativa.

En conjunto, el proceso de preedición resultó útil para eliminar errores, mejorar el estilo y reducir la ambigüedad léxica en el texto fuente. Por su parte, aunque la aplicación del lenguaje controlado tuvo efectos positivos, la mejora de la traducibilidad automática (medida indirectamente a través del esfuerzo de posesición) fue inferior a la esperada en el texto controlado con respecto al texto bruto, especialmente en lo que concierne al control gramatical. Esto podría deberse a varias causas: (1) no es posible evaluar directamente la efectividad de todas las reglas (CG1 *acortar las frases*); (2) algunas reglas pueden introducir un gran volumen de errores (CG10 *evitar la voz pasiva*); (3) algunas reglas fueron innecesarias o tuvieron una utilidad residual (CG8 *evitar la doble negación* o CG7 *uso de tiempos verbales simples*, respectivamente); (4) la traducción automática genera errores inevitables, imprevisibles en origen y arbitrarios de variada índole, como la adición de espacios, las omisiones del artículo definido en inglés, o los errores como consecuencia de la polisemia preposicional, aparte del clásico escollo de la traducción automática: la falta de naturalidad en el lenguaje. Por el contrario, el control léxico-terminológico sí tuvo una efectividad notable, aunque esta podría haber sido mayor de haber contado con herramientas informáticas de apoyo para poder detectar el léxico problemático en origen y prever todas

sus posibles traducciones durante la preedición. Sin embargo, esto tampoco permitiría eliminar al completo el error de imprecisión semántica, dado que también existe léxico imposible de desambiguar en origen aun cuando se detecta. Por último, la traducción automática no fue capaz de emplear la puntuación, las convenciones y demás normas de forma adecuada al texto y al idioma de destino, sino que las reprodujo con cierta fidelidad desde el texto fuente y además introdujo errores arbitrarios (adición de espacios y mayúsculas). Como consecuencia de todo ello, el proceso de posesición resultó ciertamente indispensable, tal como adelantaban O'Brien (2004) y Parra Escartín *et al.* (2017) en sus respectivos trabajos.

El control gramatical del lenguaje no resultó de utilidad al nivel esperado, especialmente al tener en cuenta el gran esfuerzo técnico de preedición y el esfuerzo cognitivo que conlleva su aplicación (Ramírez Polo, 2012) ante la falta de herramientas específicas para el español (González Sigüero, 2017). Además, su implementación reveló cuestiones a explorar en mayor profundidad con respecto a la interferencia entre la calidad del texto fuente y su traducibilidad a través de motores de traducción automática. Tal sería el caso de la eliminación de las voces pasivas, no deseables en el texto en español (Amador Iscla y Alexandre Benavent, 2002), pero que produjeron una mejora de la traducibilidad automática al inglés, ya que, como puso de manifiesto este trabajo, la reescritura a la voz activa es susceptible de introducir nuevos errores gramaticales. Aceptar estos anglicismos sintácticos en los textos científicos redactados en español quizá sea, por tanto, una concesión necesaria para la defensa del español desde una perspectiva más global a través de la reivindicación del multilingüismo científico con ayuda de la traducción automática.

Por otro lado, aunque el control léxico-terminológico resultó mucho más efectivo, el estudio puso de manifiesto la necesidad de seguir explorando esta cuestión con el fin de desarrollar herramientas que permitan su implementación de forma verdaderamente eficiente en español y así detectar los problemas de traducción automática en origen, tales como las traducciones imprevisibles para el traductor humano o las cuestiones de ambigüedad irresoluble. Este asunto tampoco estaría libre de futuras controversias en el marco de la defensa del español científico, sobre todo al tener en cuenta la tendencia actual hacia el empleo de un lenguaje

más claro (*plain language*) o más próximo al lenguaje natural, dado que un léxico más natural también introduciría un mayor grado de ambigüedad y reduciría la traducibilidad automática.

Aunque este trabajo sienta los cimientos para comenzar a llenar la laguna existente en el estudio del lenguaje controlado en español y de la traducibilidad automática al inglés, la metodología empleada no está exenta de limitaciones. En primer lugar, los resultados no son extrapolables ni generalizables, ya que el estudio se centró sobre un solo texto científico, con unas características únicas. Por otro lado, el trabajo fue realizado al completo por un único sujeto (la autora). Si bien esto presenta ciertas fortalezas, ya que se eliminan sesgos derivados de la intervención de diferentes lingüistas como consecuencia de una variación en sus destrezas y experiencia o problemas de consistencia metodológica, también introduce otros sesgos imposibles de eliminar. Al realizar la misma persona todo el flujo de preedición, posesición, revisión y anotación, intervienen procesos de familiarización con el texto y cuestiones de subjetividad. Además, todo el flujo de trabajo, incluido el proceso de anotación, que fue triple, es una tarea que requiere de un gran esfuerzo cognitivo, especialmente cuando se realiza de forma manual, como en este caso. Es altamente probable que el apoyo de herramientas informáticas para la aplicación del lenguaje controlado y la restricción léxico-terminológica hubieran permitido realizar un análisis más exhaustivo y robusto, de haber estado disponibles en español.

Los resultados de esta investigación, aparte de novedosos, pueden ser útiles como punto de partida para el estudio de nuevas técnicas y la creación de herramientas que permitan mejorar la traducibilidad automática de los textos científicos del español al inglés de forma general, independientemente de si se busca preservar la calidad del texto de partida.

A pesar del salto cualitativo que han experimentado en los últimos años los motores de traducción automática, este trabajo también refleja el elevado esfuerzo técnico necesario para traducir un artículo científico del español al inglés con las herramientas y técnicas disponibles actualmente. La incorporación de la traducción automática como el posible próximo vehículo de transmisión del conocimiento científico en detrimento del inglés como lengua franca, no solo no amenaza a la profesión del traductor científico, sino que le ofrece a este y al resto de los mediadores lingüísticos un mundo de posibilidades.

7. CONCLUSIONES

A pesar de que las conclusiones del presente trabajo no son extrapolables ni generalizables y que el estudio no está exento de limitaciones metodológicas, los resultados son novedosos y constituyen un punto de partida para continuar investigando la mejora de la traducibilidad automática al inglés de los textos científicos redactados en español.

La aplicación del lenguaje controlado resultó ser un proceso arduo y la evaluación de su eficacia precisó de un enorme esfuerzo cognitivo que no queda suficientemente reflejado en términos de magnitud a través de los esfuerzos técnicos cuantificados en este trabajo. Adicionalmente, la diversidad estilística de las secciones de un artículo científico requiere de una implementación del control del lenguaje muy desigual, ya que cada apartado constituye prácticamente un subgénero textual.

Aunque el esfuerzo técnico de posesición del texto controlado fue menor que el del texto bruto, la propuesta de lenguaje controlado de este trabajo no demostró la eficiencia esperada por la autora. La utilidad del control gramatical a la hora de prevenir errores fue baja de forma generalizada o difícil de cuantificar. Asimismo, existen errores inevitables, como la adición o la omisión de palabras, y que parecen responder a procesos arbitrarios, o la elección de las preposiciones como consecuencia de la polisemia. La naturalidad de la traducción, el principal escollo de los motores de traducción automática hasta la fecha, sí se vio considerablemente beneficiada, pero esto podría ser consecuencia de la reescritura de las frases requerida para implementar las reglas de control gramatical y no a la propia aplicación de estas. La preedición mejoró notablemente la imprecisión léxica gracias al control léxico-terminológico, que tuvo los mayores efectos y redujo el error de esta tipología a la mitad. Sin embargo, su eficiencia se encuentra limitada por la imposibilidad de desambiguar y detectar todo el léxico problemático durante la preedición sin una herramienta informática de apoyo. Este estudio reveló que el lenguaje controlado también puede introducir errores, tal y como quedó demostrado con la eliminación del indeseable abuso de la voz pasiva y la consiguiente generación de errores de tiempo verbal en la traducción del texto controlado. Por último, la traducción automática no consiguió adaptar la puntuación y las convenciones al idioma meta y además introdujo errores ortotipográficos de forma masiva y arbitraria. Por

todo ello, independientemente de si se edita previamente el texto, o de si se aplica lenguaje controlado, los resultados confirman que, por el momento, los textos resultantes de la traducción automática siguen requiriendo de la supervisión humana mediante un proceso de posesición.

Si bien el presente estudio confirma la hipótesis inicial, la mejora de la traducibilidad del texto no fue tan elevada como se esperaba. No obstante, los resultados de esta investigación arrojan luz sobre importantes áreas a trabajar con el fin de mejorar la traducibilidad automática del artículo científico en español, ahondar en la interferencia con otros paradigmas lingüísticos emergentes (defensa del español científico y lenguaje claro) y explorar la traducción automática como el próximo vehículo de transmisión del conocimiento científico para contribuir así a la democratización de la ciencia.

REFERENCIAS

- Adriaens, G. (1995). Simplified English grammar and style correction in an MT framework: The LRE SECC project. *Aslib Proceedings*. Vol 47 Nro. 3, pp. 73-82.
<https://doi.org/10.1108/eb051383>
- Aleixandre Benavent, R. y Amador Iscla, A. (2001a). Problemas del lenguaje médico actual (I). Extranjerismos y falsos amigos. *Papeles Médicos*, 10(3), 144-149.
- Aleixandre Benavent, R. y Amador Iscla, A. (2001b). Problemas del lenguaje médico actual (II). Abreviaciones y epónimos. *Papeles Médicos*, 10(4), 170-176.
- Amador Iscla, A. y Aleixandre Benavent, R. (2002). Problemas del lenguaje médico actual (III). Gramática y estilo. *Papeles Médicos*, 11(1), 18-23.
- Amano, T., González-Varo, J. P. y Sutherland, W. J. (2016). Languages Are Still a Major Barrier to Global Science. *PLOS Biology*, 14(12), e2000933.
<https://doi.org/10.1371/journal.pbio.2000933>
- Artigas Guillamón, M. del C. (2000). Especificidades de las lenguas científicas, técnicas y profesionales. *Anales de Filología Francesa*, 9, 23-44.
- Benfield, J. R. y Feak, C. B. (2006). How Authors Can Cope With the Burden of English as an International Language*. *CHEST*, 129(6), 1728-1730.
<https://doi.org/10.1378/chest.129.6.1728>
- Carmona Sandoval, J. C. (2013). Discurso y artículo científico. Una aproximación retórica. *Ra Ximhai*, 117-152. <https://doi.org/10.35197/rx.09.01.2013.07.jc>
- Claros Díaz, M. G. (2016). *Como traducir y redactar textos científicos en español. Reglas, ideas y consejos*. Fundación Dr. Antonio Esteve.
- Day, R. A. (2005). *Cómo escribir y publicar trabajos científicos*. Asociación Panamericana de la Salud.

- de Almeida, G. (2013). *Translating the post-editor: An investigation of post-editing changes and correlations with professional experience across two Romance languages*. [Tesis doctoral, Dublin City University]. http://doras.dcu.ie/17732/1/THESIS_G_de_Almeida.pdf
- De Biase, T., Escorsell, J., Gresta, C. y Rabasseda, X. (2009). La redacción médica en el seno de una agencia de información. En *La redacción médica como profesión* (pp. 29-38). Fundación Dr. Antonio Esteve.
- Domínguez García, I. (2009). Un acercamiento al lenguaje del texto científico. *Varona*, 48-49, 67-72.
- Drubin, D. G. y Kellogg, D. R. (2012). English as the universal language of science: Opportunities and challenges. *Molecular Biology of the Cell*, 23(8), 1399. <https://doi.org/10.1091/mbc.E12-02-0108>
- Englander, K. (2014). The Rise of English as the Language of Science. En *Writing and Publishing Science Research Papers in English: A Global Perspective* (p. 3). https://doi.org/10.1007/978-94-007-7714-9_1
- Escobar, G. Á. (2007). Importancia del lenguaje en el conocimiento y la ciencia. *Revista Virtual de Estudos da Linguagem - ReVEL*, 5(8), 16.
- Fernández Polo, F. y Cal Varela, M. (2009). English for research purposes at the University of Santiago de Compostela: A survey. *Journal of English for Academic Purposes*, 8, 152-164. <https://doi.org/10.1016/j.jeap.2009.05.003>
- Giba, J. (2014). *Developing skills in scientific writing*. Fundación Dr. Antonio Esteve.
- Gibbs, W. (1995). Lost Science in the Third World. *Scientific American*, 273(2), 92-99. <https://doi.org/10.1038/scientificamerican0895-92>

- Gil-Alberdi González, B. (2009). El redactor de textos médicos como profesional independiente. En *La redacción médica como profesión* (pp. 39-45). Fundación Dr. Antonio Esteve.
- Giménez Toledo, E., Kulczycki, E., Pölönen, J. y Sivertsen, G. (5 de diciembre de 2019). Bibliodiversity – What it is and why it is essential to creating situated knowledge. *LSE Impact Blog - Impact of Social Sciences*. <https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/05/bibliodiversity-what-it-is-and-why-it-is-essential-to-creating-situated-knowledge/>
- Gobbi, I. (2015). *Español Técnico Simplificado. Un lenguaje controlado para la redacción de manuales técnicos en español*. Eus - Ediz.
- González Sigüero, R. (2017). *State of the Art of Controlled Language*. [Trabajo de Fin de Máster, Universidad Politécnica de Madrid].
- Kancewicz-Hoffman, N. y Pölönen, J. (2020). Does excellence have to be in English? Language diversity and internationalisation in SSH research evaluation. *Overview of Peer Review Practices in the SSH*, 32-41. <https://doi.org/10.6084/M9.FIGSHARE.12032589>
- Kuhn, T. (2014). A Survey and Classification of Controlled Natural Languages. *Computational Linguistics*, 40(1), 121-170. https://doi.org/10.1162/COLI_a_00168
- Kulczycki, E., Mustajoki, H., Pölönen, J. y Røeggen, V. (6 de junio de 2019). *Polyglots need protection*. Research Professional News. <https://researchprofessionalnews.com/rr-news-europe-views-of-europe-2019-6-polyglots-need-protection/>
- Martínez Llenas, B. (2009). *Las perífrasis como elementos únicos del español*. Universitat Jaume I.
- Meza, P. (2017). *Ciencia y publicidad a través de las disciplinas: Estrategias de autopromoción en artículos de investigación de autores consolidados y en formación*, 32.

- Mitamura, T. y Nyberg, E. (1995). *Controlled English for Knowledge-Based MT: Experience with the KANT System*. 158-172.
- Navarro, F. A. (2001). El inglés, idioma internacional de la medicina. *Panace@*, 2(3), 35-51.
- Navarro, F. A. (2009). La precisión del lenguaje en la redacción médica. En *La redacción médica como profesión* (pp. 89-105). Fundación Dr. Antonio Esteve.
- O'Brien, S. (2019). Controlled language and writing for an international audience. En B. Maylath y K. St. Amant (Eds.), *Translation and Localization: A Guide for Technical and Professional Communicators*. Routledge.
- O'Brien, S. (2010). Controlled language and readability. En G. M. Shreve y E. Angelone (Eds.), *American Translators Association Scholarly Monograph Series: Vol. XV* (pp. 143-165). John Benjamins Publishing Company. <https://doi.org/10.1075/ata.xv.08obr>
- O'Brien, S. (2003). Controlling controlled English: An Analysis of Several Controlled Language Rule Sets. *EAMT/CLAW 2003 Conference Proceedings*, 105-114.
- O'Brien, S. (2004). Machine translatability and post-editing effort: How do they relate? *Translating and the Computer* 26, 31 pp. <https://aclanthology.org/2004.tc-1.3.pdf>
- Parra Escartín, C. y Goulet, M.-J. (2020). *When the Post-Editor is not a Translator: Can machine translation be post-edited by academics to prepare their publications in English?* En M. Koponen, B. Mossop, I.S. Robert y G. Scocchera (Eds.) *Translation Revision and Post-Editing*; Routledge. <https://doi.org/10.4324/9781003096962-8>
- Parra Escartín, C., O'Brien, S., Goulet, M.-J. y Simard, M. (2017). Machine Translation as an Academic Writing Aid for Medical Practitioners. *Research Track, 1*, 254-267.

- Pulido, M. (2010). El estilo científico. En *Publicación Científica Biomédica* (pp. 97-107). Elsevier. <https://doi.org/10.1016/B978-84-8086-461-9.50010-2>
- Ramírez Polo, L. (2012). Los lenguajes controlados y la documentación técnica: Mejorando la traducibilidad. *Tradumàtica: tecnologies de la traducció*, 10, 192-204.
- Rodríguez-Tapia, S. (2020). El grado de especialización textual del discurso médico en español: Análisis basado en corpus desde la teoría de la percepción. *Panacea*, 21(51), 75-92.
- Rosales, S. (2009). Perspectiva histórica de la redacción médica profesional. En *La redacción médica como profesión* (Vol. 17, pp. 1-6). Fundación Dr. Antonio Esteve.
- Rossetti, A., & O'Brien, S. (2019). Helping the helpers: Evaluating the impact of a controlled language checker on the intralingual and interlingual translation tasks involving volunteer health professionals. *Translation Studies*, 12(2), 253-271. <https://doi.org/10.1080/14781700.2019.1689161>
- Ruiz Cascales, R., & Sutcliffe, R. F. E. (2003). A Specification and validating parser for simplified technical Spanish. *EAMT Workshop: Improving MT through other language technology tools: resources and tools for building MT*.
- Saggion, H., Štanjer, S., Bott, S., Mille, S., Rello, L., & Drndarevic, B. (2015). Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish. *ACM Transactions on Accessible Computing (TACCESS) - Special Issue on Speech and Language Processing for AT (Part 2)*. https://www.academia.edu/38619748/Making_It_Simplext_Implementation_and_Evaluation_of_a_Text_Simplification_System_for_Spanish
- Taşkın, Z., Doğan, G., Kulczycki, E., & Zuccala, A. A. (2020, junio 18). Long read | Science needs to inform the public. That can't be done solely in English. *LSE Covid-19*. <https://blogs.lse.ac.uk/covid19/2020/06/18/long-read-science-needs-to-inform-the-public-that-cant-be-done-solely-in-english/>

UNGIS. (2019). *High-Level Political Forum 2019*.
<https://sustainabledevelopment.un.org/hlpf/2019>

Villagrán T., A. y Harris D., P. R. (2009). Algunas claves para escribir correctamente un artículo científico. *Revista Chilena de Pediatría*, 80(1), 70-78.
<https://doi.org/10.4067/S0370-41062009000100010>

WHO. (2019). *Ten health issues WHO will tackle this year*. WHO.
<https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019>

ANEXOS

ANEXO I. TEXTO DEL ARTÍCULO CIENTÍFICO ORIGINAL O TEXTO BRUTO

Propuestas para el control y la mitigación de la diseminación de COVID-19: un manejo estratégico de la enfermedad

RESUMEN

La velocidad de diseminación del COVID-19 en el mundo llevó a que los países afectados cerraran sus fronteras y tomaran medidas de distanciamiento social. Después de seis meses de que la enfermedad fuera declarada pandemia, muchos países están tomando medidas de flexibilización del aislamiento, aunque sin una vacuna o un medicamento capaz de enfrentar la infección por el SARS-CoV-2, la situación podría revertirse en cualquier momento. El objetivo del presente trabajo fue proponer un algoritmo de decisión tendiente a optimizar las detecciones de casos asintomáticos y administrar la cuarentena de una manera estratégica, para así evitar la diseminación del virus y tender hacia una normalidad administrada. Se elaboró una propuesta tentativa de optimización y ordenamiento de pruebas de detección del SARS-CoV-2, basada en el análisis de muestras compuestas reunidas a partir de aquellas tomadas de manera individual a personas asintomáticas que integran cohortes de interés. Se definieron cohortes según su función en la sociedad o grado de vulnerabilidad. El algoritmo contempla variables como la prioridad de la cohorte, el número de integrantes de los grupos de análisis dentro de cada cohorte, el contacto intragrupal e intergrupalo, la vulnerabilidad al contagio por la actividad desarrollada y el tiempo transcurrido desde que se realizó la prueba por última vez. Se ilustró la propuesta con cohortes hipotéticas definidas, con un único grupo de análisis para simplificar, y se comprobó que la aplicación de la herramienta permite establecer de una manera racional un orden de prioridad para realizar las pruebas en grupos críticos de la sociedad. Esta herramienta permitirá optimizar recursos y disminuir el impacto de la enfermedad en la salud, la sociedad y la economía de una región.

PALABRAS CLAVE: COVID-19, Distanciamiento social, SARS-CoV-2, Algoritmo de decisión, Cuarentena

INTRODUCCIÓN

Los coronavirus son una familia de virus que infectan animales, principalmente aves y mamíferos, en los que causan enfermedades; esto se traduce en pérdidas económicas asociadas a sus actividades productivas. También pueden saltar de especies e infectar a los seres humanos, en los que causan enfermedades de distinta gravedad (resfrío, bronquitis, neumonía, entre otras), dependiendo del virus en particular y de las condiciones de salud del huésped. Si bien la mayoría de las personas se infectan con estos virus alguna vez en la vida sin padecer grandes consecuencias, algunos coronavirus han causado enfermedades que se han diseminado rápidamente y han provocado brotes y epidemias preocupantes. Como ejemplos se pueden citar el síndrome respiratorio agudo grave (SARS) originado en Yunan, China, en 2001; el síndrome respiratorio de Oriente Medio (MERS) causado por el MERS-CoV, iniciado en Arabia Saudita en 2012, y actualmente el COVID-19, causado por un nuevo coronavirus, denominado SARS-CoV-2.

Se cree que esta nueva enfermedad, COVID-19, surgió en diciembre de 2019 en un mercado de mariscos (en donde se vendían animales exóticos vivos y muertos) ubicado en Wuhan, provincia de Hubei, China, aunque hay algunas controversias sobre esto. La velocidad de diseminación del virus fue tan elevada que, en poco tiempo, las autoridades de China cerraron la ciudad y dispusieron cuarentena obligatoria. A pesar de ello, el contagio se extendió con rapidez por los

países limítrofes primero, y, paulatinamente, por los demás, facilitado por la gran movilidad de viajeros. Tan solo después de casi tres meses de haberse detectado la enfermedad, el 11 de marzo de 2020, la Organización Mundial de la Salud (OMS) la declaró una pandemia.

El rápido avance de la enfermedad puso en evidencia la falta de preparación de los sistemas de salud de la mayoría de los países para afrontar la situación. El número de enfermos sobrepasó con creces las capacidades disponibles y el mundo observó con estupor la creciente cifra de muertos en países desarrollados de Europa. Ante ese panorama y con el objeto de contener el contagio, en la mayoría de los países se fueron estableciendo medidas de distanciamiento social (más conocidas como cuarentena), con distintas modalidades. A la fecha de escritura de este artículo (19 de mayo de 2020), el número de casos confirmados en el mundo se acercaba rápidamente a los cinco millones, de los cuales aproximadamente el 7% correspondía a fallecimientos. (A la fecha de la corrección de este manuscrito, 15 de setiembre de 2020, hay casi 30 millones de casos confirmados en el mundo y el número de fallecimientos se acerca al millón).

De manera muy temprana (20 de marzo de 2020), el gobierno argentino estableció el “aislamiento social preventivo y obligatorio” (ASPO). Esto permitió evitar el contagio masivo y ganar tiempo para que el sistema de salud pudiera mejorar su capacidad de respuesta. Así, se montaron nuevos hospitales o centros de atención; se adquirieron y/o fabricaron respiradores, elementos de protección personal y demás insumos para poder atender la emergencia, a la vez que se convocó al sistema científico a contribuir en la generación de conocimientos para afrontar la pandemia.

