

Psicología crítica y lingüística computacional: Acercamiento crítico a la traducción automática estadística

**Critical psychology and computational linguistics:
Critical approach to statistical machine translation**

Julio Cesar Corona Arias

Instituto Michoacano de Ciencias de la Educación (Morelia, México)

Resumen. El presente trabajo criticará el proceso estadístico y los conceptos psicológicos involucrados en la traducción automática estadística, basados en la suposición de la imposibilidad de traducción del lenguaje metafórico. Algunas preguntas iniciales son ¿cómo entiende la traducción automática estadística al lenguaje?, ¿cómo lidia con la metáfora y sus consecuencias?, y ¿qué puede decir la psicología crítica al respecto?

Palabras clave: traducción automática, lenguaje, metáfora, estadística

Abstract. This paper will criticize the statistical process and psychological concepts involved on Machine Translation and its critical implications, based on the assumption of the impossibility of metaphorical language translation. Some initial questions are ¿how does statistical machine translation understand language?, ¿how does it deal with the metaphor and its consequences? and ¿what does critical psychology can say about it?

Keywords: Machine translation, language, metaphor, statistics

Hay máquinas de rimar, pero no de poetizar

Octavio Paz

Introducción

Aproximación y justificación crítica

La primera interrogante al realizar este trabajo es acerca de la relación entre un traductor automático y la psicología. En el mundo actual hay gente siendo medicada, hospitalizada por diagnósticos psiquiátricos o neurológicos, compañías editoriales tratando de imponer prácticas u objetivaciones conceptuales y, en ese sentido, cabe preguntarse por la pertinencia de un análisis informático desde la psicología crítica, pudiendo haber cosas más cercanas al mundo ordinario por las cuales debería preguntarse.

La psicología cuenta con una gran habilidad para que sus discursos, prácticas y teorías psicológicas penetren en los más insospechados rincones estructurales; públicos o privados. La psicología crítica no sólo debe estar atenta a los departamentos de psicología, psiquiatría o neurología, es decir, aquéllos donde se supone se estudia al sujeto y lo que lo determina, sino con mayor razón debe estar atenta a los espacios donde se supone que no hay un sujeto o una disciplina psicológica; a las prácticas y procesos que tienen como objetivo eliminar y suplantar la presencia de los individuos, procesos donde domina la automatización y reproducción de las objetivaciones lingüísticas psicologizadas del sujeto.

La psicología conceptualiza parcializaciones del sujeto, lo divide en partes y formula cómo cree que funcionan; a las vivencias las agrupa bajo la noción de 'memoria' y a los acontecimientos orales bajo la idea de 'lenguaje'. Les asigna reglas y determinaciones para posteriormente ser tomados por otras disciplinas y explicarse desde esquemas ajenos a la psicología (De Vos, 2012, p. 1); si la psicología explica estos conceptos, entonces la lingüística computacional los predice y reproduce.

La lingüística computacional es uno de los resultados de aplicar la tecnología en programación, informática, matemática estadística o computación al estudio, predicción o procesamiento de textos y palabras en lenguaje natural (Bolshakov & Gelbukh, 2004, p. 25; Sidorov, 2009, p. 50). Lo anterior no quiere decir que en la traducción automática no haya una teoría psicológica formal y académica como las que ordinariamente ataca la psicología crítica, sino que a partir de estas objetivaciones de la lingüística computacional, podemos esbozar una idea de sujeto y encontrar una más de las "modalidades de vigilancia y regulación de la vida cotidiana" (Parker, 2009, p. 41) que se forman desde una "meta-perspectiva" (De Vos, 2012, p. 65) del individuo. Es necesaria una psicología crítica que haga una lectura

del lenguaje informático desde una posición metafórica y fundamentalmente significativa, esta psicología deberá "estar en oposición a cualquier reducción psicológica del sujeto", sin importar si el carácter es "computacional", "gramatical" o "predictivo" (Pavón-Cuéllar, 2009, p. 48) y develarnos "las estructuras de poder que mantienen a esas formas de lenguaje" (Parker, 2003, p. 203).