Al igual que en el resto del mundo, la detección del SARS-CoV-2 se lleva a cabo en Argentina mediante pruebas de reacción en cadena de la polimerasa en tiempo real con retrotranscripción (RT-qPCR). En principio, esta prueba se le realiza a cada individuo sospechoso de ser portador debido a que presenta síntomas relacionados con la enfermedad o por conexión epidemiológica, aunque esto ha ido cambiando con la evolución de la pandemia. Otros test moleculares recientemente desarrollados por investigadores argentinos permiten detectar la presencia de fragmentos genéticos y han mostrado resultados prometedores; estos son el NEOKIT y el ELA-CHEMSTRIP. La ventaja de estos últimos métodos es que permitirían identificar al virus de manera más rápida, lo que haría posible su detección en un grupo poblacional más grande. A pesar de esto, en la mayoría de los países en desarrollo la capacidad para realizar dichos análisis es reducida, por lo que no se puede evaluar a la población que se supone sana o asintomática (aunque puede incluir individuos contagiosos) y, por lo tanto, se deben tomar medidas generalizadas para controlar la diseminación del virus.

Teniendo en cuenta el contexto sanitario que se vive y que el distanciamiento social ha cumplido su propósito en cuanto al enlentecimiento de la diseminación de la enfermedad, se considera necesario volver paulatinamente y de manera controlada a la actividad. No obstante, se sabe que, en caso de deteriorarse la situación sanitaria, las medidas de flexibilización podrían ser revertidas. Por ello, el objetivo del presente trabajo fue proponer un algoritmo de decisión tendiente a optimizar las detecciones de casos asintomáticos y administrar la cuarentena de una manera estratégica, para así evitar la diseminación del virus y permitir una nueva normalidad administrada.

MATERIALES Y MÉTODOS

Se elaboró una herramienta tentativa para utilizar estratégicamente las pruebas de detección del virus SARS-CoV-2 con el objetivo de controlar la diseminación de la enfermedad en múltiples cohortes. Para ello, y como ya fue indicado por otros autores, se sugirió analizar muestras compuestas (análisis grupal) combinando aquellas tomadas de manera individual a distintas personas asintomáticas que integran la cohorte de interés. De este modo, se podría evaluar la presencia del virus en mayor cantidad de personas simultáneamente, con un ahorro considerable de insumos y de tiempo. Se propuso comenzar analizando cohortes de personas cuya función fuera crítica para el normal funcionamiento de la sociedad o pertenecientes a sectores muy vulnerables, para luego continuar con cohortes no críticas, según su importancia y grado de vulnerabilidad.

Definición de cohortes críticas

Las cohortes que aquí se proponen como críticas y que deberían ser las primeras en someterse a la detección del virus abarcarían los siguientes rubros: salud, seguridad, comercio de bienes de consumo esenciales, tercera edad (geriátricos), servicios e industrias esenciales (fabricación de respiradores, de alimentos, etc.). Es fundamental estudiar a los miembros de estos sectores, al ser potenciales agentes de contagio. El personal de salud (médicos, enfermeros, personal de limpieza, etc.) debería mantenerse en óptimas condiciones, por lo que, además de proveerles los elementos de protección personal, se les debería realizar el análisis de manera sistemática y periódica. Lo mismo vale para el personal de seguridad, ya que está muy expuesto a personas potencialmente infectadas. Los comerciantes de la cadena de suministros, incluyendo los grandes mercados de frutas y verduras, deberían ser analizados regularmente también, ya que una persona infectada podría hacer que se paralice toda la cadena y se generen graves problemas sociales. Otros sectores, como el de trabajadores de la construcción, administrativos, comerciantes de elementos no esenciales, habitantes de zonas residenciales, por mencionar algunos, corresponden a sectores no críticos, por lo que se analizarían cuando la cantidad de pruebas disponibles fuera suficiente, sin comprometer el control de los sectores que sí lo son. Sin embargo, aquellas cohortes no críticas que tengan relación con integrantes de cohortes críticas deberían tener mayor prioridad de análisis.

Realización de los análisis

Luego de definir los grupos (o *pools*) de las distintas cohortes, se propone tomar muestras individuales (hisopados nasofaríngeos) a todas las personas asintomáticas que los integran. Una parte de la muestra podría ser mezclada en el laboratorio para conformar una muestra compuesta de cada grupo para su análisis. Inicialmente, se debería aproximar la tasa de infectividad y la prevalencia a partir de datos oficiales de los contagios en el lugar geográfico analizado, con el objetivo de definir el tamaño del *pool* y la metodología de detección por grupos.

El análisis por *pools* de muestras tiene sentido como ahorro de insumos y de tiempo, siempre y cuando la prevalencia sea menor que el 10%, valor a partir del cual conviene realizar los testeos individuales debido a la alta cantidad de *pools* que darían positivos (y tendrían que “abrirse” para estudiar las muestras individuales). El tamaño de los *pools* estaría definido por la prevalencia: mientras mayor sea esta, menor será el número de muestras que conformarían un *pool*. La prevalencia de la enfermedad en la población y también en las cohortes conformadas por personal esencial se consideraría información epidemiológica conocida, que podría calcularse a partir de los casos (individuales) confirmados en los grupos de interés en relación con la población o un tamaño definido de la población.

Cabe aclarar que antes de iniciar la detección mediante *poles* de muestras habría que evaluar algunos aspectos metodológicos críticos para asegurar el éxito de esta modalidad diagnóstica. Lo primero que se debería verificar a través de la bioinformática es que las secuencias de los oligonucleótidos que se usarán en la detección son específicas. Luego, correspondería determinar la eficiencia de la amplificación de la reacción de RT-qPCR y verificar que la eficiencia sea máxima, aun a bajas concentraciones de ARN. Además, sería imprescindible determinar el límite de detección del método empleado, ya que de eso dependerá la clasificación de los resultados entre “positivos” y “no detectables (negativos)”. Finalmente, se debería realizar una validación de la técnica de *poles* en un grupo de la población y comparar esos resultados con los que se obtienen por testeos individuales para determinar la probabilidad de una inadecuada clasificación de un resultado. Recién al contar con todo el conocimiento mencionado se podría empezar con la detección en *poles* de muestras.

El número máximo de individuos que podrían conformar un grupo debería definirse teniendo en cuenta el volumen por muestra individual que se debe agregar a la reacción de detección sin comprometer la sensibilidad, pero permitiendo suficiente carga viral para lograrla. Además, se debería tener en cuenta la prevalencia del virus o la relación entre la cantidad de individuos infectados y la población total analizada para optimizar el tamaño de los grupos desde un punto de vista estadístico. Estos valores servirían de guía, pero la definición de los grupos debería realizarse de acuerdo con la conveniencia práctica. Por ejemplo, si el tamaño óptimo de grupos fuera 15 personas y en un geriátrico hubiera 10 personas, no sería necesario ni conveniente sumar gente externa al grupo para llegar al valor de tamaño óptimo. Por otro lado, si se analizase a la cohorte de un barrio completo, los grupos deberían contar con el tamaño óptimo y ser seleccionados según su proximidad, con el objetivo de ahorrar recursos.

Los grupos con resultados negativos en pruebas de SARS-CoV-2 se deberían analizar cada cierto período de tiempo, el cual tendría que optimizarse según un algoritmo que resolvería el orden de prioridad de cada grupo. De resultar positiva alguna de las pruebas realizadas, se debería aislar al grupo implicado y a las personas con las que hayan tenido contacto estrecho para evitar contagios, hasta completar el análisis. Para ello, el laboratorio debería “abrir los *poles*” positivos y profundizar en el análisis de las muestras individuales para identificar al individuo infectado, quien luego debería permanecer aislado (fig. 1). Las personas con resultados individuales negativos podrían salir del aislamiento.

Definir las cohortes a analizar

Tomar muestras de los individuos de la cohorte y combinarlas en una grupal

Realizar análisis de muestra grupal

Calcular tiempo a esperar según prioridad de cohorte

Negativo

Resultado del análisis del Grupo

Positivo

Mantener alerta y distanciamiento por 2 semanas

Analizar las muestras individualmente y/o por subgrupos

Negativo

Resultado del análisis individual

Positivo

Figura 1. Diagrama de flujo que indica los pasos a seguir para mitigar la diseminación de la enfermedad en la población y optimizar el uso de los insumos de diagnóstico.

Optimización de recursos

Se desarrolló un algoritmo de decisión para optimizar el testeo estableciendo un orden lógico para su realización, con el fin de modelar y evaluar el comportamiento de la transmisión de la enfermedad, en función del impacto que las cohortes definidas tienen sobre la salud, la seguridad y la economía. Idealmente, se requiere un conocimiento profundo de los integrantes de las cohortes de análisis, por lo que la implementación de este algoritmo a nivel municipal podría resultar más beneficiosa.

Se consideró necesario, aunque no excluyente, incluir las siguientes variables generales en el modelo, ya que se supone que afectan a todos los grupos de todas las cohortes, y estas deberían actualizarse diariamente:

Cantidad de pruebas disponibles: condiciona la cantidad de grupos a los cuales se les puede realizar el análisis y, por lo tanto, la cantidad de cohortes que pueden ser analizadas periódicamente.

Función de contagio general: se define como la relación entre la población que es susceptible de contagiarse, la infectada y la recuperada. Se aplica en una zona que abarca a las cohortes analizadas y depende de los valores históricos de esas variables.

Otras variables importantes son las relacionadas con características particulares del grupo a analizar:

Ubicación/coordenadas: permite conocer la localización del grupo en cuestión y puede ser útil para determinar el nivel de contacto intergrupar o ayudar a decidir cuál es la vulnerabilidad al contagio por la actividad. Por ejemplo, podría existir mayor riesgo de contagio para un grupo de seguridad que comparta la ubicación con un grupo de un hospital donde hay individuos positivos para SARS-CoV-2.

Grado de prioridad del grupo (A): alude a la prioridad de un grupo por sobre otros para realizar el análisis. Por ejemplo, grupos de la cohorte del sector salud (médicos y enfermeros dedicados al tratamiento de la enfermedad ocasionada por el SARS-CoV-2) tendrían prioridad de ser analizados por sobre un grupo de una cohorte integrada por personas dedicadas a la producción de bienes de consumo. Para que esta decisión sea lo menos subjetiva posible, se definió una escala de 1 a 10, en la que 1 significa nada prioritario y 10 indica máxima prioridad (tabla 1). Se establecieron rangos específicos para cada grupo hipotético incluido en el algoritmo.

Tabla 1. Escala de prioridad de grupos para el testeo

Prioridad del grupo (A)

Rango

Médicos dedicados al tratamiento directo de infectados por SARS-CoV-2

Otros médicos

Miembros de un geriátrico

Personal de seguridad

Comerciantes de bienes de consumo
Industria
Grupos residenciales

Número de integrantes de cada grupo (B): la cantidad de integrantes de cada grupo se definiría de acuerdo con la conveniencia y debería ser tal que los negativos del análisis sean negativos verdaderos (que no comprometan el límite de detección de la muestra), para evitar que se continúe diseminando la infección. Esto significa que si el análisis de un grupo infectado se hace, por ejemplo, por RT-qPCR, la sensibilidad dependerá de la fase de infección en curso, de la calidad de las muestras extraídas y de la cantidad de muestras individuales que formen parte de la muestra compuesta (a mayor número de muestras, menor cantidad de cada blanco y, por tanto, menor sensibilidad).

Contacto intragrupal (C): esta variable indica el nivel de contacto (relacionado con la distancia social y la frecuencia del contacto) entre los miembros del grupo analizado. Se estableció para su valoración una escala entre 1 y 10, donde valores altos indican mayor contacto y valores bajos menor contacto (al contacto nulo se le adjudicó el valor 1). Los valores de contacto para diferentes distancias pueden coincidir a causa de un cambio en la frecuencia del contacto (tabla 2).

Tabla 2. Escala para el contacto intragrupal (C)

Distancia (m)

Frecuencia

Rango

Baja

Media

Alta

Baja

Media

Alta

Baja

Media

Alta

Vulnerabilidad al contagio por la actividad (D): hace referencia al grado de vulnerabilidad de los grupos estudiados dentro de una cohorte a causa de la actividad. Se estableció también en este caso una escala entre 1 y 10, en la que un mayor valor indica mayor nivel de vulnerabilidad al contagio por la actividad (tabla 3). Los valores se podrían obtener a partir de estudios del nivel de contagio en estos grupos y deberían ser actualizados periódicamente.

Tabla 3. Escala para la vulnerabilidad de contagio (D)

Actividad

Valor

Médicos de familia

Enfermeros especializados

Técnicos de farmacia

Vigilante de prisión

Residenciales

Vulnerabilidad por el tiempo transcurrido (E): valor calculado (ecuación 1) entre 0 y 10 que indica el riesgo de que exista una infección no detectada en el tiempo transcurrido.

donde t (en días) es el tiempo transcurrido desde la última prueba realizada y t_i es el período de incubación máximo del virus (14 días).

Activación intergrupala (F): valor que se utiliza para indicar si aumentará (o no) el orden de prioridad a causa de que existe relación con un grupo cuyo análisis dio positivo. Es decir que a causa de la relación intergrupala con grupos infectados, el grupo analizado aumentará su orden de prioridad y, por lo tanto, podría implicar que sea analizado primero. Así F tomará el valor de 1 si la prueba fuera positiva y 0 si resultara negativa (no se activa la relación intergrupala).

Contacto intergrupala (G): variable que establece el nivel de contacto entre grupos distintos (de una misma cohorte o de una diferente). Se definió una escala entre 1 y 10 (al contacto nulo se le adjudicó el valor de 1), donde valores altos indican un contacto más estrecho entre los miembros de distintos grupos (la relación del grupo i con el grupo j se define como relación ij , con $i \neq j$). Los valores de contacto para diferentes distancias pueden coincidir a causa de un cambio en la frecuencia del contacto (tabla 4).

Tabla 4. Escala para contacto intergrupala (G)

Distancia (m)

Frecuencia

Rango

Baja

Media

Alta

Baja

Media

Alta

Baja

Media

Alta

Orden de análisis general (O): valor cuyo resultado permitiría ordenar la prioridad de realización de las pruebas según el modelo actual. Primero, teniendo en cuenta las variables antes establecidas, se calcula un valor local para cada grupo (L_i , ecuación 2) y se identifica el mayor valor ($L_{máx}$). Los valores locales representarían la importancia en el análisis y no tienen una escala definida. Luego, para visualizar mejor cuáles grupos deben analizarse primero, se ponderarán los valores locales respecto del valor local máximo ($L_{máx}$) y se obtendrá el orden de prioridad de análisis general para cada grupo (O_i , ecuación 3).

donde j varía desde 1 hasta j máximo (cantidad total de grupos definidos), con $i \neq j$:

Una vez definido el orden de análisis general de los grupos (O), se podrá realizar la detección siguiendo dicho orden (de mayor a menor) y se actualizará la cantidad de pruebas disponibles. Luego, se retroalimenta la información obtenida al algoritmo con las variables E (vulnerabilidad por tiempo transcurrido) y F (activación intergrupala) (fig. 2).

GRUPO j
GRUPO i
(G) Contacto intergrupala
Ordenar análisis
VARIABLES DE GRUPO
Ubicación / coordenadas
(A) Prioridad del grupo
(B) Número de integrantes
(C) Contacto intragrupal
(D) Vulnerabilidad a contagio por la actividad
(O) Orden de análisis
Analizar grupos
VARIABLES DE RETROALIMENTACIÓN
(E) Vulnerabilidad por tiempo transcurrido
(F) Activación intergrupala
VARIABLES GENERALES
Cantidad de pruebas disponibles
Función del contagio general

Figura 2. Algoritmo general para calcular el orden de análisis de los grupos y hacer un uso racional de las pruebas de detección disponibles. Los recuadros blancos indican las variables de entrada y de salida del modelo, mientras que los bloques azules son las operaciones.

Alcance temporal y costos

Según la OMS, se espera que esta pandemia se extienda por lo menos por dos años. Aunque ya existen muchos proyectos tanto de vacunas como de tratamientos paliativos, todavía ninguno ha sido implementado de forma eficaz y masiva, por lo que el tiempo de duración de la pandemia puede resultar difícil de predecir. Del mismo modo, el tiempo de aplicación de la herramienta propuesta en este trabajo es difícil de establecer, ya que depende de muchas variables y está íntimamente relacionado con la duración de la pandemia. La herramienta es muy flexible, ya que permite operar con diferentes cantidades de cohortes (según se requiera o según el presupuesto del que se disponga) y también permite hacer los testeos a la frecuencia conveniente para cada localidad en la que se aplique. Por ello, el costo de implementación dependerá, principalmente, de la capacidad de testeo disponible. Otra variable que podría influir en los costos es la posible implementación de test rápidos, los cuales han sido desarrollados y estarán disponibles (principalmente para el sistema de salud) en los próximos meses. Es importante remarcar que el análisis de muestras agrupadas permite obtener una gran cantidad de información y minimizar los costos al analizar numerosas muestras individuales en una sola reacción.

RESULTADOS Y DISCUSIÓN

Para ilustrar la aplicación del algoritmo y evaluar su versatilidad y factibilidad se llevó a cabo una simulación. Debido a que no se contaba con datos reales al momento de formular este algoritmo,

fue necesario definir algunos de manera arbitraria (intentando acercarnos a la realidad) y realizar diferentes suposiciones.

En primer lugar, se definieron ocho grupos hipotéticos (con fines ilustrativos, ya que en la realidad serán muchos más): GRUPO 1, médicos pertenecientes al hospital 1 (m1); GRUPO 2, personal de seguridad del hospital 1 (s1); GRUPO 3, médicos del hospital 2 (m2); GRUPO 4, personal de seguridad del hospital 2 (s2); GRUPO 5, miembros de un geriátrico (ancianos y personal) que se encuentra en una zona urbana (g1), bajo el supuesto de que no se registraron casos en la zona; GRUPO 6, personal del comercio mayorista 1 de bienes de consumo, ubicado en una zona donde existen otros comercios mayoristas (c1); GRUPO 7, industria de fabricación de insumos ubicada en una zona en donde sí se registraron casos de COVID-19 (i1); GRUPO 8, familia numerosa de un barrio residencial, ubicado en una zona urbana donde no se registraron casos de COVID-19 (r1). Para cada grupo se estableció una cantidad específica de integrantes (tablas 5 y 6).

Tabla 5. Ejemplo hipotético de aplicación del algoritmo de decisión desarrollado

Grupos

Prioridad del grupo (*A*)

Número de integrantes de cada grupo (*B*)

Contacto intragrupal (*C*)

Vulnerabilidad al contagio por la actividad (*D*)

Días transcurridos desde el análisis

Vulnerabilidad por tiempo transcurrido (*E*)^a

Orden de análisis local (*Li*)^b

Orden de análisis general (*Oi*)^c

Orden de prioridad de pruebas

^{a,b,c} Calculados mediante las ecuaciones 1, 2 y 3.

Grupos definidos. m1: médicos del hospital 1; s1: seguridad del hospital 1; m2: médicos del hospital 2; s2: seguridad del hospital 2; g1: miembros en el geriátrico 1; c1: personal del comercio 1; i1: personal de la industria 1; r1: vecinos 1 de la zona residencial. En negritas se marca el grupo que se supone infectado.

Tabla 6. Ejemplo hipotético para ilustrar una matriz de relaciones intergrupales

Grupos

F: la activación intergrupala (1: infectado, 0: no infectado); NA: no aplica el contacto intergrupala para la relación de un grupo consigo mismo.

Grupos definidos. m1: médicos del hospital 1; s1: seguridad del hospital 1; m2: médicos del hospital 2; s2: seguridad del hospital 2. En negritas se marca el grupo que se supone infectado.

Con respecto a la variable prioridad del grupo (*A*), se asignó un valor de 10 a los médicos del hospital 1 (m1) (que registra mayor cantidad de pacientes con COVID-19) y 9 a los del hospital 2 (m2). Los altos valores asignados se deben a que ambos son grupos prioritarios (al formar parte de la cohorte “médicos de hospitales”), en concordancia con su función de atender a personas con COVID-19. Para los grupos de la cohorte seguridad, se estableció un valor de 6 tanto para s1 como para s2, ya que a pesar de que son necesarios, podrían ser reemplazados si sucediese un contagio. Los miembros del geriátrico g1 son un grupo de alto riesgo, debido a que la mortalidad por contagio es mucho mayor en este grupo, por lo que se adjudicó un valor de 8. El comercio

mayorista de bienes de consumo $c1$ tiene una prioridad de 4 (relativamente baja), ya que existen otros comercios mayoristas alternativos en la zona. Para la industria $i1$, el valor asignado fue 7, porque se encuentra en una zona de riesgo entre moderado y alto. Al grupo residencial $r1$ se le asignó un valor de 3, ya que se encuentra en una zona urbana con tránsito moderado, donde no se registraron casos.

Para la variable vulnerabilidad al contagio por la actividad (D), el hospital $h1$ es uno de los designados para tratar a personas con coronavirus, mientras que el hospital $h2$ no lo es. Por lo tanto, se adjudicó un puntaje de 10 a $m1$ y de 8 a $m2$ para ponderar el riesgo de contagio. Por la misma razón, el riesgo de contagio de $s1$ se estableció en 6 y el de $s2$ en 4. Aunque la actividad de $g1$ no es riesgosa, sí se trata de un grupo muy vulnerable, por lo que se le asignó un valor de 5. El grupo $c1$ se encuentra en una zona libre de COVID-19 y tiene un valor de 6 debido a que existe un flujo considerable de personas que no corresponden a la zona y un contacto cercano a los clientes. Finalmente, el grupo $i1$ desarrolla una actividad en contacto mínimo con personas externas, por lo tanto, se le asignó un valor de 3.

Continuando el análisis con la variable contacto intragrupal (C), los médicos $m1$ tienen reuniones diarias en el hospital 1 y con una frecuencia media trabajan en conjunto y a distancias menores de 2 metros. Esta suposición es válida también para los médicos $m2$ del hospital 2. De aquí que el contacto intragrupal para ambos casos fue puntuado con 8. El contacto entre los individuos de los grupos de seguridad ($s1$ y $s2$) se da, principalmente, a una distancia menor de 2 metros y con baja frecuencia, en situaciones relacionadas con mantener el orden, por lo que se asignó un valor de 6. Los ancianos y personal del geriátrico ($g1$) se reúnen a distancias entre 1 y 3 metros y a una frecuencia alta, por lo que se asignó un valor de 9. El personal tanto de $c1$ como de $i1$ se reúne a frecuencia y distancia media, por lo que se asignó en ambos casos un valor de 6. Finalmente, el grupo residencial $r1$ se reúne a distancias bajas y con frecuencia alta, por lo que se le asignó un valor de 10.

Terminando con las variables, el contacto intergrupala (G) de los médicos $m1$ y $m2$ es a frecuencia baja y a distancias mayores de 4 metros debido a solicitudes de participación especial o en cursos de actualización, por lo que el valor para la relación intergrupala en este caso es 2. Esta suposición es válida también para la relación entre los grupos de seguridad $s1$ y $s2$. El resto de las relaciones intergrupales es inexistente, por lo que su valor se estableció en 1 ([tabla 5](#)).

Suponiendo que en el último análisis realizado 6 días antes, al grupo $s1$ le hubiera dado positivo ([tabla 5](#)), habría dos consecuencias directas. En primer lugar, ese grupo debería enviarse a aislamiento preventivo hasta que se pueda identificar al individuo infectado (o a los individuos infectados), para lo cual se tendría que “abrir” el pool y habría que analizar las muestras individuales, y luego proceder a los correspondientes reemplazos laborales ([fig. 1](#)). En segundo lugar, esto ocasionaría un cambio en el orden de análisis para los otros grupos por el contacto intergrupala. En este ejemplo no se analizó al grupo de seguridad $s1$ por subgrupos, ya que se prevé suficiente disponibilidad de reemplazos hasta que $s1$ vuelva de la cuarentena; no obstante, si eso se quisiera hacer, deberían seguirse los pasos establecidos ([fig. 1](#)).

Se puede observar, entonces, que se debería realizar el análisis en primer lugar a los médicos del hospital 1 ($m1$; $O_i = 100$), seguido por el grupo del geriátrico 1 ($g1$; $O_i = 67$), los médicos del

hospital 2 (m2; $O_i = 61$), la seguridad del hospital 2 (s2; $O_i = 33$), comercio mayorista (c1; $O_i = 18$), industria (i1; $O_i = 9$) y, finalmente, al grupo residencial (r1; $O_i = 3$). Se debe tener en cuenta que el grupo de seguridad s1 se encontraría en el sexto día del aislamiento preventivo (cuarentena), por lo que debería ser analizado nuevamente una vez transcurrido el período de incubación del virus (tabla 5). Si bien el grupo médico del hospital 1 (m1) tuvo un contacto intergrupar con el grupo de seguridad infectado (s1), transcurrieron 6 días desde su último análisis, mientras que fueron 12 días para el grupo m2 del hospital 2. Además, el hecho de que no existan contagios grupales en el hospital 2 no implica que ese hospital no esté recibiendo personas enfermas que podrían contagiar a los grupos analizados en dicho hospital. Si, en cambio, el grupo de seguridad s1 hubiese dado negativo, se debería realizar el análisis, en primer lugar, al grupo del geriátrico 1 (g1; $O_i = 100$), seguido por los médicos del hospital 2 (m2; $O_i = 91$), comercio mayorista (c1; $O_i = 26$), la seguridad del hospital 2 (s2; $O_i = 25$), los médicos del hospital 1 (m1; $O_i = 25$), industria (i1; $O_i = 14$), el grupo de seguridad s1 (s1; $O_i = 7$), y, finalmente, al grupo residencial (r1, $O_i = 4$).

Este ejemplo permite visualizar el potencial para el ahorro en pruebas cuando se analiza la diseminación del COVID-19 en un determinado municipio. Se podrían analizar hasta 55 individuos distribuidos en 8 grupos a través de 8 pruebas, lo que significaría un ahorro del 85,45% de las pruebas si se procede con el método de análisis grupal, respecto del individual.

En otro trabajo de investigación en donde se realizaron experimentos a través de RT-qPCR, se encontró que se pueden analizar grupos de hasta 32 individuos, con un 10% de falsos negativos para la presencia de SARS-CoV-2¹⁴. Por otro lado, Hogan et al. (2020) analizaron mediante la técnica de RT-qPCR muestras almacenadas en un hospital tomadas de personas en cuyas historias clínicas se consignaban síntomas de infección por coronavirus, en grupos de hasta 10 individuos. Tres de los 292 grupos analizados por estos autores dieron resultados positivos para infección con SARS-CoV-2. Con los análisis individuales, luego se comprobó que dos de los tres grupos estaban integrados por personas infectadas¹². Sin embargo, como la calidad de las pruebas de detección y los protocolos varían internacionalmente, se recomienda realizar experimentos independientes antes de llevar a la práctica la propuesta de este trabajo.

El algoritmo presentado permite establecer un orden racional de prioridad para realizar el testeo de grupos críticos de la sociedad (fig. 3). Como una limitación relevante de esta propuesta, cabe destacar que la asignación de los valores de las variables de los grupos se basa en el criterio y la experiencia del operador de la herramienta, además de las particularidades epidemiológicas analizadas. Por esta razón, de aplicarse esta herramienta, su potencial irá aumentando a medida que avance la infección en la comunidad y el conocimiento de las particularidades que con ella se presentan.

Definir grupos
Ordenar análisis
Analizar grupos
GRUPO 1
GRUPO 2
GRUPO 3
GRUPO 4
GRUPO i

VARIABLES GENERALES
Función de contagio general
Cantidad de pruebas disponibles
ORDEN DE ANÁLISIS
GRUPO 1
GRUPO 2
GRUPO 3
GRUPO 4
GRUPO i
VARIABLES DE GRUPO
Prioridad del grupo
Número de integrantes
Contacto intragrupal
Vulnerabilidad por actividad
Contacto intergrupalo
GRUPO 2
Muestra combinada
Análisis GRUPO 2
Actividad normal
GRUPO 4
Muestra combinada
Análisis GRUPO 4
Cuarentena
VARIABLES DE RETROALIMENTACIÓN
Vulnerabilidad por tiempo transcurrido
Activación intergrupalo
Figura 3. Esquema de trabajo propuesto, en el que se indican los pasos a seguir con las variables correspondientes a cada etapa.

CONCLUSIONES

Mediante este análisis simulado demostramos que el algoritmo propuesto es una opción interesante y puede ser una herramienta para el control de la diseminación del SARS-CoV-2, por lo que merece ser explorada y, eventualmente, desarrollada, con el fin último de lograr una mejor administración del distanciamiento social empleando la menor cantidad del insumo limitante (pruebas de detección). Esta herramienta presenta un alto potencial para mitigar los problemas que genera la pandemia, como así también para disminuir la probabilidad de rebrotes cuando se cuenta con recursos escasos, con la consiguiente reducción del impacto en la salud, la sociedad y la economía de una región.

ANEXO II. TEXTO CONTROLADO

Propuestas para el control y la mitigación de la diseminación de COVID-19: un manejo estratégico de la enfermedad

RESUMEN

La velocidad de diseminación de la COVID-19 en el mundo ocasionó que los países afectados cerraran sus fronteras y tomaran medidas de distanciamiento social. Seis meses después de declarar la enfermedad como pandemia, muchos países tomaron medidas de flexibilización del aislamiento. Sin embargo, sin una vacuna o un medicamento capaz de frenar la infección por el SARS-CoV-2, la situación podría revertirse en cualquier momento. El objetivo del presente trabajo fue proponer un algoritmo de decisión para optimizar la detección de casos asintomáticos y prescribir la cuarentena de una manera estratégica. Medidas como estas previenen la diseminación del virus y permiten avanzar hacia una normalidad supervisada. Elaboramos una propuesta tentativa de optimización y ordenamiento de pruebas de detección del SARS-CoV-2 basada en el análisis de muestras compuestas a partir de muestras individuales de personas asintomáticas de las cohortes de interés. Definimos las cohortes según su función en la sociedad o el grado de vulnerabilidad. El algoritmo contempla variables como la prioridad de la cohorte, el número de integrantes de los grupos de análisis dentro de cada cohorte, el contacto intragrupal e intergrupala, la vulnerabilidad al contagio por la actividad desarrollada y el tiempo transcurrido desde que se realizó la prueba por última vez. Ilustramos la propuesta con cohortes hipotéticas definidas, con un único grupo de análisis para simplificar. Comprobamos que la aplicación de la herramienta permite establecer un orden de prioridad para realizar las pruebas en grupos críticos de la sociedad de una manera racional. Esta herramienta permitirá optimizar recursos y disminuir el impacto de la enfermedad en la salud, la sociedad y la economía de una región.

PALABRAS CLAVE: COVID-19, Distanciamiento social, SARS-CoV-2, Algoritmo de decisión, Cuarentena

- Comentado [MAP1]: CE
- Comentado [MAP2]: CG6
- Comentado [MAP3]: EE
- Comentado [MAP4]: CG3
- Comentado [MAP5]: CG10
- Comentado [MAP6]: CE
- Comentado [MAP7]: CG1
- Comentado [MAP8]: EE
- Comentado [MAP9]: CLT
- Comentado [MAP10]: CG6
- Comentado [MAP11]: CE
- Comentado [MAP12]: CLT
- Comentado [MAP13]: CG1
- Comentado [MAP14]: EE
- Comentado [MAP15]: CLT
- Comentado [MAP16]: CLT
- Comentado [MAP17]: CG10
- Comentado [MAP18]: CO
- Comentado [MAP19]: CG3
- Comentado [MAP20]: CG1
- Comentado [MAP21]: CG10
- Comentado [MAP22]: CE
- Comentado [MAP23]: CG10
- Comentado [MAP24]: CG1
- Comentado [MAP25]: CG10
- Comentado [MAP26]: CG2

INTRODUCCIÓN

Los coronavirus son una familia de virus que infectan a animales, principalmente aves y mamíferos, en los que causan enfermedades y producen pérdidas económicas asociadas a las actividades productivas. También pueden saltar de ciertas especies e infectar a los seres humanos y causar enfermedades de distinta gravedad (resfriado, bronquitis, neumonía, etc.), según el virus y la salud del huésped. La mayoría de las personas se infectan con estos virus alguna vez en la vida sin padecer grandes consecuencias. No obstante, algunos coronavirus han causado enfermedades que se han diseminado rápidamente y han provocado brotes y epidemias preocupantes. Como ejemplos se pueden citar el síndrome respiratorio agudo grave (SARS) originado en Yunan, China, en 2001; el síndrome respiratorio de Oriente Medio (MERS) causado por el MERS-CoV, iniciado en Arabia Saudita en 2012 y, actualmente, el COVID-19, causado por un nuevo coronavirus, denominado SARS-CoV-2.

Se cree que la COVID-19 surgió en diciembre de 2019 en un mercado de mariscos donde se vendían animales exóticos vivos y muertos ubicado en Wuhan, provincia de Hubei, China. Sin embargo, hay algunas controversias acerca de esta teoría. La velocidad de diseminación del virus fue tan elevada que, en poco tiempo, las autoridades de China cerraron la ciudad y dispusieron cuarentena obligatoria. Incluso con estas medidas, el contagio se extendió con rapidez por los países limítrofes primero y, paulatinamente, por los demás, facilitado por la gran movilidad de viajeros. Casi tres meses después de detectar la enfermedad, el 11 de marzo de 2020, la Organización Mundial de la Salud (OMS) la declaró como pandemia.

El rápido avance de la enfermedad puso en evidencia la falta de preparación de los sistemas de salud de la mayoría de los países para afrontar la situación. El número de enfermos sobrepasó con creces las capacidades disponibles y el mundo observó con estupor la creciente cifra de muertos en países desarrollados de Europa. Ante ese panorama y con el objeto de contener el contagio, en la mayoría de los países se establecieron medidas de distanciamiento social (más conocidas como cuarentena), con distintas modalidades. A la fecha de escritura de este artículo (19 de mayo de 2020), el número de casos confirmados en el mundo se acercaba rápidamente a los cinco millones, de los cuales aproximadamente el 7% correspondía a fallecimientos. A fecha de corrección del manuscrito, 15 de septiembre de 2020, hay casi 30 millones de casos confirmados en el mundo y el número de fallecimientos se acerca al millón.

- Comentado [MAP27]: EE
- Comentado [MAP28]: EE
- Comentado [MAP29]: CG5
- Comentado [MAP30]: CG6
- Comentado [MAP31]: EE
- Comentado [MAP32]: EE
- Comentado [MAP33]: EE
- Comentado [MAP34]: CL1
- Comentado [MAP35]: CG9
- Comentado [MAP36]: CG1
- Comentado [MAP37]: EE
- Comentado [MAP38]: CG1
- Comentado [MAP39]: EE
- Comentado [MAP40]: CO
- Comentado [MAP41]: EE
- Comentado [MAP42]: CG1
- Comentado [MAP43]: EE
- Comentado [MAP44]: CG5
- Comentado [MAP45]: CG5
- Comentado [MAP46]: EE
- Comentado [MAP47]: CO
- Comentado [MAP48]: EE
- Comentado [MAP49]: CG7
- Comentado [MAP50]: EE
- Comentado [MAP51]: CG7
- Comentado [MAP52]: CO
- Comentado [MAP53]: CO
- Comentado [MAP54]: EE
- Comentado [MAP55]: EE
- Comentado [MAP56]: EE
- Comentado [MAP57]: CG1

De manera muy temprana (20 de marzo de 2020), el gobierno argentino estableció el aislamiento social preventivo y obligatorio (ASPO). Esta medida permitió evitar el contagio masivo y ganar tiempo para que el sistema de salud pudiera mejorar su capacidad de respuesta. Así, se montaron nuevos hospitales o centros de atención y se adquirieron y fabricaron respiradores, elementos de protección personal y demás productos para poder atender la emergencia. También se instó a la comunidad científica a contribuir en la generación de conocimientos para afrontar la pandemia. Como en el resto del mundo, la detección del SARS-CoV-2 se realiza en Argentina mediante pruebas de reacción en cadena de la polimerasa en tiempo real con retrotranscripción (RT-qPCR). En principio, esta prueba se le realiza a cada individuo sospechoso de ser portador debido a que presenta síntomas relacionados con la enfermedad o por conexión epidemiológica. Pero estos protocolos han ido cambiando con la evolución de la pandemia. Otras pruebas moleculares recientemente desarrollados por investigadores argentinos permiten detectar la presencia de fragmentos genéticos con resultados prometedores: estos son el NEOKIT y el ELA-CHEMSTRIP. La ventaja de estos últimos métodos es que permiten identificar al virus de manera más rápida y posibilitan su detección en un grupo poblacional más grande. Sin embargo, en la mayoría de los países en desarrollo la capacidad para realizar dichos análisis es reducida. No se puede evaluar a la población que se supone sana o asintomática (aunque puede incluir individuos contagiosos) y se deben tomar medidas generalizadas para controlar la diseminación del virus.

- Comentado [MAP58]: CO
- Comentado [MAP59]: CG5
- Comentado [MAP60]: EE
- Comentado [MAP61]: CE
- Comentado [MAP62]: CLT
- Comentado [MAP63]: CG1
- Comentado [MAP64]: EE
- Comentado [MAP65]: CLT
- Comentado [MAP66]: CLT
- Comentado [MAP67]: EE
- Comentado [MAP68]: CG6
- Comentado [MAP69]: CG1
- Comentado [MAP70]: CG1
- Comentado [MAP71]: EE
- Comentado [MAP72]: CG5
- Comentado [MAP73]: CLT ANGLICISMO LÉXICO
- Comentado [MAP74]: CG6
- Comentado [MAP75]: EE
- Comentado [MAP76]: CG7
- Comentado [MAP77]: CG3
- Comentado [MAP78]: CG5
- Comentado [MAP79]: CG6
- Comentado [MAP80]: CG5
- Comentado [MAP81]: CG1
- Comentado [MAP82]: CG5
- Comentado [MAP83]: EE
- Comentado [MAP84]: CG1

Debido al contexto sanitario actual y a que el distanciamiento social ha resultado efectivo para ralentizar la diseminación de la enfermedad, resulta necesario volver paulatina y controladamente a la actividad. No obstante, en caso de deteriorarse la situación sanitaria, las medidas de flexibilización podrían ser revertidas. El objetivo del presente trabajo fue proponer un algoritmo de decisión para optimizar la detección de casos asintomáticos y prescribir la cuarentena de una manera estratégica, prevenir la diseminación del virus y permitir una nueva normalidad supervisada.

MATERIALES Y MÉTODOS

Elaboramos una herramienta tentativa para utilizar estratégicamente las pruebas de detección del virus SARS-CoV-2 y controlar la diseminación de la enfermedad en múltiples cohortes. Con este fin, como ya fue indicado por otros autores, sugerimos analizar muestras compuestas (análisis grupal) y combinar las muestras individuales de distintas personas asintomáticas de la cohorte de interés. Esta metodología permitiría evaluar la presencia del virus en mayor cantidad de personas simultáneamente, con un ahorro considerable de recursos y de tiempo. Propusimos analizar en primer lugar las cohortes de personas cuya función fuera crítica para el funcionamiento normal de la sociedad o pertenecientes a sectores muy vulnerables. El análisis continuaría con las cohortes no críticas, según su importancia y grado de vulnerabilidad.

- Comentado [MAP85]: EE
- Comentado [MAP86]: CG10
- Comentado [MAP87]: CG6
- Comentado [MAP88]: CG10
- Comentado [MAP89]: EE
- Comentado [MAP90]: CG1
- Comentado [MAP91]: CG10
- Comentado [MAP92]: CG5
- Comentado [MAP93]: CG6
- Comentado [MAP94]: EE
- Comentado [MAP95]: CLT
- Comentado [MAP96]: CG5
- Comentado [MAP97]: CLT
- Comentado [MAP98]: CG1
- Comentado [MAP99]: CG10
- Comentado [MAP100]: CG3
- Comentado [MAP101]: CG5
- Comentado [MAP102]: CG10
- Comentado [MAP103]: CG9
- Comentado [MAP104]: EE
- Comentado [MAP105]: CG3
- Comentado [MAP106]: CG5
- Comentado [MAP107]: EE
- Comentado [MAP108]: CLT
- Comentado [MAP109]: CG10
- Comentado [MAP110]: CG6
- Comentado [MAP111]: EE
- Comentado [MAP112]: CG7
- Comentado [MAP113]: CE
- Comentado [MAP114]: CG1
- Comentado [MAP115]: EE
- Comentado [MAP116]: CE

Definición de cohortes críticas

Las cohortes que aquí proponemos como críticas y que deberían ser las primeras en someterse a la detección del virus abarcarían los siguientes rubros: salud, seguridad, comercio de bienes de consumo esenciales, tercera edad (geriátricos), servicios e industrias esenciales (fabricación de respiradores, de alimentos, etc.). Es fundamental estudiar a los miembros de estos sectores, al ser potenciales vectores de contagio. El personal sanitario (médicos, enfermeros, personal de limpieza, etc.) debería mantenerse en óptimas condiciones y, además de proveerles elementos de protección personal, deberían recibir análisis de manera sistemática y periódica. Estos protocolos son necesarios también para el personal de seguridad, ya que está muy expuesto a personas potencialmente infectadas. Los comerciantes de la cadena de suministros, incluidos los grandes mercados de frutas y verduras, deberían recibir análisis regularmente también. En este ámbito, una persona infectada podría hacer que se paralice toda la cadena y se generen graves problemas sociales. Otros sectores, como el de los trabajadores de la construcción, administrativos, comerciantes de elementos no esenciales o habitantes de zonas residenciales, por mencionar algunos, corresponden a sectores no críticos. Estos grupos se analizarían cuando la cantidad de pruebas disponibles fuera suficiente, sin comprometer el control de los sectores que son críticos. Sin embargo, aquellas cohortes no críticas que tengan relación con integrantes de cohortes críticas deberían tener mayor prioridad de análisis.

Comentado [MAP117]: CG10

Comentado [MAP118]: CLT

Comentado [MAP119]: EE

Comentado [MAP120]: CG10

Comentado [MAP121]: CG1

Comentado [MAP122]: CG5

Comentado [MAP123]: CG9

Comentado [MAP124]: CG10

Comentado [MAP125]: CG1

Comentado [MAP126]: EE

Comentado [MAP127]: CE

Comentado [MAP128]: CE

Comentado [MAP129]: CG1

Comentado [MAP130]: CG5

Comentado [MAP131]: CG4

Realización de los análisis

Luego de definir los grupos de las distintas cohortes, proponemos tomar muestras individuales (hisopados nasofaríngeos) a todas las personas asintomáticas de dichos grupos. Una parte de la muestra podría ser mezclada en el laboratorio para conformar una muestra compuesta de cada grupo para su análisis. Inicialmente, deberíamos estimar la tasa de infectividad y la prevalencia a partir de los datos oficiales de contagio en el lugar geográfico analizado. Estos valores permitirían definir el tamaño del grupo y la metodología de detección por grupos.

El análisis por grupos de muestras tiene sentido como ahorro de recursos y de tiempo, siempre y cuando la prevalencia sea inferior al 10 %. A partir de este valor conviene realizar los análisis individuales debido a la alta cantidad de grupos que darían positivo (sería necesario el estudio de las muestras individuales). El tamaño de los grupos estaría definido por la prevalencia: a mayor prevalencia, menor será el número de muestras de un grupo. Consideramos la prevalencia de la enfermedad en la población y también en las cohortes conformadas por personal esencial como información epidemiológica conocida. La prevalencia podría calcularse a partir de los casos (individuales) confirmados en los grupos de interés en relación con la población o un tamaño definido de la población.

Comentado [MAP132]: CLT ANGLICISMO LÉXICO

Comentado [MAP133]: CG10

Comentado [MAP134]: CG5

Comentado [MAP135]: CG10

Comentado [MAP136]: CLT

Comentado [MAP137]: CE

Comentado [MAP138]: CE

Comentado [MAP139]: CG1

Comentado [MAP140]: EE

Comentado [MAP141]: CLT ANGLICISMO LÉXICO

Comentado [MAP142]: CLT ANGLICISMO LÉXICO

Comentado [MAP143]: CLT

Comentado [MAP144]: CE

Comentado [MAP145]: CG1

Comentado [MAP146]: CO

Comentado [MAP147]: CG5

Comentado [MAP148]: CLT ANGLICISMO LÉXICO

Comentado [MAP149]: CLT ANGLICISMO LÉXICO

Comentado [MAP150]: CE

Comentado [MAP151]: CLT ANGLICISMO LÉXICO

Comentado [MAP152]: EE

Comentado [MAP153]: CLT ANGLICISMO LÉXICO

Comentado [MAP154]: EE

Comentado [MAP155]: CG5

Comentado [MAP156]: CG3

Comentado [MAP157]: CLT ANGLICISMO LÉXICO

Comentado [MAP158]: CG10

Comentado [MAP159]: EE

Comentado [MAP160]: CG1

Comentado [MAP161]: EE

Cabe aclarar que, antes de iniciar la detección mediante grupos de muestras, habría que evaluar algunos aspectos metodológicos críticos para asegurar el éxito de esta modalidad diagnóstica. La primera cuestión a verificar a través de la bioinformática es que las secuencias de los oligonucleótidos para la detección sean específicas. Luego, correspondería determinar la eficiencia de la amplificación de la reacción de RT-qPCR y verificar que la eficiencia sea máxima, aun a bajas concentraciones de ARN. Además, sería imprescindible determinar el límite de detección del método empleado, ya que esta variable afecta a la clasificación de los resultados como “positivos” y “no detectables (negativos)”. Finalmente, deberíamos realizar una validación de la técnica de grupos en un grupo de la población y comparar los resultados con los que se obtienen por análisis individuales. Esta comparación permite determinar la probabilidad de clasificación errónea de un resultado. Recién al contar con todo el conocimiento mencionado podríamos empezar con la detección en grupos de muestras.

El número máximo de individuos de un grupo debería definirse según el volumen por muestra individual a agregar a la reacción de detección sin comprometer la sensibilidad, pero permitiendo suficiente carga viral para lograrla. Además, debemos considerar la prevalencia del virus o la relación entre la cantidad de individuos infectados y la población total analizada para optimizar el tamaño de los grupos desde un punto de vista estadístico. Estos valores servirían de guía, pero la definición de los grupos debería realizarse de acuerdo con la conveniencia práctica. Si el tamaño óptimo de grupo fuera 15 personas y en un geriátrico hubiera 10 personas, no sería necesario ni conveniente sumar gente externa para llegar al tamaño óptimo. Por otro lado, al analizar la cohorte de un barrio completo, los grupos deberían tener el tamaño óptimo y ser seleccionados según su proximidad para ahorrar recursos.