Acercamiento histórico

Históricamente, podemos ubicar el nacimiento de la traducción automática en la relación entre dos momentos particulares: El *Memorandum* de Weaver y el nombramiento de Yenoshua Bar-Hillel. En julio de 1949 el director de la fundación Rockefeller, Warren Weaver, emite un memorándum que comienza a llamar la atención sobre el problema de la ambigüedad en la TA. En 1951, a Yenoshua Bar-Hillel, un teórico pionero en la TA, se le reconoce como el primer investigador de tiempo completo en abordar los problemas en esa área.

Warren Weaver fue un académico estadounidense, director de la unidad de ciencias naturales de la fundación Rockefeller, con amplio conocimiento en criptografía, estadística y probabilidad. Fue impulsor en la creación de la corporación de investigación y desarrollo del ejército estadounidense (Hutchins, 2000, p. 18), conocida con las iniciales de RAND, 'Research and development'; distribuye de manera privada y personal un memorándum (Weaver, 1949), el cual se considera el "documento con mayor influencia en la historia temprana de la traducción automática" y el primero que propone el "uso de las computadoras para traducir" de un lenguaje a otro (Hutchins, 2000, p. 17); en ese trabajo inaugura la propuesta de una metodología y objetivos propios para el campo de la traducción automática a través del uso de herramientas y técnicas específicas.

En el *Memorandum*, Weaver propone 4 puntos básicos:

- La tarea de resolver "el problema de la ambigüedad" del lenguaje se debe hacer examinando el contexto inmediato de cada palabra dentro del texto, una palabra nos puede ayudar a determinar el significado de otra, por lo tanto la respuesta a tal problema radica en determinar esa contigüidad entre una palabra y otra; qué "valor mínimo" de contigüidad nos lleva a la "correcta elección de un significado" (p. 9).
- El aprovechamiento de sistemas de cómputo con capacidad lógica que aborden al lenguaje desde estas propiedades y puedan "asumir que el problema es formalmente solucionable" (*ibídem*)
- Uso de técnicas criptográficas para "enfaticar el estudio de la semántica estadística como un paso preliminar" (p. 12) en la interpretación del lenguaje; aceptaba la imposibilidad de perfección en la traducción, pero entendía los errores del lenguaje como ocurrencias estadísticas. La

criptografía se propone encontrar regularidades estadísticas en la “frecuencia de letras, combinaciones de letras, intervalos entre letras, patrones, etc.” (p. 3) y usarlos para ver generalidades y recurrencias en los diferentes idiomas.

- Consideraba la existencia de una base de “rasgos comunes” e “invariantes” entre todas las lenguas. No se aceptaba la posibilidad de traducir directamente de un idioma a otro, sino “descender desde cada lenguaje hasta la base de la comunicación humana” (p. 12), donde todos los lenguajes pueden entenderse desde un sólo lugar, un tipo de interlingua estadística.

En gran medida la propuesta de Weaver sigue aplicándose hoy en día: formalidad lógica, interpretaciones estadísticas, análisis de la contigüidad entre palabras. Weaver entendía la cuestión de la TA, como un problema de criptografía, un código que debía ser descifrado; tal código debía romperse a través de herramientas lógicas, estadísticas y un robusto equipo de cómputo.

Yenoshua Bar-Hillel fue un lingüista y lógico estadounidense de origen israelí. En 1951 comienza a trabajar en el departamento de investigación electrónica del Instituto Tecnológico de Massachusetts, su trabajo estaba enfocado en la aplicación de las computadoras a la investigación lingüística, especialmente en la traducción automática y la recuperación de información (Stranzy, 2005, p. 125). Se le considera el primer investigador en tener un puesto en el área de la traducción automática (Hutchins, 2000, p. 2328; Hutchins, 2010 p.1). Además de su trabajo en el Instituto Tecnológico de Massachusetts, colaboró con la oficina de investigación naval de los Estados Unidos (Hutchins, 2000, p. 305), RAND (*Idem*, p. 302) y la OTAN (p. 308). En el desarrollo de su trabajo se mostró crítico hacia algunos enfoques en la investigación en TA; para él, la posibilidad de una traducción totalmente automática de alta calidad no tenía posibilidades de realizarse (Bar-Hillel, 1952, p. 230) y se mostró crítico con los que perseguían ese objetivo. El problema principal consistía en que no existían métodos a través de los cuales fuera posible "eliminar las ambigüedades semánticas" (Bar-Hillel, 1951, p. 22), por lo menos no a través de la TA pura, o aquella que en su proceso no cuenta con intervención humana.