- Comentado [MAP162]: CO
- Comentado [MAP163]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP164]: CO
- Comentado [MAP165]: CG5
- Comentado [MAP166]: CG10
- Comentado [MAP167]: CG10
- Comentado [MAP168]: CE
- Comentado [MAP169]: CG5
- Comentado [MAP170]: CG10
- Comentado [MAP171]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP172]: CG5
- Comentado [MAP173]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP174]: CG1
- Comentado [MAP175]: EE
- Comentado [MAP176]: CE
- Comentado [MAP177]: CLT
- Comentado [MAP178]: CG10
- Comentado [MAP179]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP180]: EE
- Comentado [MAP181]: CG6
- Comentado [MAP182]: CG10
- Comentado [MAP183]: CG1
- Comentado [MAP184]: CG10
- Comentado [MAP185]: CG6
- Comentado [MAP186]: EE
- Comentado [MAP187]: CE
- Comentado [MAP188]: CG1
- Comentado [MAP189]: CG10
- Comentado [MAP190]: CG6
- Comentado [MAP191]: CG1
- Comentado [MAP192]: EE

Los grupos con resultados negativos en pruebas de SARS-CoV-2 se deberían analizar cada cierto tiempo, optimizado según un algoritmo que resuelva el orden de prioridad de cada grupo. De resultar positiva alguna de las pruebas realizadas, se debería aislar al grupo y a las personas con las que hayan tenido contacto estrecho para evitar contagios hasta completar el análisis. El laboratorio deberá estudiar los grupos positivos y profundizar en el análisis de las muestras individuales para identificar al individuo infectado, quien luego debería permanecer aislado (fig. 1). Las personas con resultados individuales negativos podrían salir del aislamiento.

Definir las cohortes a analizar

Tomar muestras de los individuos de la cohorte y combinarlas en una muestra grupal

Realizar análisis de muestra grupal

Calcular tiempo a esperar según prioridad de cohorte

Negativo

Resultado del análisis del Grupo

Positivo

Mantener alerta y distanciamiento por 2 semanas

Analizar las muestras individualmente o por subgrupos

Negativo

Resultado del análisis individual

Positivo

Figura 1. Diagrama de flujo que indica los pasos a seguir para mitigar la diseminación de la enfermedad en la población y optimizar el uso de los recursos de diagnóstico.

Optimización de recursos

Desarrollamos un algoritmo de decisión para optimizar el análisis y establecer un orden lógico para su realización. El algoritmo permite modelar y evaluar el comportamiento de la transmisión de la enfermedad en función del impacto de las cohortes definidas sobre la salud, la seguridad y la economía. Idealmente, es necesario un conocimiento profundo de los integrantes de las cohortes de análisis, por lo que la implementación de este algoritmo a nivel municipal podría resultar más beneficiosa.

Consideramos necesario, aunque no excluyente, incluir las siguientes variables generales en el modelo, ya que afectan a todos los grupos de todas las cohortes, y estas variables deberían actualizarse diariamente:

Comentado [MAP193]: EE

Comentado [MAP194]: CG5

Comentado [MAP195]: CE

Comentado [MAP196]: EE

Comentado [MAP197]: CO

Comentado [MAP198]: CG1

Comentado [MAP199]: CG5

Comentado [MAP200]: EE

Comentado [MAP201]: CLT. ANGLICISMO LÉXICO

Comentado [MAP202]: CLT. ANGLICISMO LÉXICO

Comentado [MAP203]: CG4

Comentado [MAP204]: CO

Comentado [MAP205]: CE

Comentado [MAP206]: CLT

Comentado [MAP207]: CG10

Comentado [MAP208]: CLT. ANGLICISMO LÉXICO

Comentado [MAP209]: CG9

Comentado [MAP210]: CG1

Comentado [MAP211]: EE

Comentado [MAP212]: CO

Comentado [MAP213]: CG3

Comentado [MAP214]: CG10

Comentado [MAP215]: CG10

Comentado [MAP216]: CG10

Comentado [MAP217]: CG4

Cantidad de pruebas disponibles: condiciona la cantidad de grupos a los cuales se les puede realizar el análisis y, por tanto, la cantidad de cohortes que pueden ser analizadas periódicamente.

Función de contagio general: definida como la relación entre la población susceptible de contagiarse, la población infectada y la población recuperada. Esta variable se aplica en una zona que abarca a las cohortes analizadas y depende de los valores históricos de esas variables.

Otras variables importantes están relacionadas con características particulares del grupo a analizar:

Ubicación/coordenadas: permite conocer la localización del grupo y puede ser útil para determinar el nivel de contacto intergrupual o ayudar a decidir la vulnerabilidad al contagio por la actividad. Por ejemplo, podría existir mayor riesgo de contagio para un grupo de seguridad que comparta la ubicación con un grupo de un hospital donde haya individuos positivos para SARS-CoV-2.

Grado de prioridad del grupo (A): alude a la prioridad de un grupo sobre otros para realizar el análisis. Por ejemplo, los grupos de la cohorte del sector sanitario (médicos y enfermeros dedicados al tratamiento del SARS-CoV-2) tendrían prioridad de análisis sobre un grupo de una cohorte de personas dedicadas a la producción de bienes de consumo. Para que esta decisión sea lo menos subjetiva posible, definimos una escala de 1 a 10, en la que 1 significa “nada prioritario” y 10 indica “máxima prioridad” (tabla 1). Establecimos rangos específicos para cada grupo hipotético incluido en el algoritmo.

Tabla 1. Escala de prioridad de grupos para el análisis

Prioridad del grupo (A)

Rango

Médicos dedicados al tratamiento directo de infectados por SARS-CoV-2

Otros médicos

Población de un geriátrico

Personal de seguridad

Comerciantes de bienes de consumo

Industria

Grupos residenciales

Número de integrantes de cada grupo (B): la cantidad de integrantes de cada grupo se definiría de acuerdo con la conveniencia. Los negativos del análisis deben ser negativos verdaderos (que no comprometan el límite de detección de la muestra), para evitar la diseminación la infección. Si el análisis de un grupo infectado se hace por RT-qPCR, la sensibilidad dependerá de la fase de infección en curso, de la calidad de las muestras extraídas y de la cantidad de muestras individuales de la muestra compuesta (a mayor número de muestras, menor cantidad de cada blanco y menor sensibilidad).

Contacto intragrupal (C): esta variable indica el nivel de contacto (relacionado con la distancia social y la frecuencia del contacto) entre los miembros del grupo analizado. Para su valoración establecimos una escala entre 1 y 10, donde valores altos indican mayor contacto y valores bajos menor contacto (al contacto nulo le adjudicamos el valor 1). Los valores de contacto para diferentes distancias pueden coincidir a causa de un cambio en la frecuencia de contacto (tabla 2).

Tabla 2. Escala para el contacto intragrupal (C)

Distancia (m)

Frecuencia

Rango

Baja

Media

Alta

- Comentado [MAP218]: EE
- Comentado [MAP219]: CG1
- Comentado [MAP220]: CG1
- Comentado [MAP221]: CG4
- Comentado [MAP222]: CG4
- Comentado [MAP223]: CG4
- Comentado [MAP224]: CG5
- Comentado [MAP225]: EE
- Comentado [MAP226]: CG1
- Comentado [MAP227]: CE
- Comentado [MAP228]: EE
- Comentado [MAP229]: CLT
- Comentado [MAP230]: CG10
- Comentado [MAP231]: CE
- Comentado [MAP232]: EE
- Comentado [MAP233]: CG1
- Comentado [MAP234]: CG10
- Comentado [MAP235]: CO
- Comentado [MAP236]: CO
- Comentado [MAP237]: CG10
- Comentado [MAP238]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP239]: CLT
- Comentado [MAP240]: CG1
- Comentado [MAP241]: EE
- Comentado [MAP242]: CG7
- Comentado [MAP243]: CG3
- Comentado [MAP244]: EE
- Comentado [MAP245]: CG1
- Comentado [MAP246]: CG2
- Comentado [MAP247]: CG10
- Comentado [MAP248]: CG10
- Comentado [MAP249]: CG1
- Comentado [MAP250]: CE

Baja
Media
Alta
Baja
Media
Alta

Vulnerabilidad al contagio por la actividad (D): hace referencia al grado de vulnerabilidad de los grupos estudiados dentro de una cohorte a causa de la actividad. Establecimos también en este caso una escala entre 1 y 10, en la que un mayor valor indica mayor nivel de vulnerabilidad al contagio por la actividad (tabla 3). Los valores se podrían obtener a partir de estudios del nivel de contagio en estos grupos y deberían ser actualizados periódicamente.

Tabla 3. Escala para la vulnerabilidad al contagio (D)

Actividad

Valor

Médicos de familia

Enfermeros especializados

Técnicos de farmacia

Vigilante de prisión

Residenciales

Vulnerabilidad por el tiempo transcurrido (E): valor calculado (ecuación 1) entre 0 y 10, que indica el riesgo de que exista una infección no detectada en el tiempo transcurrido.

donde t (en días) es el tiempo transcurrido desde la última prueba realizada y t_i es el período de incubación máximo del virus (14 días).

Activación intergrupala (F): valor empleado para indicar si aumentará (o no) el orden de prioridad como consecuencia del contacto con un grupo cuyo análisis dio positivo. Es decir, a causa de la relación intergrupala con grupos infectados, el grupo analizado aumentará su orden de prioridad y, por lo tanto, podría conllevar que sea analizado primero. F tomará el valor de 1 si la prueba fuera positiva y el valor de 0 si resultara negativa (no se activa la relación intergrupala).

Contacto intergrupala (G): variable que establece el nivel de contacto entre grupos distintos (de una misma cohorte o de una cohorte diferente). Definimos una escala entre 1 y 10 (al contacto nulo le adjudicamos el valor de 1). Los valores altos indican un contacto más estrecho entre los miembros de distintos grupos (la relación del grupo i con el grupo j se define como relación ij , con $i \neq j$). Los valores de contacto para diferentes distancias pueden coincidir a causa de un cambio en la frecuencia de contacto (tabla 4).

Comentado [MAP251]: CG10

Comentado [MAP252]: CE

Comentado [MAP253]: CO

Comentado [MAP254]: CG10

Comentado [MAP255]: CLT

Comentado [MAP256]: CO

Comentado [MAP257]: CLT

Comentado [MAP258]: CG5

Comentado [MAP259]: CG4

Comentado [MAP260]: CG4

Comentado [MAP261]: CG10

Comentado [MAP262]: CG10

Comentado [MAP263]: CG1

Comentado [MAP264]: CO

Comentado [MAP265]: EE

Comentado [MAP266]: CE

Tabla 4. Escala para contacto intergrupales (G)

Distancia (m)

Frecuencia

Rango

Baja

Media

Alta

Baja

Media

Alta

Baja

Media

Alta

Orden de análisis general (O): valor cuyo resultado permitiría ordenar la prioridad de realización de las pruebas según el modelo actual. Primero, teniendo en cuenta las variables antes establecidas, calculamos un valor local para cada grupo (L_i , ecuación 2) e identificamos el mayor valor ($L_{máx}$). Los valores locales representarían la importancia en el análisis y no tienen una escala definida. Luego, para visualizar mejor cuáles grupos deben analizarse primero, ponderaremos los valores locales respecto al valor local máximo ($L_{máx}$) para obtener el orden de prioridad de análisis general para cada grupo (O_i , ecuación 3).

donde j varía desde 1 hasta j máximo (cantidad total de grupos definidos), con $i \neq j$:

Una vez definido el orden de análisis general de los grupos (O), podremos realizar la detección siguiendo dicho orden (de mayor a menor) y actualizar la cantidad de pruebas disponibles. Luego, retroalimentamos la información obtenida al algoritmo con las variables E (vulnerabilidad por el tiempo transcurrido) y F (activación intergrupales) (fig. 2).

GRUPO j

GRUPO i

(G) Contacto intergrupales

Ordenar análisis

VARIABLES DE GRUPO

Ubicación / coordenadas

(A) Prioridad del grupo

(B) Número de integrantes

(C) Contacto intragrupal

(D) Vulnerabilidad al contagio por la actividad

(O) Orden de análisis

Analizar grupos

VARIABLES DE RETROALIMENTACIÓN

(E) Vulnerabilidad por el tiempo transcurrido

(F) Activación intergrupales

VARIABLES GENERALES

Cantidad de pruebas disponibles

Función de contagio general

Comentado [MAP267]: CG10

Comentado [MAP268]: CG10

Comentado [MAP269]: CG10

Comentado [MAP270]: CE

Comentado [MAP271]: CG10

Comentado [MAP272]: CG10

Comentado [MAP273]: CG10

Comentado [MAP274]: CG10

Comentado [MAP275]: CE

Comentado [MAP276]: CE

Comentado [MAP277]: CE

Comentado [MAP278]: CE

Figura 2. Algoritmo general para calcular el orden de análisis de los grupos y hacer un uso racional de las pruebas de detección disponibles. Los recuadros blancos indican las variables de entrada y de salida del modelo, mientras que los bloques azules son las operaciones.

Alcance temporal y costos

Según la OMS, se espera que esta pandemia se extienda por lo menos por dos años. Aunque existen muchos proyectos de vacunas y tratamientos paliativos, todavía ninguno se ha implementado de forma eficaz y masiva, por lo que la duración de la pandemia es impredecible. El tiempo de aplicación de la herramienta propuesta en este trabajo es difícil de establecer, pues depende de muchas variables y está íntimamente relacionado con la duración de la pandemia. La herramienta es muy flexible: permite operar con diferentes cantidades de cohortes (según los requerimientos o el presupuesto) y permite hacer los análisis a la frecuencia conveniente para cada localidad. El costo de implementación dependerá, principalmente, de la capacidad de análisis disponible. Otra variable que podría influir en los costos es la posible implementación de pruebas rápidas, que estarán disponibles (principalmente para el sistema sanitario) en los próximos meses. Es importante remarcar que el análisis de muestras agrupadas permite obtener una gran cantidad de información y minimizar los costos al analizar numerosas muestras individuales en una sola reacción.

- Comentado [MAP279]: CG10
- Comentado [MAP280]: CE
- Comentado [MAP281]: EE
- Comentado [MAP282]: CG1
- Comentado [MAP283]: EE
- Comentado [MAP284]: CG1
- Comentado [MAP285]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP286]: CG1
- Comentado [MAP287]: EE
- Comentado [MAP288]: CG5
- Comentado [MAP289]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP290]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP291]: EE
- Comentado [MAP292]: CLT
- Comentado [MAP293]: CG1

RESULTADOS Y DISCUSIÓN

Para ilustrar la aplicación del algoritmo y evaluar su versatilidad y factibilidad, realizamos una simulación. Ya que no contábamos con datos reales al formular este algoritmo, fue necesario definir algunos datos de manera arbitraria (intentando acercarnos a la realidad) y realizar diferentes suposiciones.

En primer lugar, definimos ocho grupos hipotéticos con fines ilustrativos, ya que en la realidad serán muchos más: GRUPO 1, médicos pertenecientes al hospital 1 (m1); GRUPO 2, personal de seguridad del hospital 1 (s1); GRUPO 3, médicos del hospital 2 (m2); GRUPO 4, personal de seguridad del hospital 2 (s2); GRUPO 5, población de un geriátrico (ancianos y personal) que se encuentra en una zona urbana (g1), bajo el supuesto de que no había casos en la zona; GRUPO 6, personal del comercio mayorista 1 de bienes de consumo, ubicado en una zona donde existen otros comercios mayoristas (c1); GRUPO 7, industria de fabricación de bienes ubicada en una zona en donde sí había casos de COVID-19 (i1); GRUPO 8, familia numerosa de un barrio residencial, ubicado en una zona urbana donde no había casos de COVID-19 (r1). Establecimos una cantidad específica de integrantes para cada grupo (tablas 5 y 6).

Tabla 5. Ejemplo hipotético de aplicación del algoritmo de decisión desarrollado

Grupos

Prioridad del grupo (A)

Número de integrantes de cada grupo (B)

Contacto intragrupal (C)

Vulnerabilidad al contagio por la actividad (D)

Días transcurridos desde el análisis

Vulnerabilidad por tiempo transcurrido (E)^a

Orden de análisis local (L)^b

Orden de análisis general (O)^c

Orden de prioridad de pruebas

^{a,b,c} Calculados mediante las ecuaciones 1, 2 y 3.

Grupos definidos. m1: médicos del hospital 1; s1: seguridad del hospital 1; m2: médicos del hospital 2; s2: seguridad del hospital 2; g1: población del geriátrico 1; c1: personal del comercio 1; i1: personal de la industria 1; r1: vecinos de la zona residencial 1. El grupo supuestamente infectado está marcado en negrita.

Tabla 6. Ejemplo hipotético para ilustrar una matriz de relaciones intergrupales

Grupos

F: activación intergrupala (1: infectado, 0: no infectado); NA: el contacto intergrupala no es una variable aplicable a la relación de un grupo consigo mismo.

Grupos definidos. m1: médicos del hospital 1; s1: seguridad del hospital 1; m2: médicos del hospital 2; s2: seguridad del hospital 2.

El grupo supuestamente infectado está marcado en negrita.

Con respecto a la variable “prioridad del grupo (A)”, asignamos un valor de 10 a los médicos del hospital 1 (m1), que registra mayor cantidad de pacientes con COVID-19, y 9 a los médicos del hospital 2 (m2). Los altos valores asignados se deben a que ambos son grupos prioritarios (pertenecen a la cohorte “médicos de hospitales”) cuya función es atender a personas con COVID-19.

Para los grupos de la cohorte “seguridad” establecimos un valor de 6 tanto para s1 como para s2. Aunque este grupo es personal esencial, podrían reemplazarse si sucediese un contagio. La población del geriátrico g1 es un grupo de alto riesgo debido a que la mortalidad por contagio es

Comentado [MAP294]: CG6

Comentado [MAP295]: CG10

Comentado [MAP296]: EE

Comentado [MAP297]: CG1

Comentado [MAP298]: CG10

Comentado [MAP299]: CG5

Comentado [MAP300]: CG10

Comentado [MAP301]: EE

Comentado [MAP302]: CLT

Comentado [MAP303]: CG10

Comentado [MAP304]: CLT

Comentado [MAP305]: CG10

Comentado [MAP306]: CG10

Comentado [MAP307]: CG10

Comentado [MAP308]: CG2

Comentado [MAP309]: CLT

Comentado [MAP310]: EE

Comentado [MAP311]: CO

Comentado [MAP312]: CG10

Comentado [MAP313]: CG2

Comentado [MAP314]: CE

Comentado [MAP315]: CG2

Comentado [MAP316]: EE

Comentado [MAP317]: CG10

Comentado [MAP318]: CG2

Comentado [MAP319]: CO

Comentado [MAP320]: CG10

Comentado [MAP321]: CO

Comentado [MAP322]: CG4

Comentado [MAP323]: CG6

Comentado [MAP324]: EE

Comentado [MAP325]: CG1

Comentado [MAP326]: CO

Comentado [MAP327]: CG10

Comentado [MAP328]: CG1

Comentado [MAP329]: EE

Comentado [MAP330]: CG4

Comentado [MAP331]: CG3

Comentado [MAP332]: CLT

Comentado [MAP333]: CG10

Comentado [MAP334]: CLT

Comentado [MAP335]: EE

Comentado [MAP336]: CO

mucho mayor, por lo que le adjudicamos un valor de 8. El comercio mayorista de bienes de consumo c1 tiene una prioridad de 4 (relativamente baja), ya que existen otros comercios mayoristas alternativos en la zona. Para la industria i1 el valor asignado fue 7, porque se encuentra en una zona de riesgo entre moderado y alto. Al grupo residencial r1 le asignamos un valor de 3, ya que se encuentra en una zona urbana con tránsito moderado, donde no había casos. Para la variable “vulnerabilidad al contagio por la actividad (D)”, el hospital h1 es uno de los centros designados para tratar a personas con coronavirus, mientras que el hospital h2 no lo es. Por lo tanto, adjudicamos un puntaje de 10 a m1 y de 8 a m2 para ponderar el riesgo de contagio. Por la misma razón, establecimos el riesgo de contagio de s1 en 6 y el de s2 en 4. Aunque la actividad de g1 no es riesgosa, sí se trata de un grupo muy vulnerable y le asignamos un valor de 5. El grupo c1 se encuentra en una zona libre de COVID-19 y tiene un valor de 6 debido a que existe un flujo considerable de personas que no corresponden a la zona y un contacto cercano con los clientes. Finalmente, el grupo i1 desarrolla una actividad en contacto mínimo con personas externas, por lo tanto le asignamos un valor de 3.

- Comentado [MAP337]:** CG1
- Comentado [MAP338]:** CG10
- Comentado [MAP339]:** CO
- Comentado [MAP340]:** CG10
- Comentado [MAP341]:** CG10
- Comentado [MAP342]:** CO
- Comentado [MAP343]:** CG4
- Comentado [MAP344]:** CG10
- Comentado [MAP345]:** CG10
- Comentado [MAP346]:** CG3
- Comentado [MAP347]:** CG10
- Comentado [MAP348]:** CE
- Comentado [MAP349]:** CG10

En cuanto a la variable “contacto intragrupal (C)”, los médicos m1 tienen reuniones diarias en el hospital 1 y con una frecuencia media trabajan en conjunto y a distancias menores de 2 metros. Esta suposición es válida también para los médicos m2 del hospital 2. Por esta razón, puntuamos el contacto intragrupal para ambos casos con 8. El contacto entre los individuos de los grupos de seguridad (s1 y s2) se da, principalmente, a una distancia menor de 2 metros y con baja frecuencia, en situaciones relacionadas con mantener el orden. Por todo ello, le asignamos un valor de 6. Los ancianos y el personal del geriátrico (g1) se reúnen a distancias entre 1 y 3 metros con una frecuencia alta, por lo que asignamos un valor de 9. El personal tanto de c1 como de i1 se reúne a frecuencia y distancia medias, por lo que asignamos un valor de 6 en ambos casos. Finalmente, el grupo residencial r1 se reúne a distancias bajas y con frecuencia alta, por lo que le asignamos un valor de 10. Para terminar con las variables, el “contacto intergrupar (G)” de los médicos m1 y m2 es a frecuencia baja y a distancias mayores de 4 metros. El contacto ocurre debido a solicitudes de participación especial o en cursos de actualización, por lo que el valor para la relación intergrupar en este caso es 2. Esta suposición es válida también para la relación entre los grupos de seguridad s1 y s2. El resto de las relaciones intergrupales es inexistente, por lo que establecimos su valor en 1 (tabla 5).

- Comentado [MAP350]: CG9
- Comentado [MAP351]: CO
- Comentado [MAP352]: CG5
- Comentado [MAP353]: CG10
- Comentado [MAP354]: CG2
- Comentado [MAP355]: CG1
- Comentado [MAP356]: EE
- Comentado [MAP357]: CG10
- Comentado [MAP358]: CE
- Comentado [MAP359]: CG1
- Comentado [MAP360]: CG10
- Comentado [MAP361]: CE
- Comentado [MAP362]: CG10
- Comentado [MAP363]: CG2
- Comentado [MAP364]: CG10
- Comentado [MAP365]: CG9
- Comentado [MAP366]: CO
- Comentado [MAP367]: CG1
- Comentado [MAP368]: EE
- Comentado [MAP369]: CG10

Un positivo en el grupo s1 en el último análisis realizado 6 días antes (tabla 5) tendría dos consecuencias directas. En primer lugar, ese grupo debería enviarse a aislamiento preventivo hasta identificar al individuo infectado (o a los individuos infectados). Por tanto, sería necesario analizar las muestras individuales y luego proceder a los correspondientes reemplazos laborales (fig. 1). En segundo lugar, esto ocasionaría un cambio en el orden de análisis para los otros grupos debido al contacto intergrupar. En este ejemplo no analizamos al grupo de seguridad s1 por subgrupos, ya que hay prevista suficiente disponibilidad de reemplazos hasta que s1 vuelva de la cuarentena. No obstante, de realizarse el análisis, deberían seguirse los pasos establecidos (fig. 1). Podemos observar, entonces, que sería necesario realizar el análisis en primer lugar a los médicos del hospital 1 (m1; $O_i = 100$), seguido por el grupo del geriátrico 1 (g1; $O_i = 67$), los médicos del hospital 2 (m2; $O_i = 61$), la seguridad del hospital 2 (s2; $O_i = 33$), el comercio mayorista (c1; $O_i = 18$), la industria (i1; $O_i = 9$) y, finalmente, el grupo residencial (r1; $O_i = 3$). Debemos tener en cuenta que el grupo de seguridad s1 se encontraría en el sexto día del aislamiento preventivo (cuarentena), por lo que deberíamos analizarlo nuevamente una vez transcurrido el período de incubación del virus (tabla 5). Si bien el grupo médico del hospital 1 (m1) tuvo un contacto intergrupar con el grupo de seguridad infectado (s1), transcurrieron 6 días desde su último análisis.