Modelado

Modelo de lenguaje

Cuando se hace mención a un traductor estadístico automático, se refiere a un sistema informático, un programa, una aplicación o extensión con diferentes componentes lógicos, computacionales y algorítmicos, donde

cada uno de ellos realiza una tarea específica y determinada con el propósito de traducir un texto.

La traducción estadística asigna a cada palabra o frase requerida un valor estadístico. La palabra o frase se elige dentro de otras, de acuerdo a su probabilidad de aparecer en un texto dado. Veamos cómo es el proceso que designa 'today' y no 'bird' o 'pizza' cuando traducimos la palabra 'hoy': lo anterior se realiza a través de un modelo de traducción, un modelo de lenguaje y un decodificador. A través de estos modelos el sistema elige basándose en la información que tiene en su corpus, esto es un conjunto de diferentes textos de donde se realizan estadísticas a muchos niveles; se hace una estadística de la frecuencia de ocurrencia de todas las palabras y si de toda esa información la probabilidad de que ocurra 'today' es 0.0075% y de que ocurra 'pizza' es 0.000.60%, entonces el resultado asignado es 'today'.

Para realizar una traducción, el modelo de lenguaje recibe una información de entrada (*input*), esto es, la palabra o frase que se quiere traducir; una palabra o frase en un idioma de origen. En este caso el *input* que usaremos de ejemplo es la palabra 'hoy'; el modelo de lenguaje asigna una probabilidad a esa palabra, basándose en la información del corpus: ¿qué probabilidades tiene la palabra 'hoy' de aparecer en el idioma español? Generalmente los corpus se construyen de textos oficiales: documentos parlamentarios, documentos de las páginas web oficiales de los gobiernos, manuales técnicos o empresariales, etc. Si una palabra se encuentra 200 ocasiones y otra está 100 ocasiones, entonces la primera tiene el doble de probabilidades que la segunda; así, la función del modelo de lenguaje es recibir el *input* en el idioma original y darle una asignación estadística correspondiente a la probabilidad de ocurrencia de la frase o palabra en ese idioma.

Markov y la falta de memoria

Ahora bien ¿qué sucede cuando no se requiere la probabilidad de ocurrencia de una palabra, sino de una frase o un grupo de palabras? En ese caso es útil recurrir a un modelo de gramas. Un grama es un conjunto determinado de palabras, en general un grupo de elementos cuya longitud y agrupación es variable y determinada; un bigrama es un conjunto de dos elementos, un trigrama un conjunto de tres; en un modelo bigramático de traducción una palabra depende de la contigua, en el trigramático cada palabra depende de la contigua y la que está junto a esa misma. Veamos la frase 'hoy lloverá mucho', en un modelo bigramático las probabilidades de ocurrencia de la palabra 'mucho' están determinadas por las probabilidades de ocurrencia de la palabra 'lloverá', el sistema buscará dentro de la información obtenida en el corpus y asignará más probabilidades a la palabra 'mucho'; es más probable que después de 'lloverá' encontremos la palabra 'mucho', en lugar de la palabra 'pizza'. Para que el sistema arroje

como resultado la palabra 'pizza', entonces en nuestro corpus tendría que aparecer la palabra 'pizza' junto a 'lloverá' más veces que la palabra 'mucho', el bigrama 'lloverá pizza' tendría que ser más probable que el bigrama 'lloverá mucho'. Si el modelo es trigramático, entonces la palabra 'mucho' no sólo depende de la anterior, o sea de 'lloverá', sino también depende a su vez de la palabra 'hoy'.

El proceso por el que una palabra es designada a su vez por otra, puede llevar a un problema de dispersión; si nuestra frase completa a traducir es 'Veo el cielo nublado, hoy lloverá mucho', la palabra 'mucho' se determina por la palabra 'lloverá', ésta por la palabra 'hoy', ésta a su vez por 'nublado'. Para evitar el problema de dispersión o *ad infinitum*, se usa el modelo de Markov, éste es un modelo que permite que una palabra determinada, sólo dependa de la anterior, es decir, delimita la causalidad de la probabilidad entre las palabras y así pretende evitar causalidades *ad infinitum*.

El modelo de Markov es un procedimiento estadístico con el cual se pueden modelar aquellos procesos que satisfacen la propiedad de Markov, esta característica o propiedad consiste en asumir que un estado futuro puede ser predecible a partir únicamente del estado presente, esto se conoce como un proceso de *falta de memoria*. En nuestro caso, cuando se habla de estados, se refiere a palabras; afirmar que se puede predecir un estado futuro, es afirmar que se puede *predecir* cuál es la siguiente palabra dentro de una oración o serie de palabras.

En los modelos de Markov, primeramente, se requiere saber qué probabilidades hay de empezar en el estado actual. Si nuestra serie de palabras es 'hoy lloverá mucho' y queremos saber cuál es la palabra que le sigue a 'hoy', primero tenemos que saber qué probabilidades iniciales hay de comenzar con 'hoy', a partir de allí cuál es la palabra que tiene más probabilidades de ocurrencia; si no tuviéramos este modelo estadístico únicamente sabríamos las probabilidades de ocurrencia de las palabras a partir del corpus en general y de esta forma podemos saber qué probabilidades hay, pero a partir de una palabra en particular.

Finalmente tenemos las probabilidades de transición, esto es el ritmo de ocurrencia de una palabra; puede ser que en un corpus dos palabras tengan casi la misma probabilidad, pero que una de ellas aparezca con mayor frecuencia al inicio de los textos, mientras que la otra aparezca de manera homogénea en cualquier parte del texto, esa estadística puede hacer la diferencia a la hora de arrojar un resultado.

Modelo de traducción

El modelo de traducción es bastante similar al modelo de lenguaje, al menos en su funcionamiento estadístico. Se necesita un corpus que sea paralelo o alineado, es decir que el mismo texto esté en por lo menos dos idio-

mas. Los textos paralelos o alineados datan de antes de Cristo, el primero que se conoce es la Piedra de Roseta, un bloque de piedra del año 196 AC, con el mismo mensaje escrito en 3 diferentes idiomas; egipcio, demótico y griego. Al igual que los corpus del modelo de lenguaje, los modelos de traducción automática buscan en internet textos alineados; se hace un conteo general de las palabras en el corpus, genera el promedio de ocurrencia de cada una y establece una relación entre la alineación de palabras, esto es, la relación entre cada palabra en ambos idiomas y ese es el resultado que arroja; el sistema va recorriendo las páginas y documentos en la web, generando estadísticas y relaciones de las palabras y frases.

Así como el modelo de lenguaje muestra las probabilidades de que una palabra determinada aparezca en un lenguaje origen, el modelo de traducción indica las probabilidades de ocurrencia en el idioma destino, allí radica la importancia de los textos alineados; en un corpus alineado la palabra 'hoy' debe tener una probabilidad de aparecer relacionada a la palabra 'today'. Finalmente, el sistema se encarga de hacer una valoración general y reordenamiento de las palabras, recibe como información de entrada la probabilidad de ocurrencia en el idioma original proveniente del modelo de lenguaje, del modelo de traducción recibe la probabilidad de alineación entre dos palabras y un decodificador finalmente reordena las palabras en caso de problemas sintácticos.

Edición, suavizado y entrenamiento

La pre y post edición así como la delimitación del dominio del texto son una práctica no sólo común, sino importante. La pre edición de textos permite que el sistema traduzca el texto de origen desde un sólo sentido; antes de introducir el texto de origen en el traductor, se le pueden hacer modificaciones sintácticas, gramáticas o morfológicas que eviten una traducción de baja calidad.