- Comentado [MAP370]: CG2
- Comentado [MAP371]: CG9
- Comentado [MAP372]: EE
- Comentado [MAP373]: CG6
- Comentado [MAP374]: CG1
- Comentado [MAP375]: EE
- Comentado [MAP376]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP377]: CLT. ANGLICISMO LÉXICO
- Comentado [MAP378]: CE
- Comentado [MAP379]: EE
- Comentado [MAP380]: CG10
- Comentado [MAP381]: CG10
- Comentado [MAP382]: CG1
- Comentado [MAP383]: CG10
- Comentado [MAP384]: CG10
- Comentado [MAP385]: CG10
- Comentado [MAP386]: CE
- Comentado [MAP387]: CE
- Comentado [MAP388]: CE
- Comentado [MAP389]: CG10
- Comentado [MAP390]: CG10
- Comentado [MAP391]: CG1

El tiempo transcurrido para el grupo m2 del hospital 2 fue de 12 días. Además, el hecho de que no existan contagios grupales en el hospital 2 no implica que ese hospital no esté recibiendo personas enfermas que podrían contagiar a los grupos analizados. En cambio, si el grupo de seguridad s1 diese negativo, deberíamos realizar el análisis al grupo del geriátrico 1 (g1; $O_i = 100$) en primer lugar, seguido por los médicos del hospital 2 (m2; $O_i = 91$), el comercio mayorista (c1; $O_i = 26$), la seguridad del hospital 2 (s2; $O_i = 25$), los médicos del hospital 1 (m1; $O_i = 25$), la industria (i1; $O_i = 14$), el grupo de seguridad s1 (s1; $O_i = 7$) y, finalmente, el grupo residencial (r1; $O_i = 4$). Este ejemplo permite visualizar el potencial para el ahorro en pruebas cuando analizamos la diseminación del COVID-19 en un determinado municipio. Podríamos analizar hasta 55 individuos distribuidos en 8 grupos con 8 pruebas, lo que significaría un ahorro del 85,45% de las pruebas con respecto al análisis de muestras individuales.

En otro trabajo de investigación con experimentos a través de RT-qPCR, los autores determinaron que es posible analizar grupos de hasta 32 individuos, con un 10 % de falsos negativos para la presencia de SARS-CoV-2. Por otro lado, Hogan et al. (2020) analizaron mediante la técnica de RT-qPCR las muestras almacenadas en un hospital de personas cuyas historias clínicas incluían síntomas de infección por coronavirus, en grupos de hasta 10 individuos. Tres de los 292 grupos analizados por estos autores dieron resultados positivos para infección con SARS-CoV-2. Con los análisis individuales, comprobaron que dos de los tres grupos estaban integrados por personas infectadas. Sin embargo, como la calidad de las pruebas de detección y los protocolos varían internacionalmente, es recomendable realizar experimentos independientes antes de implantar la propuesta de este trabajo.

- Comentado [MAP392]: EE
- Comentado [MAP393]: CG4
- Comentado [MAP394]: CG2
- Comentado [MAP395]: CG1
- Comentado [MAP396]: CG2
- Comentado [MAP397]: CG7
- Comentado [MAP398]: CG10
- Comentado [MAP399]: EE
- Comentado [MAP400]: CG2
- Comentado [MAP401]: CE
- Comentado [MAP402]: CE
- Comentado [MAP403]: CO
- Comentado [MAP404]: CE
- Comentado [MAP405]: CO
- Comentado [MAP406]: CG10
- Comentado [MAP407]: CG10
- Comentado [MAP408]: CG1
- Comentado [MAP409]: EE
- Comentado [MAP410]: CG10
- Comentado [MAP411]: CG10
- Comentado [MAP412]: CG10
- Comentado [MAP413]: CO
- Comentado [MAP414]: CE
- Comentado [MAP415]: EE
- Comentado [MAP416]: CG1
- Comentado [MAP417]: CG10
- Comentado [MAP418]: CG10
- Comentado [MAP419]: CG1

El algoritmo presentado permite establecer un orden racional de prioridad para realizar el análisis de grupos críticos de la sociedad (fig. 3). Como una limitación relevante de esta propuesta, cabe destacar que la asignación de los valores de las variables de los grupos está basada en el criterio y la experiencia del operador de la herramienta, además de las particularidades epidemiológicas analizadas. Por esta razón, de aplicar esta herramienta, su potencial aumentará a medida que avance la infección en la comunidad y el conocimiento de sus particularidades.

Definir grupos

Ordenar análisis

Analizar grupos

GRUPO 1

GRUPO 2

GRUPO 3

GRUPO 4

GRUPO i

Variables generales

Función de contagio general

Cantidad de pruebas disponibles

ORDEN DE ANÁLISIS

GRUPO 1

GRUPO 2

GRUPO 3

GRUPO 4

GRUPO i

Variables de grupo

Prioridad del grupo

Número de integrantes

Contacto intragrupal

Vulnerabilidad por la actividad

Contacto intergrupalo

GRUPO 2

Muestra combinada

Análisis GRUPO 2

Actividad normal

GRUPO 4

Muestra combinada

Análisis GRUPO 4

Cuarentena

Variables de retroalimentación

Vulnerabilidad por tiempo transcurrido

Activación intergrupalo

Figura 3. Esquema de trabajo propuesto de los pasos a seguir con las variables correspondientes a cada etapa.

Comentado [MAP420]: CLT. ANGLICISMO LÉXICO

Comentado [MAP421]: CG10

Comentado [MAP422]: CG10

Comentado [MAP423]: CG6

Comentado [MAP424]: EE

Comentado [MAP425]: CG1

Comentado [MAP426]: CE

Comentado [MAP427]: CG10

CONCLUSIONES

Mediante este análisis simulado demostramos que el algoritmo propuesto es una opción interesante y puede ser una herramienta para el control de la diseminación del SARS-CoV-2. Consideramos que esta propuesta merece ser explorada y desarrollada, con el fin de mejorar la administración del distanciamiento social y optimizar el uso de los recursos limitantes (pruebas de detección). Esta herramienta presenta un alto potencial para mitigar los problemas que genera la pandemia y disminuir la probabilidad de rebrote cuando se cuenta con recursos escasos, con la consiguiente reducción del impacto sobre la salud, la sociedad y la economía de una región.

- Comentado [MAP428]: CG1
- Comentado [MAP429]: EE
- Comentado [MAP430]: CLT, ANGLICISMO LÉXICO
- Comentado [MAP431]: CLT
- Comentado [MAP432]: CG1
- Comentado [MAP433]: EE
- Comentado [MAP434]: EE
- Comentado [MAP435]: EE
- Comentado [MAP436]: CGS
- Comentado [MAP437]: CLT
- Comentado [MAP438]: EE
- Comentado [MAP439]: CE
- Comentado [MAP440]: CE

ANEXO III. POSEDICIÓN DEL TEXTO BRUTO

Proposals for the control and mitigation of the spread of COVID-19: a strategic management of the disease

ABSTRACT

The speed of spread of COVID-19 in the world led the affected countries to close their borders and take social distancing measures. Six months after the disease was declared a pandemic, many countries started taking measures to make confinement more flexible, although without a vaccine or a medicine capable of dealing with the SARS-CoV-2 infection, the situation could be reversed at any time. The objective of this work was to propose a decision algorithm aimed at optimizing the detection of asymptomatic cases and managing quarantine in a strategic way in order to avoid the spread of the virus and advance towards a supervised normality. A tentative proposal for optimization and organization of SARS-CoV-2 detection tests was developed, based on the analysis of composite samples collected from those taken individually from asymptomatic people who are part of the cohorts of interest. Cohorts were defined according to their function in society or degree of vulnerability. The algorithm considers variables such as the priority of the cohort, the number of members of the analysis groups within each cohort, intragroup and intergroup contact, vulnerability to contagion due to the activity carried out and the time elapsed since the test was last carried out. The proposal was illustrated with defined hypothetical cohorts, with a single analysis group to simplify, and it was found that the application of the tool allows to establish a rational order of priority to carry out tests in critical groups of society. This tool will enable optimizing resources and reducing the impact of the disease on health, society, and the economy of a region.

KEYWORDS: COVID-19, Social distancing, SARS-CoV-2, Decision algorithm, Quarantine

INTRODUCTION

Coronaviruses are a family of viruses that infect animals, mainly birds and mammals, in which they cause disease. This translates into economic losses associated with productive activities¹. They can also jump from certain species and infect humans, causing diseases of varying severity (cold, bronchitis, pneumonia, among others), depending on the particular virus and the host's health conditions. While most people are infected with these viruses sometime in their lives without great consequences, some coronaviruses have caused diseases that have spread rapidly and have caused worrisome outbreaks and epidemics. Examples include the severe acute respiratory syndrome (SARS) originating in Yunnan, China, in 2001; the Middle East respiratory syndrome (MERS) caused by MERS-CoV, which started in Saudi Arabia in 2012¹⁸, and currently COVID-19, caused by a new coronavirus, called SARS-CoV-2⁶.

This new disease, COVID-19, is believed to have emerged in December 2019 at a seafood market (where live and dead exotic animals were sold) located in Wuhan, Hubei province, China, although there are some controversies about this¹². The speed of spread of the virus was so high that, in a short time, the authorities of China closed the city and ordered mandatory quarantine. Despite this, the contagion spread rapidly through the neighboring countries first, and, gradually, to the others, facilitated by the great mobility of travelers. Only almost three months after detecting the disease, on March 11, 2020, the World Health Organization (WHO) declared it a pandemic.

The rapid advance of the disease highlighted the lack of preparation of the healthcare systems of most countries to face the situation. The number of sick people far exceeded available capacities

Comentado [MAP1]: Adición palabras. TA

Comentado [MAP2]: Adición palabras. TB

Comentado [MAP3]: Precisión léxica. TA

Comentado [MAP4]: Estructura natural. TB

Comentado [MAP5]: Puntuación. TB

Comentado [MAP6]: Precisión léxica. TB

Comentado [MAP7]: Precisión léxica. TB

Comentado [MAP8]: Precisión léxica. TB

Comentado [MAP9]: Estructura natural. TA

Comentado [MAP10]: Precisión léxica. TA

Comentado [MAP11]: Estructura natural. TA

Comentado [MAP12]: Tiempo verbal. TA

Comentado [MAP13]: Puntuación. TA

Comentado [MAP14]: Puntuación. TB

Comentado [MAP15]: Semántica. TB

Comentado [MAP16]: Semántica. TB

Comentado [MAP17]: Puntuación. TB

Comentado [MAP18]: Estructura natural. TA

Comentado [MAP19]: Precisión léxica. TB

and the world watched in amazement at the rising death toll in developed countries in Europe. Given this scenario, and in order to contain the contagion, in most countries social distancing measures (better known as quarantine) were established, with different modalities. As of the writing of this article (May 19, 2020), the number of confirmed cases in the world was rapidly approaching five million, of which approximately 7% corresponded to deaths³. As of the date of the correction of this manuscript, September 15, 2020, there are almost 30 million confirmed cases in the world and the number of deaths is close to one million.

Early on (March 20, 2020), the government of Argentina established the preventive and compulsory social distancing (aislamiento social preventivo y obligatorio, ASPO). This made it possible to avoid massive contagion and to buy time for the healthcare system to improve its response capacity. Thus, new hospitals or care centers were set up, respirators, personal protection elements and other supplies were acquired and/or manufactured to be able to attend the emergency, while the scientific community was called upon to contribute to the generation of knowledge to face the pandemic.

As in the rest of the world, the detection of SARS-CoV-2 is carried out in Argentina using real-time reverse transcription polymerase chain reaction (RT-qPCR) tests. In principle, this test is performed on each individual suspected of being a carrier because they present symptoms related to the disease or due to an epidemiological connection, although this has changed with the evolution of the pandemic. Other molecular tests recently developed by Argentinian researchers allow detecting the presence of genetic fragments and have shown promising results: these are the NEOKIT and the ELA-CHEMSTRIP. The advantage of the latter methods is that they would allow the virus to be identified more quickly, which would make it possible to detect it in a larger population group. Despite this, in most developing countries the capacity to perform such analyses is limited, so it is not possible to assess the population that is assumed to be healthy or asymptomatic (although it may include contagious individuals) and, therefore, widespread measures should be taken to control the spread of the virus.

Taking into account the healthcare context in which we live, and that social distancing has served its purpose in terms of slowing down the spread of the disease, it is considered necessary to return to activity gradually and in a controlled manner. However, it is known that, should the healthcare situation deteriorate, the flexibilization measures could be reversed. Therefore, the objective of this work was to propose a decision algorithm aimed at optimizing the detection of asymptomatic cases and managing quarantine in a strategic way, in order to avoid the spread of the virus and allow a new supervised normality.

Comentado [MAP20]: Puntuación. TA

Comentado [MAP21]: Puntuación. TB

Comentado [MAP22]: Estructura natural. TA

Comentado [MAP23]: Precisión léxica. TA

Comentado [MAP24]: Omisión. TA

Comentado [MAP25]: Precisión léxica. TB

Comentado [MAP26]: Puntuación. TB

Comentado [MAP27]: Convención. TA

Comentado [MAP28]: Precisión léxica. TB

Comentado [MAP29]: Puntuación. TB

Comentado [MAP30]: Mayúscula. TA

Comentado [MAP31]: Precisión léxica. TB

Comentado [MAP32]: Adición espacio. TA

Comentado [MAP33]: Adición espacio. TA

Comentado [MAP34]: Precisión léxica. TA

Comentado [MAP35]: Puntuación. TB

Comentado [MAP36]: Mayúscula. TA

Comentado [MAP37]: Precisión léxica. TA

Comentado [MAP38]: Adición espacio. TA

Comentado [MAP39]: Precisión léxica. TB

Comentado [MAP40]: Precisión léxica. TB

Comentado [MAP41]: Precisión léxica. TB

MATERIALS AND METHODS

A tentative tool was developed to strategically use SARS-CoV-2 virus detection tests in order to control the spread of the disease in multiple cohorts. To this aim, and as already indicated by other authors, we suggested analyzing composite samples (pool analysis) by combining the individual samples of different asymptomatic people of the cohort of interest. In this way, the presence of the virus could be evaluated in a greater number of people simultaneously, with a considerable saving of resources and time. We proposed to begin by analyzing the cohorts of people whose function was critical to the normal functioning of society, or who belonged to very vulnerable sectors, and then to continue with non-critical cohorts, according to their importance and degree of vulnerability.

- Comentado [MAP42]: Estructura natural. TA
- Comentado [MAP43]: Estructura natural. TA
- Comentado [MAP44]: Precisión léxica. TA
- Comentado [MAP45]: Estructura natural. TA
- Comentado [MAP46]: Estructura natural. TA
- Comentado [MAP47]: Preposición. TA
- Comentado [MAP48]: Estructura natural. TA
- Comentado [MAP49]: Estructura natural. TA
- Comentado [MAP50]: Preposición. TA
- Comentado [MAP51]: Precisión léxica. TB
- Comentado [MAP52]: Estructura natural. TA
- Comentado [MAP53]: Omisión. TA
- Comentado [MAP54]: Puntuación. TA

Definition of critical cohorts

The cohorts proposed here as critical, and that should be the first to undergo the virus detection, would cover the following sectors: healthcare, security, trade of essential consumer goods, the elderly (nursing home), essential services and industries (manufacturing of respirators, food, etc.). It is essential to study the population of these sectors, as they are potential contagion vectors. Healthcare personnel (doctors, nurses, cleaning personnel, etc.) should be kept in optimal condition, so that, in addition to providing them with personal protection elements, they should be tested systematically and periodically. This is also necessary for security personnel, as they are highly exposed to potentially infected people. Supply chain traders, including large fruit and vegetable markets, should be regularly screened as well, as an infected person could bring the entire chain to a standstill and lead to serious social problems. Other sectors, such as construction workers, administrative workers, traders of non-essential goods, inhabitants of residential areas, to name a few, correspond to non-critical sectors, so they should be analyzed whenever the number of available tests is sufficient, without compromising the control of the sectors that are critical. However, those non-critical cohorts that are related to individuals of critical cohorts should have a higher priority for analysis.

- Comentado [MAP55]: Estructura natural. TA
- Comentado [MAP56]: Puntuación. TA
- Comentado [MAP57]: Estructura natural. TA
- Comentado [MAP58]: Puntuación. TA
- Comentado [MAP59]: Precisión léxica. TA
- Comentado [MAP60]: Precisión léxica. TB
- Comentado [MAP61]: Precisión léxica. TA
- Comentado [MAP62]: Preposición. TA
- Comentado [MAP63]: Precisión léxica. TA
- Comentado [MAP64]: Precisión léxica. TB
- Comentado [MAP65]: Precisión léxica. TB
- Comentado [MAP66]: Precisión léxica. TB
- Comentado [MAP67]: Precisión léxica. TA
- Comentado [MAP68]: Estructura natural. TA
- Comentado [MAP69]: Estructura natural. TA
- Comentado [MAP70]: Precisión léxica. TA
- Comentado [MAP71]: Precisión léxica. TB
- Comentado [MAP72]: Tiempo verbal. TA
- Comentado [MAP73]: Estructura natural. TA
- Comentado [MAP74]: Precisión léxica. TA
- Comentado [MAP75]: Precisión léxica. TA
- Comentado [MAP76]: Tiempo verbal. TA
- Comentado [MAP77]: Semántica. TB
- Comentado [MAP78]: Precisión léxica. TB
- Comentado [MAP79]: Estructura natural. TA

Carrying out the analyses

After defining the groups (or pools) of the different cohorts, we propose taking individual samples (nasopharyngeal swabs) from all the asymptomatic people who comprise them. A part of the sample could be mixed in the laboratory to create a composite sample from each group for analysis. Initially, the infectivity rate and prevalence should be estimated from the official contagion data on in the analyzed geographic location, with the aim of defining the group size and the detection methodology.

Pool sample analysis makes sense to save resources and time, as long as the prevalence is lower than 10%, a value from which it is convenient to carry out individual tests due to the high number of groups that would have positive results (and they would have to be opened to study the individual samples). The size of the groups would be defined by the prevalence: the higher the prevalence, the smaller the number of samples that would comprise a group. The prevalence of the disease in the population, and also in the cohorts comprised of essential personnel, would be considered as known epidemiological information, and could be calculated from the confirmed (individual) cases in the groups of interest in relation to the population, or a defined population.

- Comentado [MAP80]: Precisión léxica. TA
- Comentado [MAP81]: Convención. TA
- Comentado [MAP82]: Estructura natural. TA
- Comentado [MAP83]: Precisión léxica. TA
- Comentado [MAP84]: Precisión léxica. TB
- Comentado [MAP85]: Omisión. TA
- Comentado [MAP86]: Precisión léxica. TA
- Comentado [MAP87]: Precisión léxica. TA
- Comentado [MAP88]: Estructura natural. TA
- Comentado [MAP89]: Precisión léxica. TA
- Comentado [MAP90]: Estructura natural. TA
- Comentado [MAP91]: Precisión léxica. TB
- Comentado [MAP92]: Precisión léxica. TA
- Comentado [MAP93]: Precisión léxica. TA
- Comentado [MAP94]: Precisión léxica. TB
- Comentado [MAP95]: Puntuación. TA
- Comentado [MAP96]: Precisión léxica. TA
- Comentado [MAP97]: Precisión léxica. TA
- Comentado [MAP98]: Precisión léxica. TA
- Comentado [MAP99]: Puntuación. TA
- Comentado [MAP100]: Precisión léxica. TA
- Comentado [MAP101]: Puntuación. TA
- Comentado [MAP102]: Estructura natural. TA
- Comentado [MAP103]: Estructura natural. TA
- Comentado [MAP104]: Puntuación. TA
- Comentado [MAP105]: Estructura natural. TA

It should be clarified that, before starting the detection through pools of samples, some critical methodological aspects should be evaluated to ensure the success of this diagnostic modality. The first thing that should be verified through bioinformatics is that the sequences of the oligonucleotides that will be used in the detection are specific. Then, it would be necessary to determine the efficiency of the amplification of the RT-qPCR reaction and verify that the efficiency is at maximum, even at low concentrations of RNA. Furthermore, it would be essential to determine the detection limit of the method used, since the classification of the results between positive and not detectable (negative) will depend on that. Finally, a validation of the pool technique should be carried out in a group of the population, and these results should be compared with those obtained by individual tests to determine the probability of a misclassification of a result. Only after having all the aforementioned knowledge could begin with the detection in pools of samples.

The maximum number of individuals in a group should be defined taking into account the volume per individual sample that must be added to the detection reaction without compromising the sensitivity, while allowing enough viral load to reach it. In addition, the prevalence of the virus or the ratio between the number of infected individuals and the total population analyzed should be taken into account to optimize the size of the groups from a statistical point of view. These values would serve as a guide, but the definition of the groups should be done according to practical reasons. For example, if the optimal group size were 15 people, and in a nursing home there were 10 people, it would not be necessary or convenient to add people from outside the group to reach the optimal size value. On the other hand, if the cohort of a complete neighborhood is analyzed, the groups should have the optimal size and be selected according to proximity in order to save resources.

Groups with negative results in SARS-CoV-2 tests should be tested periodically, and the procedure should be optimized according to an algorithm that resolves the order of priority of each group. If any of the tests carried out are positive, the group involved and the people with whom they have had close contact should be isolated to avoid contagion, until the analysis is completed. To this end, the laboratory should open the positive pools and focus on the analysis of the individual samples to identify the infected individual, who should then remain isolated (fig. 1). People with negative individual results could come out of isolation.

- Comentado [MAP106]: Convención. TA
- Comentado [MAP107]: Estructura natural. TA
- Comentado [MAP108]: Adición de espacio. TA
- Comentado [MAP109]: Estructura natural. TA
- Comentado [MAP110]: Puntuación. TA
- Comentado [MAP111]: Puntuación. TA
- Comentado [MAP112]: Convención. TA
- Comentado [MAP113]: Estructura natural. TA
- Comentado [MAP114]: Puntuación. TA
- Comentado [MAP115]: Precisión léxica. TB
- Comentado [MAP116]: Estructura natural. TA
- Comentado [MAP117]: Estructura natural. TA
- Comentado [MAP118]: Estructura natural. TA
- Comentado [MAP119]: Omisión. TA
- Comentado [MAP120]: Estructura natural. TA
- Comentado [MAP121]: Precisión léxica. TA
- Comentado [MAP122]: Precisión léxica. TB
- Comentado [MAP123]: Estructura natural. TA
- Comentado [MAP124]: Puntuación. TA
- Comentado [MAP125]: Precisión léxica. TA
- Comentado [MAP126]: Estructura natural. TA
- Comentado [MAP127]: Puntuación. TA
- Comentado [MAP128]: Precisión léxica. TB
- Comentado [MAP129]: Estructura natural. TA
- Comentado [MAP130]: Estructura natural. TA
- Comentado [MAP131]: Tiempo verbal. TB
- Comentado [MAP132]: Estructura natural. TA
- Comentado [MAP133]: Puntuación. TA
- Comentado [MAP134]: Convención. TA
- Comentado [MAP135]: Adición de espacio. TA
- Comentado [MAP136]: Precisión léxica. TB
- Comentado [MAP137]: Mayúscula. TA

Define the cohorts to analyze

Take samples of the individuals in the cohort and combine them into a pool sample

Perform pool sample analysis

Calculate waiting time according to cohort priority

Negative

Pool analysis result

Positive

Maintain alert and distancing for 2 weeks

Analyze samples individually or by subgroups

Negative

Individual analysis result

Positive

Figure 1. Flow chart that indicates the steps to follow to mitigate the spread of the disease in the population and optimize the use of diagnostic resources.