Los modelos que se usan en lenguajes cerrados, se restringen manipulando los datos que recaban a través de los corpus. Un lenguaje cerrado es aquél que está controlado y generalmente es muy técnico: corpus médicos, manuales de empresas o científicos. Si una palabra se traduce de 3 o 4 maneras diferentes, el modelo de traducción se programa para que ciertas palabras no aparezcan como resultado. Un ejemplo de lo anterior es la palabra 'light' que puede traducirse al español como el adjetivo 'ligero', pero también como el sustantivo 'luz'; si la palabra 'light' se usa dentro de un contexto determinado, entonces se programa el sistema para que se anule la palabra 'ligero' y sólo muestre como resultado la palabra 'luz' y así el resultado se da en un lenguaje controlado que permite su lectura desde un sólo sentido.

Si la probabilidad de una palabra, depende de su ocurrencia en el corpus ¿qué sucede con aquellas palabras cuya ocurrencia en el corpus es

nula? Si los corpus son creados con base en documentos oficiales o empresariales, ¿qué ocurre cuando el sistema se encuentra con una palabra hasta entonces desconocida? Para evitar ese problema, el sistema estadístico usa una técnica de suavizado que asigna de manera automática una probabilidad mínima a todas las palabras que no aparecen en el corpus, de esta manera todas las palabras tienen, aunque sea un mínimo de posibilidades de ocurrencia; si esto no se hace, las palabras desconocidas tendrían probabilidad cero y no se podría asignar un valor estadístico positivo a palabras cuya ocurrencia es cero.

Imposibilidades críticas

El significante como predicción, el lenguaje como mercancía

Los sistemas de traducción toman nuestras palabras, las convierten en *inputs* que traducen en *outputs* y los venden como si fueran la representación real de nuestra subjetividad en el mundo digital. Analizan las palabras que queremos traducir y su relación con el momento en que escribimos, su relación con nuestra ubicación geográfica, edad o los lugares que visitamos. Los sistemas informáticos automáticos se *entrenan* para analizar cuando entramos a nuestro correo, obtienen estadísticas de los movimientos con el puntero del 'mouse', si damos 'enter' o damos 'click' y cuánto duramos en hacerlo; las páginas a las que damos 'like', el tiempo que nos detenemos en una página antes de subir o bajar en la pantalla. Hay estadísticas para saber si las imágenes en que cada sujeto se detiene son de libros, computadoras o algún multivitamínico. Desde que nos conectamos a internet hasta que el aparato está totalmente apagado, todo lo que hacemos se adapta a un tipo de información estadística, rastreable, contable, reproducible y predecible.

Toda esa información acerca de lo que hacemos, leemos, escribimos o compramos, sirve para hacer una representación digital de cada uno de los sujetos; una digitalización de nuestra subjetividad. Nuestras individualidades son ignoradas y los sistemas informáticos nos reconocen como individuos sólo en la medida que podemos verificar nuestra identidad digitalizada; tenemos que comprobar que no somos robots para poder acceder a nuestra cuenta de correo o a alguna página con requerimientos de seguridad de alto nivel. A las empresas les interesa conocer nuestros hábitos de consumo y *quiénes somos*, al menos de manera digital. Toda esa información se recolecta y se vende para hacer llegar la publicidad a un comprador específico, tan específico como la segmentación del mercado basado en la cantidad de información que se posea de cada uno de los sujetos.

Al momento de contratar servicios de publicidad a través de grandes compañías como Facebook o Google, la plataforma donde se contrata el servicio pregunta hacia quién queremos dirigir nuestra publicidad o vender nuestro producto; horarios, edades, ciudades de residencia, nivel de estudios. Se intenta objetivar a los sujetos haciendo una relación entre la velocidad de conexión, modelo de celular, páginas que visita por la mañana y aplicaciones bajadas de Google play. En la medida en que esa información lo define de manera más exacta, mejor es la segmentación de mercado y mejor se ofertan los servicios publicitarios o de ventas en línea.

Una de las maneras más importantes para recopilar esta información es a través de los traductores automáticos, el más usado en el mundo es el de Google, el cual es de funcionamiento estadístico. Actualmente realiza más de 1 millón de millones de traducciones diarias (Johnson, 2017), su corpus no acepta textos que no puedan adaptarse a una reducción estadística; poesía, novela o literatura en general no son opciones aceptadas. Los corpus están totalmente hechos de textos oficiales; públicos o privados. A pesar de que Google cuenta con una de las bases de datos más grandes, Gramma Viewer, por "motivos legales" no puede usar esos libros para el corpus de su traductor (Ibídem) sí que su corpus está integrado exclusivamente por formas gramaticales, morfológicas, lógicas y sintácticas determinadas, lineales en tanto que oficiales, reduccionistas y falsamente neutrales; la importancia del lenguaje en la determinación del sujeto, está condicionada por esas formas oficiales.

Los traductores automáticos se validan como los mejores, los más precisos o verdaderos; para lograrlo se someten a otros programas evaluadores para que califiquen la calidad de las traducciones, entre más puntaje tenga un traductor, es mejor y por lo tanto, puede vender mejor nuestra información. Para evaluar la calidad de un traductor automático el método más usado es Bleu, un programa informático automático y estadístico: asigna un puntaje a la relación entre el orden y ocurrencia de las palabras en una traducción automática y la compara con la realizada por un "traductor humano profesional" (Papineni *et al*, 2002, p. 311). Para subir los puntajes de las evaluaciones, los sistemas de traducción automática *entrenan* sus corpus para asociarlos con otro tipo de información disponible y lograr una traducción de mejor calidad; la solución a la ambigüedad de una palabra, puede resolverse de manera diferente si el usuario está en España o en una Favela en Brasil.

Entre mejor entrenado está un traductor, más cerca está de llegar a una traducción perfecta, pero como el Aquiles de Zenón, jamás llegan a la meta. Cada vez más, usamos algún TAE y el resultado es mejor, nos satisface más; es más fluido, adecuado, correcto, exacto; sin embargo, entre más satisfechos, más notamos la imperfección, la falta de algo. No se entiende, algo sobra, no encaja o no se adecua en la traducción. Nuestra pequeña frase recién traducida la intentamos introducir como una pieza de

rompecabezas en un texto más amplio y algo pasa: no embona, no cabe, le falta, no sirve o caduca.

¿Qué falta y cómo se sustituye (Pavón-Cuéllar, 2009, p. 34) cuando se realiza una traducción automática estadística?, ¿qué es aquello que se quedó fuera en el modelo de Markov?, ¿qué se quedó justamente en esa palabra a la que el proceso no llegó?, ¿qué anomalías, perversiones discursivas, bordes e imposibilidades subjetivas se perdieron en el proceso de suavizado y entrenamiento?

Metáfora y psicologización

La psicología crítica debe analizar la manera en que la TAE asume que el lenguaje funciona y los términos de funcionamiento en que explica al lenguaje y al sujeto que lo articula. La TAE entiende la ambigüedad como un problema, como un 'algo' que le sobra al lenguaje y que se debe corregir para poderlo entender. Sin embargo, la ambigüedad no sólo no es un problema, es la parte fundamental del lenguaje; es una solución ante el problema de la formalización del lenguaje en tanto que impide la reducción estadística del sujeto determinado por este lenguaje.

En la lingüística computacional necesitamos ser específicos acerca del lenguaje al que nos referimos; PROLOG, JAVA, C, C++. Tenemos que decir algo más que sólo 'lenguaje', para saber de qué lenguaje hablamos. Cuando simplemente hablamos de 'el lenguaje', hacemos referencia al lenguaje ambiguo, al que determina al sujeto, al que es incontrolable, impredecible, al que se estructura como el inconsciente, al que metaforiza; lenguaje que no alcanza y que se equivoca, aquél que nos hace cuestionarnos acerca de la posibilidad de comunicarnos ante ese error sistemático (Arrivé, 1994, p. 107).

La metáfora no se puede entender y por lo tanto explicar o predecir, no se articula de la misma manera causal que en la lógica tradicional: de A no se sigue B, jamás sabemos si P, entonces Q. Sólo asumimos que las palabras dicen algo más o menos en nuestros mismos términos y nos aventuramos en la ilusión de creer que nos entendemos. En el lenguaje metafórico no hay un referente para lo que decimos, no hay un concepto o un significado ni algo en el mundo con lo que se pueda relacionar de manera necesaria; no hay objetos que representen conceptos, tampoco hay formas lógicas o lógica alguna que lo sustente. Hay significantes y efectivamente, significan (Pavón-Cuéllar, 2009, p. 37), pero jamás sabemos qué es eso que significan.