Resource Optimization

A decision algorithm was developed to optimize the analysis and establish a logical order to carry it out in order to model and evaluate the behavior of the transmission of the disease based on the impact that the defined cohorts have on the health, the safety and the economy. Ideally, a deep knowledge of the population of the analysis cohorts is required, so the implementation of this algorithm at the municipal level would be more beneficial.

It was considered necessary, although not essential, to include the following general variables in the model, since they are assumed to affect all groups in all cohorts, and these should be updated daily:

Number of tests available: it determines the number of groups that can be analyzed and, therefore, the number of cohorts that can be periodically analyzed.

General contagion function: it is defined as the relationship between the population that is susceptible to infection, the infected population and the recovered population. It is applied in an area that covers the cohorts analyzed and depends on the historical values of these variables.

Other important variables are those related to particular characteristics of the group to be analyzed: Location/coordinates: allows to know the location of the group and can be useful to determine the level of intergroup contact or help decide which is the vulnerability to contagion due to the activity. For example, there might be an increased risk of contagion for a security group that shares the location with a group in a hospital where there are individuals that tested positive for SARS-CoV-2.

- Comentado [MAP138]: Precisión léxica. TA
- Comentado [MAP139]: Precisión léxica. TB
- Comentado [MAP140]: Precisión léxica. TA
- Comentado [MAP141]: Estructura natural. TA
- Comentado [MAP142]: Precisión léxica. TA
- Comentado [MAP143]: Estructura natural. TA
- Comentado [MAP144]: Precisión léxica. TA
- Comentado [MAP145]: Precisión léxica. TA
- Comentado [MAP146]: Estructura natural. TA
- Comentado [MAP147]: Estructura natural. TA
- Comentado [MAP148]: Precisión léxica. TB
- Comentado [MAP149]: Precisión léxica. TA
- Comentado [MAP150]: Estructura natural. TA
- Comentado [MAP151]: Tiempo verbal. TB
- Comentado [MAP152]: Cambio de sentido. TA
- Comentado [MAP153]: Estructura natural. TA
- Comentado [MAP154]: Omisión. TA
- Comentado [MAP155]: Omisión. TA
- Comentado [MAP156]: Adición de palabras. TA
- Comentado [MAP157]: Precisión léxica. TB
- Comentado [MAP158]: Tiempo verbal. TA
- Comentado [MAP159]: Precisión léxica. TB
- Comentado [MAP160]: Preposición. TA
- Comentado [MAP161]: Adición de espacio. TA
- Comentado [MAP162]: Semántica. TB
- Comentado [MAP163]: Semántica. TB
- Comentado [MAP164]: Adición de espacio. TA
- Comentado [MAP165]: Adición de espacio. TA
- Comentado [MAP166]: Adición de espacio. TA
- Comentado [MAP167]: Estructura natural. TA
- Comentado [MAP168]: Estructura natural. TA
- Comentado [MAP169]: Preposición. TA
- Comentado [MAP170]: Estructura natural. TA

Group priority level (A): refers to the priority of one group over the others to be subject to analysis. For example, the groups in the healthcare sector cohort (doctors and nurses dedicated to the treatment of the disease caused by SARS-CoV-2) would have analysis priority over a group in a cohort comprised of people dedicated to the production of consumer goods. To make this decision as less subjective as possible, a scale from 1 to 10 was defined, in which 1 means no priority and 10 indicates the highest priority (table 1). Specific ranges were established for each hypothetical group included in the algorithm.

Table 1. Group priority scale for analysis

Group priority (A)

Range

Doctors dedicated to the direct treatment of those infected by SARS-CoV-2

Other doctors

Population of a nursing home

Security personnel

Consumer goods traders

Industry

Residential groups

- Comentado [MAP171]:** Adición de espacio. TA
- Comentado [MAP172]:** Adición de espacio. TA
- Comentado [MAP173]:** Estructura natural. TA
- Comentado [MAP174]:** Cambio de sentido. TA
- Comentado [MAP175]:** Omisión. TA
- Comentado [MAP176]:** Preposición. TA
- Comentado [MAP177]:** Precisión léxica. TB
- Comentado [MAP178]:** Estructura natural. TA
- Comentado [MAP179]:** Preposición. TA
- Comentado [MAP180]:** Precisión léxica. TA
- Comentado [MAP181]:** Estructura natural. TA
- Comentado [MAP182]:** Precisión léxica. TB
- Comentado [MAP183]:** Adición de espacio. TA
- Comentado [MAP184]:** Mayúscula. TA
- Comentado [MAP185]:** Adición de espacio. TA
- Comentado [MAP186]:** Precisión léxica. TA
- Comentado [MAP187]:** Estructura natural. TA
- Comentado [MAP188]:** Precisión léxica. TA
- Comentado [MAP189]:** Precisión léxica. TB
- Comentado [MAP190]:** Cambio de sentido. TA

Number of individuals of each group (*B*): the number of individuals of each group would be defined according to practical reasons and should be such that the test negatives of the analysis are true negatives (so that they do not compromise the detection limit of the sample) in order to prevent further spread of the infection. This means that if the analysis of an infected group is done, for example, through RT-qPCR, the sensitivity will depend on the phase of ongoing infection, the quality of the samples taken and the number of individual samples that are part of the composite sample (the greater the number of samples, the less amount of each target and, therefore, lower sensitivity).

Intragroup contact (*C*): this variable indicates the level of contact (related to social distance and the frequency of contact) between the individuals of the group analyzed. A scale between 1 and 10 was established for its evaluation, where high values indicate greater contact and low values less contact (null contact was assigned a value of 1). Contact values for different distances may coincide due to changes in contact frequency (table 2).

Table 2. Scale for intragroup contact (*C*)

| Distance (m) | Frequency |
|--------------|-----------|
| Range | |
| Low | |
| Medium | |
| High | |
| Low | |
| Medium | |
| High | |
| Low | |
| Medium | |
| High | |

- Comentado [MAP191]:** Precisión léxica. TB
- Comentado [MAP192]:** Adición de espacio. TA
- Comentado [MAP193]:** Adición de espacio. TA
- Comentado [MAP194]:** Precisión léxica. TB
- Comentado [MAP195]:** Estructura natural. TA
- Comentado [MAP196]:** Semántica. TB
- Comentado [MAP197]:** Estructura natural. TA
- Comentado [MAP198]:** Puntuación. TA
- Comentado [MAP199]:** Omisión. TA
- Comentado [MAP200]:** Preposición. TA
- Comentado [MAP201]:** Adición de espacio. TA
- Comentado [MAP202]:** Adición de espacio. TA
- Comentado [MAP203]:** Estructura natural. TA
- Comentado [MAP204]:** Adición de palabras. TA
- Comentado [MAP205]:** Precisión léxica. TA
- Comentado [MAP206]:** Precisión léxica. TA
- Comentado [MAP207]:** Estructura natural. TA
- Comentado [MAP208]:** Precisión léxica. TB
- Comentado [MAP209]:** Precisión léxica. TA
- Comentado [MAP210]:** Estructura natural. TA
- Comentado [MAP211]:** Estructura natural. TA
- Comentado [MAP212]:** Adición de espacio. TA
- Comentado [MAP213]:** Adición de espacio. TA
- Comentado [MAP214]:** Precisión léxica. TA
- Comentado [MAP215]:** Precisión léxica. TA
- Comentado [MAP216]:** Precisión léxica. TA
- Comentado [MAP217]:** Mayúscula. TA
- Comentado [MAP218]:** Precisión léxica. TA
- Comentado [MAP219]:** Precisión léxica. TA
- Comentado [MAP220]:** Mayúscula. TA
- Comentado [MAP221]:** Precisión léxica. TA
- Comentado [MAP222]:** Precisión léxica. TA
- Comentado [MAP223]:** Mayúscula. TA

Vulnerability to contagion due to the activity (D): refers to the degree of vulnerability of the groups studied within a cohort due to the activity. In this case, a scale between 1 and 10 was also established, in which a higher value indicates a higher level of vulnerability to contagion due to the activity (table 3). The values could be obtained from studies of the level of contagion in these groups and should be updated periodically.

Table 3. Scale for vulnerability to contagion (D)

| Activity | Value |
|----------------------|-------|
| Family doctor | |
| Specialized nurses | |
| Pharmacy technicians | |
| Prison warden | |
| Residential | |

Vulnerability due to elapsed time (E): calculated value (equation 1) between 0 and 10 that indicates the risk of an undetected infection during the elapsed time.

where t (in days) is the time elapsed since the last test performed and t_i is the maximum incubation period of the virus (14 days).

Intergroup activation (F): value used to indicate whether the order of priority will increase (or not) due to contact with a group whose analysis was positive. That is, because of the intergroup relationship with infected groups, the group analyzed increases their order of priority and, therefore, this could lead to it being analyzed first. Thus, F will take the value of 1 if the test is positive and 0 if it is negative (the intergroup relationship is not activated).

Intergroup contact (G): variable that establishes the level of contact between different groups (from the same cohort or from a different one). A scale between 1 and 10 (null contact was assigned the value of 1) was defined, where high values indicate a closer contact between the individuals of different groups (the relationship of group i with group j is defined as relationship ij , with $i \neq j$). Contact values for different distances may coincide due to changes in the contact frequency (table 4).

- Comentado [MAP224]: Adición de espacio. TA
- Comentado [MAP225]: Adición de espacio. TA
- Comentado [MAP226]: Preposición. TA
- Comentado [MAP227]: Adición de espacio. TA
- Comentado [MAP228]: Mayúscula. TA
- Comentado [MAP229]: Adición de espacio. TA
- Comentado [MAP230]: Cambio de sentido. TA
- Comentado [MAP231]: Cambio de sentido. TA
- Comentado [MAP232]: Estructura natural. TA
- Comentado [MAP233]: Precisión léxica. TA
- Comentado [MAP234]: Adición de espacio. TA
- Comentado [MAP235]: Adición de espacio. TA
- Comentado [MAP236]: Preposición. TA
- Comentado [MAP237]: Adición de espacio. TA
- Comentado [MAP238]: Adición de espacio. TA
- Comentado [MAP239]: Precisión léxica. TB
- Comentado [MAP240]: Estructura natural. TA
- Comentado [MAP241]: Estructura natural. TA
- Comentado [MAP242]: Tiempo verbal. TA
- Comentado [MAP243]: Precisión léxica. TB
- Comentado [MAP244]: Estructura natural. TA
- Comentado [MAP245]: Adición de espacio. TA
- Comentado [MAP246]: Adición de espacio. TA
- Comentado [MAP247]: Estructura natural. TA
- Comentado [MAP248]: Estructura natural. TA
- Comentado [MAP249]: Precisión léxica. TB
- Comentado [MAP250]: Precisión léxica. TA
- Comentado [MAP251]: Precisión léxica. TA
- Comentado [MAP252]: Adición de espacio. TA
- Comentado [MAP253]: Adición de espacio. TA
- Comentado [MAP254]: Estructura natural. TA
- Comentado [MAP255]: Omisión. TA
- Comentado [MAP256]: Adición de espacio. TA
- Comentado [MAP257]: Adición de espacio. TA

Table 4. Scale for intergroup contact (G)

Distance (m)

Frequency

Range

Low

Medium

High

Low

Medium

High

Low

Medium

High

General analysis order (O_j): value whose result would allow structuring the test priority according to the current model. First, taking into account the previously established variables, a local value is calculated for each group (L_i , equation 2) and the highest value is identified (L_{max}). Local values would represent relevance for analysis and do not have a defined scale. Then, to better visualize which groups should be analyzed first, the local values will be weighed with respect to the maximum local value (L_{max}) and the general analysis priority order will be obtained for each group (O_i , equation 3).

where j varies from 1 to the maximum value of j (total number of defined groups), with $i \neq j$:

| | |
|---------------------|------------------------|
| Comentado [MAP258]: | Precisión léxica. TA |
| Comentado [MAP259]: | Precisión léxica. TA |
| Comentado [MAP260]: | Precisión léxica. TA |
| Comentado [MAP261]: | Mayúscula. TA |
| Comentado [MAP262]: | Precisión léxica. TA |
| Comentado [MAP263]: | Precisión léxica. TA |
| Comentado [MAP264]: | Mayúscula. TA |
| Comentado [MAP265]: | Precisión léxica. TA |
| Comentado [MAP266]: | Precisión léxica. TA |
| Comentado [MAP267]: | Mayúscula. TA |
| Comentado [MAP268]: | Estructura natural. TA |
| Comentado [MAP269]: | Adición de espacio. TA |
| Comentado [MAP270]: | Adición de espacio. TA |
| Comentado [MAP271]: | Precisión léxica. TA |
| Comentado [MAP272]: | Estructura natural. TA |
| Comentado [MAP273]: | Estructura natural. TA |
| Comentado [MAP274]: | Adición de espacio. TA |
| Comentado [MAP275]: | Adición de espacio. TA |
| Comentado [MAP276]: | Adición de espacio. TA |
| Comentado [MAP277]: | Adición de espacio. TA |
| Comentado [MAP278]: | Adición de espacio. TA |
| Comentado [MAP279]: | Adición de espacio. TA |
| Comentado [MAP280]: | Precisión léxica. TA |
| Comentado [MAP281]: | Preposición. TB |
| Comentado [MAP282]: | Precisión léxica. TA |
| Comentado [MAP283]: | Adición de espacio. TA |
| Comentado [MAP284]: | Adición de espacio. TA |
| Comentado [MAP285]: | Adición de espacio. TA |
| Comentado [MAP286]: | Adición de espacio. TA |
| Comentado [MAP287]: | Adición de espacio. TA |
| Comentado [MAP288]: | Adición de espacio. TA |
| Comentado [MAP289]: | Estructura natural. TA |
| Comentado [MAP290]: | Adición de espacio. TA |

Once the general analysis order of the groups (O) is defined, the detection can be carried out following that order (from highest to lowest) and the number of available tests can be updated. Then, the information obtained is fed back to the algorithm with the variables E (vulnerability due to elapsed time) and F (intergroup activation) (fig. 2).

GROUP j

GROUP i

(G) Intergroup contact

Structure analysis

GROUP VARIABLES

Location / coordinates

(A) Group priority

(B) Number of individuals

(C) Intragroup contact

(D) Vulnerability to contagion due to the activity

(O) Analysis order

Analyze groups

FEEDBACK VARIABLES

(E) Vulnerability due to elapsed time

(F) Intergroup activation

GENERAL VARIABLES

Number of tests available

General contagion function

Figure 2. General algorithm for calculating the analysis order of the groups and making rational use of the available screening tests. The white boxes indicate the input and output variables of the model, while the blue blocks are the operations.

Time scope and costs

According to the WHO, this pandemic is expected to last for at least two years. Although there are already many projects for both vaccines and palliative treatments, none have yet been implemented effectively and massively, so the duration of the pandemic can be difficult to predict. In the same way, the application timeline of the tool proposed in this work is difficult to establish, since it depends on many variables and is closely related to the duration of the pandemic. The tool is very flexible, since it allows operating with various numbers of cohorts (as required or according to the budget available) and also allows testing at a convenient frequency for each location where it is applied. Therefore, the cost of implementation will depend, mainly, on the available analysis capacity. Another variable that could influence costs is the possible implementation of rapid tests, which have been developed and will be available (mainly for the healthcare system) in the coming months. It is important to note that the analysis of pooled samples allows a large amount of information to be obtained and minimizes costs by analyzing many individual samples in a single reaction.

Comentado [MAP291]: Adición de espacio. TA

Comentado [MAP292]: Adición de espacio. TA

Comentado [MAP293]: Tiempo verbal. TA

Comentado [MAP294]: Adición de espacio. TA

Comentado [MAP295]: Tiempo verbal. TA

Comentado [MAP296]: Adición de espacio. TA

Comentado [MAP297]: Mayúscula. TA

Comentado [MAP298]: Adición de espacio. TA

Comentado [MAP299]: Precisión léxica. TA

Comentado [MAP300]: Precisión léxica. TB

Comentado [MAP301]: Preposición. TA

Comentado [MAP302]: Estructura natural. TA

Comentado [MAP303]: Preposición. TA

Comentado [MAP304]: Precisión léxica. TA

Comentado [MAP305]: Cambio de sentido. TA

Comentado [MAP306]: Estructura natural. TA

Comentado [MAP307]: Precisión léxica. TA

Comentado [MAP308]: Precisión léxica. TB

Comentado [MAP309]: Estructura natural. TA

Comentado [MAP310]: Estructura natural. TA

Comentado [MAP311]: Estructura natural. TA

Comentado [MAP312]: Precisión léxica. TA

Comentado [MAP313]: Adición de espacio. TA

Comentado [MAP314]: Precisión léxica. TB

RESULTS AND DISCUSSION

To illustrate the application of the algorithm and evaluate its versatility and feasibility, a simulation was carried out. Due to the lack of real data at the time of formulating this algorithm, it was necessary to define some arbitrarily (trying to get closer to reality) and make different assumptions. In the first place, eight hypothetical groups were defined (for illustrative purposes, since in reality there will be many more): GROUP 1, doctors belonging to hospital 1 (m1); GROUP 2, hospital 1 security personnel (s1); GROUP 3, doctors from hospital 2 (m2); GROUP 4, hospital 2 security personnel (s2); GROUP 5, population of a nursing home (elderly and staff) located in an urban area (g1), under the assumption that no cases were recorded in the area; GROUP 6, personnel of the wholesale trade 1 of consumer goods, located in an area where there are other wholesale shops (c1); GROUP 7, goods manufacturing industry located in an area where COVID-19 cases were recorded (i1); GROUP 8, large family from a residential neighborhood, located in an urban area where there were no cases of COVID-19 (r1). For each group a specific number of members was established (tables 5 and 6).

Table 5. Hypothetical example of application of the developed decision algorithm

Groups

Group priority (A)

Number of members of each group (B)

Intragroup contact (C)

Vulnerability to contagion due to the activity (D)

Days elapsed since the analysis

Vulnerability due to elapsed time (E) a

Local analysis order (Li) b

General analysis order (Oi) c

Test priority order

a, b, c Calculated through equations 1, 2 and 3.

Defined groups. m1: doctors from hospital 1; s1: hospital 1 security; m2: doctors from hospital 2; s2: hospital 2 security; g1: population in nursing home 1; c1: store 1 personnel; i1: industry 1 personnel; r1: neighbors of the residential area 1. The group that is assumed to be infected is marked in bold.

Table 6. Hypothetical example to illustrate a matrix of intergroup relationships

Groups

F: intergroup activation (1: infected, 0: not infected); NA: intergroup contact does not apply to the relationship of a group with itself.

Defined groups. m1: doctors from hospital 1; s1: hospital 1 security; m2: doctors from hospital 2; s2: hospital 2 security.

The group that is assumed to be infected is marked in bold.

Regarding the variable Group priority (A), a value of 10 was assigned to doctors from hospital 1 (m1) (which registers the largest number of patients with COVID-19) and 9 to those from hospital 2 (m2). The high values assigned are due to the fact that both are priority groups (as they are part of the Hospital doctors cohort), in accordance with their function of caring for people with COVID-19. For the groups of the security cohort, a value of 6 was established for both s1 and s2, since, although they are essential, they could be replaced if contagion occurs. The population of nursing home g1 are a high-risk group, because the mortality due to contagion is much higher in this group, and therefore a value of 8 was assigned. Wholesale trade of consumer goods c1 has a priority of 4 (relatively low), as there are other alternative wholesale businesses in the area. For industry i1, the

Comentado [MAP315]: Orden. TA

Comentado [MAP316]: Orden. TA

Comentado [MAP317]: Precisión léxica. TB

Comentado [MAP318]: Precisión léxica. TB

Comentado [MAP319]: Orden. TA

Comentado [MAP320]: Mayúscula. TA

Comentado [MAP321]: Preposición. TA

Comentado [MAP322]: Omisión. TA

Comentado [MAP323]: Preposición. TA

Comentado [MAP324]: Estructura natural. TA

Comentado [MAP325]: Mayúscula. TA

Comentado [MAP326]: Mayúscula. TA

Comentado [MAP327]: Preposición. TA

Comentado [MAP328]: Orden. TA

Comentado [MAP329]: Orden. TA

Comentado [MAP330]: Precisión léxica. TB

Comentado [MAP331]: Orden. TA

Comentado [MAP332]: Orden. TA

Comentado [MAP333]: Precisión léxica. TA

Comentado [MAP334]: Estructura natural. TA

Comentado [MAP335]: Orden. TA

Comentado [MAP336]: Orden. TA

Comentado [MAP337]: Mayúscula. TA

Comentado [MAP338]: Cambio de sentido. TA

Comentado [MAP339]: Estructura natural. TA

Comentado [MAP340]: Mayúscula. TA

Comentado [MAP341]: Puntuación. TA

Comentado [MAP342]: Precisión léxica. TA

Comentado [MAP343]: Puntuación. TA

Comentado [MAP344]: Precisión léxica. TB

Comentado [MAP345]: Omisión. TA

Comentado [MAP346]: Precisión léxica. TB

Comentado [MAP347]: Omisión. TA

Comentado [MAP348]: Preposición. TA

Comentado [MAP349]: Estructura natural. TA

Comentado [MAP350]: Precisión léxica. TA

Comentado [MAP351]: Preposición. TA

assigned value was 7, because it is in a moderate-to-high risk area. Residential group r1 was assigned a value of 3, since it is located in an urban area with moderate transit, where no cases were recorded.

For the variable Vulnerability to contagion due to the activity (D), the h1 hospital is one of those designated to treat people with coronavirus, while the h2 hospital is not. Therefore, a score of 10 was assigned to m1 and of 8 to m2 to weigh the risk of contagion. For the same reason, the risk of contagion for s1 was set at 6 and that of s2 at 4. Although the activity of g1 is not risky, it is a very vulnerable group, so it was assigned a value of 5. Group c1 is in a COVID-19-free area and has a value of 6 because there is a considerable flow of people who do not correspond to the area and a close contact with customers. Finally, group i1 develops an activity in minimal contact with external people, therefore, it was assigned a value of 3.

Continuing the analysis with the Intragroup contact variable (C), the m1 doctors have daily meetings in hospital 1 and with a medium frequency they work together and at distances of less than 2 meters. This assumption is also valid for m2 doctors at hospital 2. Hence, the intragroup contact for both cases was scored with 8. The contact between the individuals of the security groups (s1 and s2) occurs, mainly, at a distance less than 2 meters and with low frequency, in situations related to maintaining order, and thus, a value of 6 was assigned. The elderly and nursing home staff (g1) meet at distances between 1 and 3 meters and at a high frequency and, therefore, a value of 9 was assigned. The personnel from both c1 and i1 meet at medium frequency and distance, so a value of 6 was assigned in both cases. Finally, the residential group r1 meets at low distances and with high frequency, so it was assigned a value of 10.

Ending with the variables, the Intergroup contact (G) of the m1 and m2 doctors is at low frequency and at distances greater than 4 meters due to requests for special participation or in refresher courses, so the value for the intergroup relationship in this case is 2. This assumption is also valid for the relationship between security groups s1 and s2. The rest of the intergroup relationships are non-existent, so their value was set at 1 (table 5).

Assuming that in the last analysis carried out 6 days before the s1 group had tested positive (table 5), there would be two direct consequences. In the first place, that group should be sent to preventive isolation until the infected individual (or infected individuals) can be identified, for which the pool would have to be opened and the individual samples analyzed, and then proceed to the corresponding job replacements (fig. 1). Second, this would cause a change in the order of analysis for the other groups due to intergroup contact. In this example, security group s1 was not analyzed by subgroups, as sufficient availability of replacements is anticipated until s1 returns from quarantine. However, should it be done, the established steps should be followed (fig. 1).

- Comentado [MAP352]: Precisión léxica. TA
- Comentado [MAP353]: Precisión léxica. TA
- Comentado [MAP354]: Mayúscula. TA
- Comentado [MAP355]: Preposición. TA
- Comentado [MAP356]: Precisión léxica. TA
- Comentado [MAP357]: Preposición. TA
- Comentado [MAP358]: Precisión léxica. TA
- Comentado [MAP359]: Estructura natural. TA
- Comentado [MAP360]: Puntuación. TA
- Comentado [MAP361]: Puntuación. TA
- Comentado [MAP362]: Precisión léxica. TA
- Comentado [MAP363]: Precisión léxica. TA
- Comentado [MAP364]: Omisión. TA
- Comentado [MAP365]: Mayúscula. TA
- Comentado [MAP366]: Precisión léxica. TA
- Comentado [MAP367]: Estructura natural. TA
- Comentado [MAP368]: Puntuación. TA
- Comentado [MAP369]: Mayúscula. TA
- Comentado [MAP370]: Mayúscula. TA
- Comentado [MAP371]: Precisión léxica. TA
- Comentado [MAP372]: Concordancia. TA
- Comentado [MAP373]: Concordancia. TA
- Comentado [MAP374]: Puntuación. TA
- Comentado [MAP375]: Mayúscula. TA
- Comentado [MAP376]: Puntuación. TA
- Comentado [MAP377]: Estructura natural. TA
- Comentado [MAP378]: Puntuación. TA
- Comentado [MAP379]: Mayúscula. TA
- Comentado [MAP380]: Estructura natural. TA

It can be observed, then, that the analysis should be carried out first on the doctors of hospital 1 (m1; Oi = 100), followed by the group of the nursing home 1 (g1; Oi = 67), the doctors of hospital 2 (m2; Oi = 61), the hospital 2 security (s2; Oi = 33), the wholesale trade (c1; Oi = 18), the industry (i1; Oi = 9) and, finally, the residential group (r1; Oi = 3). It should be taken into account that the security group s1 would be on the sixth day of preventive isolation (quarantine), so it should be analyzed again once the virus incubation period is finished (table 5). Although the medical group of hospital 1 (m1) had intergroup contact with the infected security group (s1), 6 days had elapsed since their last analysis, while it was 12 days for group m2 of hospital 2. Furthermore, the fact that there are no group infections in hospital 2 does not imply that that hospital is not receiving sick people who could infect the groups analyzed in said hospital. If, on the other hand, the security group s1 had tested negative, the analysis should be carried out, first, to the nursing home group 1 (g1; Oi = 100), followed by the doctors from hospital 2 (m2; Oi = 91), the wholesale trade (c1; Oi = 26), the hospital 2 security (s2; Oi = 25), the doctors from hospital 1 (m1; Oi = 25), the industry (i1; Oi = 14), the security group s1 (s1; Oi = 7) and, finally, the residential group (r1, Oi = 4).

This example allows to visualize the potential for saving in tests when analyzing the spread of COVID-19 in a certain municipality. Up to 55 individuals distributed in 8 groups through 8 tests could be analyzed, which would mean a saving of 85.45% of the tests if the group analysis method is used, compared to the individual one.

In another research work where experiments were carried out through RT-qPCR, it was found that groups of up to 32 individuals can be analyzed, with 10% false negatives for the presence of SARS-CoV-2. On the other hand, Hogan et al. (2020) analyzed, using the RT-qPCR technique, samples stored in a hospital taken from people whose medical records showed symptoms of coronavirus infection, in groups of up to 10 individuals. Three of the 292 groups analyzed by these authors gave positive results for infection with SARS-CoV-2. With individual analyses, it was later found that two of the three groups were included infected people. However, as the quality of screening tests and protocols vary internationally, it is recommended to perform independent experiments before putting the proposal of this work into practice.

The presented algorithm allows establishing a rational order of priority to test critical groups in society (fig. 3). As a relevant limitation of this proposal, it should be noted that the assignment of the values of the group variables is based on the criteria and experience of the operator of the tool, in addition to the epidemiological characteristics analyzed. For this reason, if this tool is applied, its potential will increase as the infection progresses in the community and the knowledge of the particularities that it presents.

Define groups

Organize analysis

Analyze groups

GROUP 1

GROUP 2

GROUP 3

GROUP 4

GROUP i

General variables

General contagion function Number of tests available

ANALYSIS ORDER

GROUP 1

GROUP 2

Comentado [MAP381]: Preposición. TA

Comentado [MAP382]: Precisión léxica. TA

Comentado [MAP383]: Omisión. TA

Comentado [MAP384]: Orden. TA

Comentado [MAP385]: Omisión. TA

Comentado [MAP386]: Omisión. TA

Comentado [MAP387]: Omisión. TA

Comentado [MAP388]: Estructura natural. TA

Comentado [MAP389]: Mayúscula. TA

Comentado [MAP390]: Adición palabras. TA

Comentado [MAP391]: Precisión léxica. TA

Comentado [MAP392]: Omisión. TA

Comentado [MAP393]: Omisión. TA

Comentado [MAP394]: Orden. TA

Comentado [MAP395]: Omisión. TA

Comentado [MAP396]: Omisión. TA

Comentado [MAP397]: Estructura natural. TA

Comentado [MAP398]: Puntuación. TA

Comentado [MAP399]: Preposición. TB

Comentado [MAP400]: Precisión léxica. TA

Comentado [MAP401]: Precisión léxica. TB

Comentado [MAP402]: Mayúscula. TA

Comentado [MAP403]: Estructura natural. TA

Comentado [MAP404]: Precisión léxica. TA

Comentado [MAP405]: Precisión léxica. TA

Comentado [MAP406]: Estructura natural. TA

GROUP 3
 GROUP 4
 GROUP i
 Group variables
 Group priority
 Number of individuals
 Intragroup contact
 Vulnerability due to the activity
 Intergroup contact
 GROUP 2
 Composite sample
 GROUP 2 analysis
 Normal activity
 GROUP 4
 Composite sample
 GROUP 4 analysis
 Quarantine
 Feedback variables
 Vulnerability due to elapsed time
 Intergroup activation

Figure 3. Proposed work scheme, in which the steps to be followed are indicated with the variables corresponding to each stage.

CONCLUSIONS

Through this simulated analysis we demonstrate that the proposed algorithm is an interesting option and can be a tool to control the spread of SARS-CoV-2, so it deserves to be explored and, eventually, developed, with the ultimate goal of achieving a better management of social distancing using the least amount of the limiting resource (screening tests). This tool has a high potential to mitigate the problems generated by the pandemic, as well as to reduce the probability of resurgence when the resources are scarce, with the consequent reduction of the impact on health, society, and the economy of a region.

Comentado [MAP407]: Precisión léxica. TA

Comentado [MAP408]: Preposición. TA

Comentado [MAP409]: Precisión léxica. TA

Comentado [MAP410]: Precisión léxica. TA

Comentado [MAP411]: Preposición. TA

Comentado [MAP412]: Precisión léxica. TA

Comentado [MAP413]: Precisión léxica. TB

Comentado [MAP414]: Precisión léxica. TA

Comentado [MAP415]: Estructura natural. TA

Comentado [MAP416]: Puntuación. TA

ANEXO IV. POSEDICIÓN DEL TEXTO CONTROLADO

Proposals for the control and mitigation of the spread of COVID-19: a strategic management of the disease

ABSTRACT

The speed of spread of COVID-19 in the world caused the affected countries to close their borders and take social distancing measures. Six months after declaring the disease a pandemic, many countries took measures to ease the confinement. However, without a vaccine or a drug capable of curbing the SARS-CoV-2 infection, the situation could be reversed at any time. The objective of the present work was to propose a decision algorithm to optimize the detection of asymptomatic cases and to prescribe quarantine in a strategic way. Measures like these prevent the spread of the virus and allow progress towards a supervised normality. We developed a tentative proposal for the optimization and structuring of SARS-CoV-2 detection tests based on the analysis of composite samples from individual samples of asymptomatic people from the cohorts of interest. We defined cohorts according to their role in society or the degree of vulnerability. The algorithm considers variables such as the priority of the cohort, the number of members of the analysis groups within each cohort, intragroup and intergroup contact, vulnerability to contagion due to the activity carried out and the time elapsed since the test was last carried out. We illustrated the proposal with defined hypothetical cohorts and with a single analysis group for simplicity. We verified that the application of the tool allows to establish an order of priority to carry out the tests in critical groups of society in a rational way. This tool will enable optimizing resources and reducing the impact of the disease on health, society, and the economy of a region.

KEYWORDS: COVID-19, Social distancing, SARS-CoV-2, Decision algorithm, Quarantine

INTRODUCTION

Coronaviruses are a family of viruses that infect animals, mainly birds and mammals, in which they cause diseases and produce economic losses associated with productive activities. They can also jump from certain species and infect humans and cause diseases of varying severity (cold, bronchitis, pneumonia, etc.), depending on the virus and the health of the host. Most people are infected with these viruses sometime in their lives without serious consequences. However, some coronaviruses have caused diseases that have spread rapidly, leading to worrisome outbreaks and epidemics. Examples include: severe acute respiratory syndrome (SARS) that originated in Yunan, China, in 2001; Middle East respiratory syndrome (MERS) caused by MERS-CoV, started in Saudi Arabia in 2012 and currently COVID-19, caused by a new coronavirus, called SARS-CoV-2.

COVID-19 is believed to have emerged in December 2019 at a seafood market selling live and dead exotic animals located in Wuhan, Hubei province, China. However, there are some controversies about this theory. The speed of spread of the virus was so high that, in a short time, the authorities of China closed the city and ordered mandatory quarantine. Even with these measures, the contagion spread rapidly through neighboring countries first and, gradually, through the others, facilitated by the great mobility of travelers. Almost three months after detecting the disease, on March 11, 2020, the World Health Organization (WHO) declared it a pandemic.

The rapid advance of the disease highlighted the lack of preparation of the healthcare systems of most countries to face the situation. The number of sick people far exceeded available capacities and the world watched in amazement at the rising death toll in developed countries in Europe.

Comentado [MAP1]: Omisión. TA

Comentado [MAP2]: Precisión léxica. TA

Comentado [MAP3]: Precisión léxica. TA

Comentado [MAP4]: Tiempo verbal. TC

Comentado [MAP5]: Estructural natural. TA

Comentado [MAP6]: Tiempo verbal. TC

Comentado [MAP7]: Precisión léxica. TA

Comentado [MAP8]: Tiempo verbal. TC

Comentado [MAP9]: Tiempo verbal. TA

Comentado [MAP10]: Puntuación. TA

Comentado [MAP11]: Precisión léxica. TB

Given this **scenario**, and in order to contain the contagion, in most countries social distancing measures (better known as quarantine) were established, with different modalities. As of the writing of this article (May 19, 2020), the number of confirmed cases in the world was rapidly approaching five million, of which approximately 7% corresponded to deaths. As of the manuscript correction date, September 15, 2020, there are almost 30 million confirmed cases in the world and the number of deaths is close to one million.

Early **on** (March 20, 2020), the government of **Argentina** established **the preventive** and compulsory social **distancing** (aislamiento social preventivo y obligatorio, ASPO). This measure made it possible to avoid massive contagion and to buy time so that the **healthcare** system could improve its response capacity. Thus, new hospitals or care centers were set up and respirators, personal protection elements and other products were acquired and manufactured to be able to attend to the emergency. The scientific community was also urged to contribute to the generation of knowledge to face the pandemic.

As in the rest of the world, the detection of SARS-CoV-2 is carried out in Argentina using real-time reverse transcription polymerase chain reaction (RT-qPCR) tests. In principle, this test is performed on each individual suspected of being a carrier because they present symptoms related to the disease or due to an epidemiological connection. But these protocols have been changing with the evolution of the pandemic. Other molecular tests recently developed by **Argentinian** researchers allow detecting the presence of genetic fragments with promising results: these are the NEOKIT and the ELA-CHEMSTRIP. The advantage of the latter methods is that they allow the virus to be identified more quickly and make it possible to detect it in a larger population group. However, in most developing countries the capacity to perform such **analyses** is limited. The presumed healthy or asymptomatic population cannot be assessed (although it may include contagious individuals) and widespread measures must be taken to control the spread of the virus. Due to the current **healthcare** context and the fact that social distancing has been effective in slowing down the spread of the disease, it is necessary to return to activity gradually and in a controlled manner. However, should the **healthcare** situation deteriorate, the easing measures could be reversed. The objective of the present work was to propose a decision algorithm to optimize the detection of asymptomatic cases and to prescribe quarantine in a strategic way, to prevent the spread of the virus and to allow a new supervised normality.

MATERIALS AND METHODS

We developed a tentative tool to strategically use SARS-CoV-2 virus testing to monitor the spread of the disease in multiple cohorts. To this end, as already indicated by other authors, we **suggested** analyzing composite samples (**pool analysis**) and combining individual samples from different asymptomatic people **of** the cohort of interest. This methodology would make it possible to evaluate the presence of the virus in a greater number of people simultaneously, with a considerable saving **in** resources and time. We proposed to analyze first the cohorts of people whose function was critical to the normal functioning of society or who belonged to highly vulnerable sectors. The analysis would continue with the non-critical cohorts, according to their importance and degree of vulnerability.

Definition of critical cohorts

The cohorts that we propose here as **critical**, and that should be the first to undergo the virus **detection**, would cover the following **sectors**: **healthcare**, **security**, **trade of essential consumer goods**, the elderly (nursing **home**), essential services and industries (manufacture of **respirators**,

Comentado [MAP12]: Puntuación. TA

Comentado [MAP13]: Estructura natural. TA

Comentado [MAP14]: Precisión léxica. TA

Comentado [MAP15]: Omisión. TA

Comentado [MAP16]: Puntuación. TA

Comentado [MAP17]: Precisión léxica. TB

Comentado [MAP18]: Convención. TA

Comentado [MAP19]: Precisión léxica. TB

Comentado [MAP20]: Precisión léxica. TA

Comentado [MAP21]: Precisión léxica. TA

Comentado [MAP22]: Precisión léxica. TB

Comentado [MAP23]: Precisión léxica. TB

Comentado [MAP24]: Tiempo verbal. TC

Comentado [MAP25]: Precisión léxica. TA

Comentado [MAP26]: Preposición. TA

Comentado [MAP27]: Preposición. TA

Comentado [MAP28]: Puntuación. TA

Comentado [MAP29]: Estructura natural. TA

Comentado [MAP30]: Puntuación. TA

Comentado [MAP31]: Precisión léxica. TA

Comentado [MAP32]: Precisión léxica. TB

Comentado [MAP33]: Precisión léxica. TA

Comentado [MAP34]: Preposición. TA

Comentado [MAP35]: Precisión léxica. TA

Comentado [MAP36]: Adición de espacio. TA

food, etc.). It is essential to study the population of these sectors, as they are potential vectors of contagion. Healthcare personnel (doctors, nurses, cleaning personnel, etc.) should be kept in optimal condition and, in addition to providing them with personal protection elements, they should be subject to systematic and periodic tests. These protocols are also necessary for security personnel, as they are highly exposed to potentially infected people. Supply chain traders, including large fruit and vegetable markets, should be tested regularly as well. In this sector, an infected person could bring the entire chain to a standstill and generate serious social problems. Other sectors, such as construction workers, administrative workers, traders of non-essential goods or inhabitants of residential areas, to name a few, correspond to non-critical sectors. These groups should be analyzed whenever the number of available tests is sufficient, without compromising the control of the sectors that are critical. However, those non-critical cohorts that are related to individuals of critical cohorts should have a higher priority for analysis.

Carrying out the analyses

After defining the groups of the different cohorts, we propose taking individual samples (nasopharyngeal swabs) from all the asymptomatic people in these groups. A part of the sample could be mixed in the laboratory to create a composite sample from each group for analysis. Initially, we should estimate the infectivity rate and prevalence from the official contagion data in the analyzed geographic location. These values would allow defining the group size and the detection methodology.

Pool sample analysis makes sense to save resources and time, as long as the prevalence is lower than 10%. Beyond this value, it is convenient to carry out individual analyses due to the high number of groups that would test positive (the study of individual samples would be necessary). The size of the groups would be defined by the prevalence: the higher the prevalence, the smaller the number of samples in a group. We considered the prevalence of the disease in the population, and also in the cohorts comprised of essential personnel, as known epidemiological information. The prevalence could be calculated from the confirmed (individual) cases in the groups of interest relative to the population, or a defined population size.

- Comentado [MAP37]: Precisión léxica. TB
- Comentado [MAP38]: Precisión léxica. TA
- Comentado [MAP39]: Estructura natural. TA
- Comentado [MAP40]: Precisión léxica. TA
- Comentado [MAP41]: Estructura natural. TA
- Comentado [MAP42]: Precisión léxica. TC
- Comentado [MAP43]: Precisión léxica. TA
- Comentado [MAP44]: Precisión léxica. TB
- Comentado [MAP45]: Tiempo verbal. TA
- Comentado [MAP46]: Estructura natural. TA
- Comentado [MAP47]: Precisión léxica. TA
- Comentado [MAP48]: Tiempo verbal. TA
- Comentado [MAP49]: Precisión léxica. TB
- Comentado [MAP50]: Estructura natural. TA
- Comentado [MAP51]: Precisión léxica. TA
- Comentado [MAP52]: Tiempo verbal. TA
- Comentado [MAP53]: Omisión. TA
- Comentado [MAP54]: Precisión léxica. TA
- Comentado [MAP55]: Omisión. TA
- Comentado [MAP56]: Precisión léxica. TA
- Comentado [MAP57]: Estructura natural. TA
- Comentado [MAP58]: Estructura natural. TA
- Comentado [MAP59]: Precisión léxica. TA
- Comentado [MAP60]: Preposición. TA
- Comentado [MAP61]: Precisión léxica. TA
- Comentado [MAP62]: Tiempo verbal. TC
- Comentado [MAP63]: Puntuación. TA
- Comentado [MAP64]: Precisión léxica. TA
- Comentado [MAP65]: Puntuación. TA
- Comentado [MAP66]: Estructura natural. TA
- Comentado [MAP67]: Puntuación. TA

It should be clarified that, before starting the detection through pools of samples, some critical methodological aspects should be evaluated to ensure the success of this diagnostic modality. The first issue to verify through bioinformatics is that the oligonucleotide sequences for detection are specific. Then, it would be necessary to determine the efficiency of the amplification of the RT-qPCR reaction and verify that the efficiency is at maximum, even at low concentrations of RNA. In addition, it would be essential to determine the detection limit of the method used, since this variable affects the classification of the results as positive and not detectable (negative). Finally, we should validate the pool technique in a group of the population, and compare the results with those obtained by individual analyses. This comparison allows determining the probability of misclassification of a result. Only after having all the aforementioned knowledge, could we begin with the detection in pools of samples.

The maximum number of individuals in a group should be defined according to the volume of individual sample to add to the detection reaction without compromising the sensitivity, while allowing sufficient viral load to reach it. In addition, we must consider the prevalence of the virus or the ratio of infected individuals to total population analyzed to optimize the size of the groups from a statistical point of view. These values would serve as a guide, but the definition of the groups should be done in a practical manner. If the optimal group size was 15 people and in a nursing home there were 10 people, it would not be necessary or convenient to add external people to reach the optimal size. On the other hand, when analyzing the cohort of an entire neighborhood, the groups should have the optimal size and be selected by proximity to save resources.

Groups with negative results in SARS-CoV-2 tests should be tested periodically, and the procedure should be optimized according to an algorithm that resolves the order of priority of each group. If any of the tests carried out are positive, the group and the people with whom they have had close contact should be isolated to avoid contagion until the analysis is completed. The laboratory should study the positive pools and focus on the analysis of the individual samples to identify the infected individual, who should then remain isolated (fig. 1). People with negative individual results could come out of isolation.

Define the cohorts to analyze

Take samples of the individuals in the cohort and combine them into a pool sample

Perform pool sample analysis

Calculate waiting time according to cohort priority

Negative

Pool analysis result

Positive

Maintain alert and distancing for 2 weeks

Analyze samples individually or by subgroups

Negative

Individual analysis result

Positive

Figure 1. Flow chart that indicates the steps to follow to mitigate the spread of the disease in the population and optimize the use of diagnostic resources.

- Comentado [MAP68]: Precisión léxica. TA
- Comentado [MAP69]: Precisión léxica. TC
- Comentado [MAP70]: Estructura natural. TA
- Comentado [MAP71]: Puntuación. TA
- Comentado [MAP72]: Puntuación. TA
- Comentado [MAP73]: Estructura natural. TA
- Comentado [MAP74]: Precisión léxica. TA
- Comentado [MAP75]: Puntuación. TA
- Comentado [MAP76]: Precisión léxica. TA
- Comentado [MAP77]: Adición palabras. TA
- Comentado [MAP78]: Tiempo verbal. TA
- Comentado [MAP79]: Precisión léxica. TA
- Comentado [MAP80]: Preposición. TA
- Comentado [MAP81]: Omisión. TA
- Comentado [MAP82]: Estructura natural. TA
- Comentado [MAP83]: Precisión léxica. TA
- Comentado [MAP84]: Precisión léxica. TB
- Comentado [MAP85]: Estructura natural. TA
- Comentado [MAP86]: Estructura natural. TA
- Comentado [MAP87]: Estructura natural. TA
- Comentado [MAP88]: Precisión léxica. TB
- Comentado [MAP89]: Estructura natural. TA
- Comentado [MAP90]: Precisión léxica. TA
- Comentado [MAP91]: Precisión léxica. TB
- Comentado [MAP92]: Mayúscula. TA
- Comentado [MAP93]: Precisión léxica. TA
- Comentado [MAP94]: Precisión léxica. TA
- Comentado [MAP95]: Estructura natural. TA
- Comentado [MAP96]: Precisión léxica. TA
- Comentado [MAP97]: Precisión léxica. TA
- Comentado [MAP98]: Precisión léxica. TA
- Comentado [MAP99]: Estructura natural. TA

Resource Optimization

We developed a decision algorithm to optimize the analysis and establish a logical order to carry it out. The algorithm allows modeling and evaluating the behavior of the transmission of the disease based on the impact of the defined cohorts on the health, the safety and the economy. Ideally, a deep knowledge of the population of the analysis cohorts is necessary, so the implementation of this algorithm at the municipal level would be more beneficial.

We considered it necessary, although not essential, to include the following general variables in the model, since they affect all groups in all cohorts, and these variables should be updated daily:
Number of tests available: determines the number of groups on which the analysis can be performed and, therefore, the number of cohorts that can be periodically analyzed.

General contagion function: defined as the relationship between the population susceptible to infection, the infected population, and the recovered population. This variable is applied in an area that encompasses the cohorts analyzed, and depends on the historical values of these variables.

Other important variables are related to particular characteristics of the group to be analyzed:

Location/coordinates: allows to know the location of the group and can be useful to determine the level of intergroup contact or help decide the vulnerability to contagion due to the activity. For example, there might be an increased risk of contagion for a security group that shares the location with a group in a hospital where there are individuals that tested positive for SARS-CoV-2.

Group priority level (A): refers to the priority of one group over the others to be subject to analysis. For example, the groups in the healthcare sector cohort (doctors and nurses dedicated to the treatment of SARS-CoV-2) would have analysis priority over a group in a cohort of people dedicated to the production of consumer goods. To make this decision as less subjective as possible, we defined a scale from 1 to 10, in which 1 means no priority and 10 indicates highest priority (table 1). We established specific ranges for each hypothetical group included in the algorithm.