El lenguaje no formalizado aún tiene metáforas que le aparecen como tumores impredecibles e incontrolables, esas metáforas logran fundir en "unidad convincente imágenes que en la experiencia pueden estar separadas e incompatibles" (Pfeiffer, 1936, p. 39) y que ningún lenguaje formal puede. La unidad no se compone de una suma de elementos o adecuación

de formas pues no hay asociación certera de palabras ni "forma verbal" alguna para "explicar al mundo" (Olvera, 2016, p. 64), las palabras se significan unas a otras; una palabra nos va llevando irremediablemente a otra y otra más.

La TAE no puede traducir la metáfora, la reduce. Desde Warren Weaver ya podíamos encontrar una versión primera acerca de las implicaciones críticas en la traducción automática; se pensaba en la criptografía como la solución a los problemas en la traducción. Como si hubiera otro lenguaje que develara el código de una lengua común a todos y cuya forma se basara en la estadística de los errores, podemos ver que 65 años después, la semilla del *Memorandum* maduró sus frutos. El planteamiento del análisis sigue siendo la búsqueda de una lengua única, formalizable a través de la reducción estadística de las generalidades azarosas de los sujetos y reducirlas a afirmaciones particulares; una especie de interlingua estadística predecible.

Las diferentes técnicas de suavizado que se usan para asignar probabilidad mínima a aquellas palabras que no tienen ocurrencia en el corpus y que se traducen por primera vez, son sólo adaptaciones estadísticas determinadas de antemano. Esas palabras no aparecen en los corpus pues no comparten el mismo contexto u objetivo oficial; no son usadas por la mayoría, no se adaptan a situaciones estandarizadas o comunes, suelen estar asociadas a otras palabras o a dimensiones obscenas, subversivas, ofensivas o informales; no se adaptan a ninguno de los estándares oficiales. Si estas palabras no aparecen en los corpus oficiales, debe ser porque no promueven el bien, la democracia, el voto libre, la productividad y la buena voluntad que invade los discursos institucionales; públicos o privados.

Estas palabras son articuladas en contextos y por subjetividades periféricas; los discursos no oficiales provienen de aquéllos que no han sido adaptados a alguna de las estructuras institucionales y sociales, sujetos generalmente patologizados, vulnerables, impredecibles y ellos son justamente los que son peligrosos por su capacidad de "perturbar el buen funcionamiento de la economía" (Parker, 2007, p. 41), se suaviza su discurso para que sirva de "soporte de estructuras fonológicas, semánticas y sintácticas" (Braunstein, 1982, p. 212) que no sólo son coherentes entre ellas, sino además con el sistema que las contiene.

La TAE no puede traducir la metáfora, la olvida. Weaver planteaba la búsqueda de un valor mínimo de contigüidad, sabía que una palabra significa a otra interminablemente, pero tenía la esperanza de que la significación entre contigüidades pudiera delimitarse a unas cuantas palabras. El modelo estadístico de Markov representa la realización de esa delimitación, la posibilidad mínima de esa contigüidad; asegura la posibilidad de significación desde la palabra anterior inmediata y se olvida de las demás y de lo que éstas podrían agregar al riesgoso acto de significar. Los modelos

de Markov *asumen* que la significación entre palabras puede delimitarse y promueven la tendencia tecnológica a la inmediatez, a lo nuevo, a la constante actualización, promueven el inmenso "poder psicológico de hacernos olvidar" (Pavón-Cuéllar, 2015, p. 564) lo que podría ser necesario recordar; cortan de tajo casi cualquier posibilidad de significación subversiva de la memoria y el lenguaje.