Table 1. Group priority scale for analysis

Group priority (A)

Range

Doctors dedicated to the direct treatment of those infected by SARS-CoV-2

Other doctors

Population of a nursing home

Security personnel

Consumer goods traders

Industry

Residential groups

- Comentado [MAP100]: Tiempo verbal. TC
- Comentado [MAP101]: Cambio de sentido. TA
- Comentado [MAP102]: Estructura natural. TA
- Comentado [MAP103]: Omisión. TA
- Comentado [MAP104]: Omisión. TA
- Comentado [MAP105]: Precisión léxica. TB
- Comentado [MAP106]: Tiempo verbal. TA
- Comentado [MAP107]: Tiempo verbal. TC
- Comentado [MAP108]: Precisión léxica. TB
- Comentado [MAP109]: Preposición. TA
- Comentado [MAP110]: Precisión léxica. TA
- Comentado [MAP111]: Preposición. TA
- Comentado [MAP112]: Puntuación. TA
- Comentado [MAP113]: Puntuación. TA
- Comentado [MAP114]: Adición de espacio. TA
- Comentado [MAP115]: Adición de espacio. TA
- Comentado [MAP116]: Preposición. TA
- Comentado [MAP117]: Estructura natural. TA
- Comentado [MAP118]: Estructura natural. TA
- Comentado [MAP119]: Cambio de sentido. TA
- Comentado [MAP120]: Preposición. TA
- Comentado [MAP121]: Precisión léxica. TB
- Comentado [MAP122]: Preposición. TA
- Comentado [MAP123]: Tiempo verbal. TC
- Comentado [MAP124]: Precisión léxica. TB
- Comentado [MAP125]: Puntuación. TA
- Comentado [MAP126]: Puntuación. TA
- Comentado [MAP127]: Mayúscula. TA
- Comentado [MAP128]: Precisión léxica. TA

Number of individuals of each group (B): the number of individuals of each group would be defined for practical reasons. Test negatives must be true negatives (so that they do not compromise the sample detection limit), to prevent the spread of the infection. If the analysis of an infected group is done through RT-qPCR, the sensitivity will depend on the phase of the ongoing infection, the quality of the samples extracted, and the number of individual samples of the composite sample (the greater the number of samples, the less amount of each target, and less sensitivity).

Intragroup contact (C): this variable indicates the level of contact (related to social distance and the frequency of contact) between the individuals of the group analyzed. For its evaluation we established a scale between 1 and 10, where high values indicate more contact, and low values less contact (we assigned the value 1 to null contact). Contact values for different distances may coincide due to changes in contact frequency (table 2).

Table 2. Scale for intragroup contact (C)

Distance (m)

Frequency

Range

Low

Medium

High

Low

Medium

High

Low

Medium

H

Vulnerability to contagion due to the activity (D): refers to the degree of vulnerability of the groups studied within a cohort due to the activity. In this case, we also established a scale between 1 and 10, in which a higher value indicates a higher level of vulnerability to contagion due to the activity (table 3). The values could be obtained from studies of the level of contagion in these groups and should be updated periodically.

Table 3. Scale for vulnerability to contagion (D)

Activity

Value

Family doctors

Specialized nurses

Pharmacy technicians

Prison warden

Residential

Vulnerability due to elapsed time (E): calculated value (equation 1) between 0 and 10, which indicates the risk of an undetected infection during the elapsed time.

where t (in days) is the time elapsed since the last test performed and t_i is the maximum incubation period of the virus (14 days).

Intergroup activation (F): value used to indicate whether the order of priority will increase (or not) as a consequence of contact with a group whose analysis was positive. That is, because of the intergroup relationship with infected groups, the analyzed group will increase its order of priority

- Comentado [MAP129]: Precisión léxica. TB
- Comentado [MAP130]: Precisión léxica. TA
- Comentado [MAP131]: Estructura natural. TA
- Comentado [MAP132]: Estructura natural. TA
- Comentado [MAP133]: Precisión léxica. TA
- Comentado [MAP134]: Omisión. TA
- Comentado [MAP135]: Omisión. TA
- Comentado [MAP136]: Preposición. TA
- Comentado [MAP137]: Omisión. TA
- Comentado [MAP138]: Estructura natural. TA
- Comentado [MAP139]: Puntuación. TA
- Comentado [MAP140]: Omisión. TA
- Comentado [MAP141]: Puntuación. TA
- Comentado [MAP142]: Precisión léxica. TB
- Comentado [MAP143]: Puntuación. TA
- Comentado [MAP144]: Tiempo verbal. TC
- Comentado [MAP145]: Estructura natural. TA
- Comentado [MAP146]: Estructura natural. TA
- Comentado [MAP147]: Precisión léxica. TA
- Comentado [MAP148]: Precisión léxica. TA
- Comentado [MAP149]: Precisión léxica. TA
- Comentado [MAP150]: Mayúscula. TA
- Comentado [MAP151]: Precisión léxica. TA
- Comentado [MAP152]: Precisión léxica. TA
- Comentado [MAP153]: Mayúscula. TA
- Comentado [MAP154]: Precisión léxica. TA
- Comentado [MAP155]: Precisión léxica. TA
- Comentado [MAP156]: Mayúscula. TA
- Comentado [MAP157]: Preposición. TA
- Comentado [MAP158]: Mayúscula. TA
- Comentado [MAP159]: Preposición. TA

and, therefore, this could lead to it being analyzed first. F will take a value of 1 if the test is positive and a value of 0 if it is negative (the intergroup relationship is not activated).

Intergroup contact (G): variable that establishes the level of contact between different groups (from the same cohort or from a different cohort). We defined a scale between 1 and 10 (we assigned the value of 1 to null contact). Higher values indicate closer contact between individuals of different groups (the relationship of group i to group j is defined as relationship ij, with $i \neq j$). Contact values for different distances may coincide due to changes in the contact frequency (table 4).

Table 4. Scale for intergroup contact (G)

Distance (m)

Frequency

Range

Low

Medium

High

Low

Medium

High

Low

Medium

High

General analysis order (O): value whose result would allow structuring the test priority according to the current model. First, taking into account the previously established variables, we calculate a local value for each group (L_i , equation 2) and identify the highest value (L_{max}). Local values would represent relevance for analysis and do not have a defined scale. Then, to better visualize which groups should be analyzed first, we will weigh the local values with respect to the maximum local value (L_{max}) to obtain the general analysis priority order for each group (O_i , equation 3).

where j varies from 1 to the maximum value of j (total number of defined groups), with $i \neq j$:

Once the general analysis order of the groups (O) is defined, we will be able to perform the detection following that order (from highest to lowest) and update the number of tests available. Then, we feed back the information obtained to the algorithm with the variables E (vulnerability due to elapsed time) and F (intergroup activation) (fig. 2).

GROUP j

GROUP i

(G) Intergroup contact

Structure analysis

GROUP VARIABLES

Location / coordinates

(A) Group priority

(B) Number of individuals

(C) Intragroup contact

(D) Vulnerability to contagion due to the activity

(O) Analysis order

Analyze groups

FEEDBACK VARIABLES

(E) Vulnerability due to elapsed time

(F) Intergroup activation

Comentado [MAP160]: Estructura natural. TA

Comentado [MAP161]: Tiempo verbal. TC

Comentado [MAP162]: Tiempo verbal. TC

Comentado [MAP163]: Estructura natural. TA

Comentado [MAP164]: Precisión léxica. TB

Comentado [MAP165]: Precisión léxica. TA

Comentado [MAP166]: Precisión léxica. TA

Comentado [MAP167]: Estructura natural. TA

Comentado [MAP168]: Omisión. TA

Comentado [MAP169]: Precisión léxica. TA

Comentado [MAP170]: Precisión léxica. TA

Comentado [MAP171]: Precisión léxica. TA

Comentado [MAP172]: Mayúscula. TA

Comentado [MAP173]: Precisión léxica. TA

Comentado [MAP174]: Precisión léxica. TA

Comentado [MAP175]: Mayúscula. TA

Comentado [MAP176]: Precisión léxica. TA

Comentado [MAP177]: Precisión léxica. TA

Comentado [MAP178]: Mayúscula. TA

Comentado [MAP179]: Estructura natural. TA

Comentado [MAP180]: Precisión léxica. TA

Comentado [MAP181]: Estructura natural. TA

Comentado [MAP182]: Precisión léxica. TA

Comentado [MAP183]: Preposición. TB

Comentado [MAP184]: Precisión léxica. TA

Comentado [MAP185]: Precisión léxica. TA

Comentado [MAP186]: Estructura natural. TA

Comentado [MAP187]: Tiempo verbal. TA

Comentado [MAP188]: Tiempo verbal. TA

Comentado [MAP189]: Mayúscula. TA

Comentado [MAP190]: Precisión léxica. TA

Comentado [MAP191]: Precisión léxica. TB

Comentado [MAP192]: Preposición. TA

Comentado [MAP193]: Estructura natural. TA

GENERAL VARIABLES

Number of tests available

General contagion function

Figure 2. General algorithm for calculating the analysis order of the groups and making rational use of the available screening tests. The white boxes indicate the input and output variables of the model, while the blue blocks are the operations.

Time scope and costs

According to the WHO, this pandemic is expected to last for at least two years. Although there are many projects for vaccines and palliative treatments, none have yet been implemented effectively and massively, so the duration of the pandemic is unpredictable. The application timeline of the tool proposed in this work is difficult to establish, as it depends on many variables and is closely related to the duration of the pandemic. The tool is very flexible: it allows to operate with various numbers of cohorts (depending on the requirements or the budget) and allows carrying out analysis at a convenient frequency for each location. The cost of implementation will depend, mainly, on the available analysis capacity. Another variable that could influence costs is the possible implementation of rapid tests, which will be available (mainly for the healthcare system) in the coming months. It is important to note that the analysis of pooled samples allows obtaining a large amount of information and minimizes costs by analyzing many individual samples in a single reaction.

RESULTS AND DISCUSSION

To illustrate the application of the algorithm and evaluate its versatility and feasibility, we performed a simulation. Since we did not have real data when formulating this algorithm, it was necessary to define some data arbitrarily (trying to get closer to reality) and make different assumptions.

First, we defined eight hypothetical groups for illustrative purposes, since in reality there will be many more: GROUP 1, doctors belonging to hospital 1 (m1); GROUP 2, hospital 1 security personnel (s1); GROUP 3, doctors from hospital 2 (m2); GROUP 4, hospital 2 security personnel (s2); GROUP 5, population of a nursing home (elderly and staff) found in an urban area (g1), under the assumption that there were no cases in the area; GROUP 6, personnel of the wholesale trade 1 of consumer goods, located in an area where there are other wholesale shops (c1); GROUP 7, goods manufacturing industry located in an area where there were cases of COVID-19 (i1); GROUP 8, large family from a residential neighborhood, located in an urban area where there were no cases of COVID-19 (r1). We established a specific number of members for each group (tables 5 and 6).

Table 5. Hypothetical example of application of the developed decision algorithm

Groups

Group priority (A)

Number of members of each group (B)

Intragroup contact (C)

Vulnerability to contagion due to the activity (D)

Days elapsed since the analysis

Vulnerability due to elapsed time (E) a

Local analysis order (Li) b

General analysis order (Oi) c

Test priority order

Comentado [MAP194]: Precisión léxica. TA

Comentado [MAP195]: Estructura natural. TA

Comentado [MAP196]: Precisión léxica. TA

Comentado [MAP197]: Precisión léxica. TB

Comentado [MAP198]: Estructura natural. TA

Comentado [MAP199]: Estructura natural. TA

Comentado [MAP200]: Estructura natural. TA

Comentado [MAP201]: Precisión léxica. TB

Comentado [MAP202]: Estructura natural. TA

Comentado [MAP203]: Tiempo verbal. TC

Comentado [MAP204]: Orden. TA

Comentado [MAP205]: Orden. TA

Comentado [MAP206]: Preposición. TA

Comentado [MAP207]: Omisión. TA

Comentado [MAP208]: Preposición. TA

Comentado [MAP209]: Estructura natural. TA

Comentado [MAP210]: Mayúscula. TA

Comentado [MAP211]: Mayúscula. TA

a, b, c Calculated through equations 1, 2 and 3.

Defined groups. m1: doctors from hospital 1; s1: hospital 1 security; m2: doctors from hospital 2; s2: hospital 2 security; g1: nursing home 1 population; c1: store 1 personnel; i1: industry 1 personnel; r1: residents of residential area 1. The group suspected of being infected is marked in bold.

Table 6. Hypothetical example to illustrate a matrix of intergroup relationships

Groups

F: intergroup activation (1: infected, 0: not infected); NA: intergroup contact is not a variable applicable to a group's relationship with itself.

Defined groups. m1: doctors from hospital 1; s1: hospital 1 security; m2: doctors from hospital 2; s2: hospital 2 security.

The group suspected of being infected is marked in bold.

Regarding the variable Group priority (A), we assigned a value of 10 to the doctors from hospital 1 (m1), which registers the largest number of patients with COVID-19, and 9 to the doctors from hospital 2 (m2). The high values assigned are due to the fact that both are priority groups (they belong to the Hospital doctors cohort) whose function is to care for people with COVID-19. For the groups in the Safety cohort, we established a value of 6 for both s1 and s2. Although this group is essential personnel, they could be replaced if contagion occurs. The nursing home population g1 is a high-risk group because the mortality due to contagion is much higher, so we assigned it a value of 8. Wholesale trade of consumer goods c1 has a priority of 4 (relatively low), since there are other alternative wholesale businesses in the area. For industry i1, the assigned value was 7, because it is in a moderate-to-high risk area. We assigned a value of 3 to residential group r1, since it is located in an urban area with moderate transit, where there were no cases.

Comentado [MAP212]: Preposición. TA

Comentado [MAP213]: Orden. TA

Comentado [MAP214]: Orden. TA

Comentado [MAP215]: Precisión léxica. TA

Comentado [MAP216]: Orden. TA

Comentado [MAP217]: Orden. TA

Comentado [MAP218]: Orden. TA

Comentado [MAP219]: Orden. TA

Comentado [MAP220]: Consistencia. TA

Comentado [MAP221]: Mayúscula. TA

Comentado [MAP222]: Cambio de sentido. TA

Comentado [MAP223]: Puntuación. TA

Comentado [MAP224]: Adición espacio. TA

Comentado [MAP225]: Mayúscula. TA

Comentado [MAP226]: Puntuación. TA

Comentado [MAP227]: Puntuación. TA

Comentado [MAP228]: Mayúscula. TA

Comentado [MAP229]: Precisión léxica. TA

Comentado [MAP230]: Omisión. TA

Comentado [MAP231]: Preposición. TA

Comentado [MAP232]: Tiempo verbal. TC

Comentado [MAP233]: Preposición. TA

Comentado [MAP234]: Precisión léxica. TA

Comentado [MAP235]: Precisión léxica. TA

For the variable Vulnerability to contagion due to the activity (D), the h1 hospital is one of the centers designated to treat people with coronavirus, while the h2 hospital is not. Therefore, we assigned a score of 10 to m1 and 8 to m2 to weigh the risk of contagion. For the same reason, we set the contagion risk of s1 at 6 and that of s2 at 4. Although the activity of g1 is not risky, it is a very vulnerable group and we assigned it a value of 5. Group c1 is located in a COVID-19-free area and has a value of 6 because there is a considerable flow of people who do not correspond to the area and a close contact with customers. Finally, group i1 develops an activity in minimal contact with external people, therefore we assigned it a value of 3.

Regarding the variable Intragroup contact (C), m1 doctors have daily meetings in hospital 1 and with a medium frequency they work together and at distances of less than 2 meters. This assumption is also valid for m2 doctors at hospital 2. For this reason, we scored the intragroup contact for both cases with 8. Contact between the individuals of the security groups (s1 and s2) occurs mainly at a distance less than 2 meters and with low frequency, in situations related to maintaining order. Therefore, we assigned a value of 6. The elderly and nursing home staff (g1) meet at distances between 1 and 3 meters with a high frequency, so we assigned a value of 9. The staff of both c1 and of i1 meets at medium frequency and distance, so we assigned a value of 6 in both cases. Finally, the residential group r1 meets at low distances and with high frequency, so we assigned it a value of 10.

To finish with the variables, the Intergroup contact (G) of doctors m1 and m2 is at low frequency and distances greater than 4 meters. The contact occurs due to requests for special participation or refresher courses, so the value for the intergroup relationship in this case is 2. This assumption is also valid for the relationship between security groups s1 and s2. The rest of the intergroup relationships are non-existent, so we set their value at 1 (table 5).

A positive in group s1 in the last analysis carried out 6 days before (table 5) would have two direct consequences. First, this group should be sent to preventive isolation until the infected individual (or infected individuals) is identified. Therefore, it would be necessary to analyze the individual samples and then proceed to the corresponding job replacements (fig. 1). Second, this would cause a change in the order of analysis for the other groups due to intergroup contact. In this example we did not analyze security group s1 by subgroups, as there are sufficient replacements available until s1 returns from quarantine. However, if the analysis is performed, the established steps should be followed (fig. 1).

- Comentado [MAP236]: Puntuación. TA
- Comentado [MAP237]: Mayúscula. TA
- Comentado [MAP238]: Omisión. TA
- Comentado [MAP239]: Tiempo verbal. TC
- Comentado [MAP240]: Precisión léxica. TA
- Comentado [MAP241]: Adición. TA
- Comentado [MAP242]: Puntuación. TA
- Comentado [MAP243]: Precisión léxica. TA
- Comentado [MAP244]: Omisión. TA
- Comentado [MAP245]: Tiempo verbal. TC
- Comentado [MAP246]: Mayúscula. TA
- Comentado [MAP247]: Puntuación. TA
- Comentado [MAP248]: Precisión léxica. TA
- Comentado [MAP249]: Tiempo verbal. TC
- Comentado [MAP250]: Preposición. TA
- Comentado [MAP251]: Tiempo verbal. TC
- Comentado [MAP252]: Tiempo verbal. TC
- Comentado [MAP253]: Precisión léxica. TA
- Comentado [MAP254]: Tiempo verbal. TC
- Comentado [MAP255]: Tiempo verbal. TC
- Comentado [MAP256]: Mayúscula. TA
- Comentado [MAP257]: Puntuación. TA
- Comentado [MAP258]: Mayúscula. TA
- Comentado [MAP259]: Tiempo verbal. TC
- Comentado [MAP260]: Estructura natural. TA
- Comentado [MAP261]: Mayúscula. TA

We can observe, then, that it would be necessary to first carry out the analysis of the doctors of hospital 1 (m1; Oi = 100), followed by the group of the nursing home 1 (g1; Oi = 67), the doctors of hospital 2 (m2; Oi = 61), the hospital 2 security (s2; Oi = 33), the wholesale trade (c1; Oi = 18), the industry (i1; Oi = 9) and, finally, the residential group (r1; Oi = 3). We must bear in mind that the security group s1 would be on the sixth day of preventive isolation (quarantine), so we should analyze it again once the virus incubation period is finished (table 5). Although the medical group from hospital 1 (m1) had intergroup contact with the infected security group (s1), 6 days elapsed since their last analysis. The time elapsed for the m2 group from hospital 2 was 12 days. In addition, the fact that there are no group infections in hospital 2 does not imply that that hospital is not receiving sick people who could infect the groups analyzed. On the other hand, if the security group s1 tested negative, we should carry out the analysis to the nursing home group 1 (g1; Oi = 100) first, followed by the doctors from hospital 2 (m2; Oi = 91), the wholesale trade (c1; Oi = 26), the hospital 2 security (s2; Oi = 25), the hospital 1 doctors (m1; Oi = 25), the industry (i1; Oi = 14), the security group s1 (s1; Oi = 7) and, finally, the residential group (r1; Oi = 4).

This example allows us to visualize the potential for saving in tests when we analyze the spread of COVID-19 in a certain municipality. We could analyze up to 55 individuals distributed in 8 groups with 8 tests, which would mean a saving of 85.45% of the tests with respect to the analysis of individual samples.

In another research work with experiments through RT-qPCR, the authors determined that it is possible to analyze groups of up to 32 individuals, with 10% false negatives for the presence of SARS-CoV-2. On the other hand, Hogan et al. (2020) analyzed the samples stored in a hospital from people whose medical records included symptoms of coronavirus infection, in groups of up to 10 individuals, using the RT-qPCR technique. Three of the 292 groups analyzed by these authors gave positive results for infection with SARS-CoV-2. With individual analyses, they found that two of the three groups included infected people. However, as the quality of detection tests and protocols vary internationally, it is advisable to carry out independent experiments before implementing the proposal in this work.

The presented algorithm allows establishing a rational order of priority to perform the analysis of critical groups in society (fig. 3). As a relevant limitation of this proposal, it should be noted that the assignment of the values of the group variables is based on the criteria and experience of the operator of the tool, in addition to the epidemiological characteristics analyzed. For this reason, if this tool is applied, its potential will increase as the infection in the community progresses and the knowledge of its particularities advances.

Define groups

Organize analysis

Analyze groups

GROUP 1

GROUP 2

GROUP 3

GROUP 4

GROUP i

General variables

General contagion function

Number of tests available

ANALYSIS ORDER

GROUP 1

Comentado [MAP262]: Estructura natural. TA

Comentado [MAP263]: Precisión léxica. TA

Comentado [MAP264]: Omisión. TA

Comentado [MAP265]: Orden. TA

Comentado [MAP266]: Omisión. TA

Comentado [MAP267]: Omisión. TA

Comentado [MAP268]: Omisión. TA

Comentado [MAP269]: Estructura natural. TA

Comentado [MAP270]: Mayúscula. TA

Comentado [MAP271]: Tiempo verbal. TA

Comentado [MAP272]: Precisión léxica. TA

Comentado [MAP273]: Precisión léxica. TA

Comentado [MAP274]: Precisión léxica. TA

Comentado [MAP275]: Omisión. TA

Comentado [MAP276]: Omisión. TA

Comentado [MAP277]: Omisión. TA

Comentado [MAP278]: Omisión. TA

Comentado [MAP279]: Estructura natural. TA

Comentado [MAP280]: Precisión léxica. TA

Comentado [MAP281]: Precisión léxica. TA

Comentado [MAP282]: Preposición. TA

Comentado [MAP283]: Mayúscula. TA

Comentado [MAP284]: Estructura natural. TA

Comentado [MAP285]: Precisión léxica. TA

Comentado [MAP286]: Precisión léxica. TB

Comentado [MAP287]: Estructura natural. TA

GROUP 2
 GROUP 3
 GROUP 4
 GROUP i
 Group variables
 Group priority
 Number of individuals
 Intragroup contact
 Vulnerability due to the activity
 Intergroup contact
 GROUP 2
 Composite sample
 GROUP 2 analysis
 Normal activity
 GROUP 4
 Composite sample
 GROUP 4 analysis
 Quarantine
 Feedback variables
 Vulnerability due to elapsed time
 Intergroup activation

Figure 3. Proposed work scheme of the steps to follow with the variables corresponding to each stage.

CONCLUSIONS

Through this simulated analysis we demonstrate that the proposed algorithm is an interesting option and can be a tool to control the spread of SARS-CoV-2. We consider that this proposal deserves to be explored and developed in order to improve the management of social distancing and optimize the use of limiting resources (screening tests). This tool has a high potential to mitigate the problems generated by the pandemic and reduce the probability of resurgence when the resources are scarce, with the resulting reduction of the impact on health, society, and the economy of a region.

Comentado [MAP288]: Precisión léxica. TA

Comentado [MAP289]: Precisión léxica. TA

Comentado [MAP290]: Precisión léxica. TA

Comentado [MAP291]: Preposición. TA

Comentado [MAP292]: Puntuación. TA

Comentado [MAP293]: Precisión léxica. TA

Comentado [MAP294]: Estructura natural. TA

Comentado [MAP295]: Omisión. TA

Comentado [MAP296]: Puntuación. TA