Las palabras no sólo son enunciadas por un sujeto, son enunciadas por un sujeto en un tiempo y en un lugar determinado y a su vez, determinante. Los procesos de significación no tienen la propiedad *de falta de memoria* o un valor mínimo, esa memoria no tiene mínimos o máximos, la metáfora no tiene "vocabulario definidor" (Gebulkh & Sidorov, 2006, p. 67), su sentido puede llegar hasta la última palabra (Braunstein, 1982, p. 187) o en lo que no se dijo; "la cadena nunca se detiene y todo lo que los une también son significantes" (Pavón-Cuéllar, 2009, p. 37).

La TAE no puede traducir la metáfora, la ignora. Al nutrir su corpus sólo desde textos oficiales, los traductores automáticos dejan fuera la posibilidad de acercamiento con los discursos no oficiales o no autorizados y asume una falsa parcialidad. La pretendida "higiene verbal" (Parker, 2007, p. 183) de seleccionar el contenido, deja fuera las formas lingüísticas de la literatura y los discursos no hegemónicos de todo aquel individuo o grupo que no tiene representación digital, ni la manera de que sus significantes sean digitalizados. No toma en cuenta lo no estandarizable, predecible o lo que no se adapte a la idea del "procesamiento" de lenguaje como un "apartado necesario para cualquier país que quiera ser competitivo (Bolshakov y Gelbukh, 2004, p. 31) de cara a las nuevas tecnologías.

Conclusión

En defensa de la lingüística computacional puede decirse que no le importa explicar el lenguaje y que sólo se enfoca en el área estadística, eso es criticable; que le corresponde ese método y objeto de estudio porque otras disciplinas se encargan de lo demás, también criticable; afirmar que aumenta y mejora las redes de comunicación o la información, incluso decir que la lingüística computacional ayuda a la democratización de la humanidad, afirmación altamente criticable; sostener que colabora en la mejora de las relaciones económicas y de comercio, criticable. Para efectos críticos, tampoco es muy seguro que a la lingüística computacional le vaya mejor si afirma que sí intenta explicar el lenguaje, que sí lo intenta predecir y controlar; criticable. ¿Qué tanto se puede criticar? La respuesta depende de la radicalidad del psicólogo que lo critique y la coherencia con la que entienda y practique la digitalización de su entorno; plataformas, sistemas operativos, y licencias bajo las cuales realice su trabajo crítico y cotidiano.

El psicólogo crítico debe promover, provocar y permitir la subversión de los corpus; violentar los algoritmos que se usan para la búsqueda de texto: usar medios digitales alternos, desde procesadores de texto hasta sistemas operativos y mantener una crítica hacia el tema: qué lenguajes de programación, con qué sistema estadístico y qué conceptos psicológicos se incluyen. Si se escribe, edita, publica, fomenta o distribuye algún contenido psicológico con fines de digitalización: ¿qué precauciones se toman?.

El problema no es que la lingüística computacional tenga un acercamiento a las preguntas por el lenguaje, sino que lo hace para convertirlo en un producto que se vende, tal producto lo exhibe de manera análoga al sujeto; dice 'lenguaje', 'memoria', 'inteligencia', 'aprendizaje' o 'mente' para explicar por qué el sujeto habla y por qué es sujeto. Para matizar la severidad de esos conceptos, usa otros como 'automático', 'artificial' o 'autónomo': 'lenguaje automático', 'inteligencia artificial', 'aprendizaje autónomo'; un concepto anula al otro.

La lingüística computacional colabora con la digitalización de la subjetividad y de la algoritmización de la vida; somos la relación entre la hora en que entramos a nuestro correo, el tiempo que duramos en dar 'clic' y los grupos de facebook a los que pertenecemos; todo lo anterior dividido entre la estadística que le corresponde en el visor de n-gramas de Google a la palabra 'mexicano', tal vez a eso hay que restarle la frecuencia de compra en amazon.com; en este giro algorítmico de la vida, los programadores serán los nuevos mineros y el código será su carbón.

Referencias

- Arrivé, M. (1994). *Lingüística e inconsciente*. Bs. As: S. XXI, 2004
- Bar-Hillel, Y. (1951). The present state of research on mechanical translation. *American Documentation*, vol. 2, pt. 4, pp. 229-237
- Bolshakov, I. & Gelbukh, A. (2004). *Computational linguistics: models, resources, applications*. Mexico: UNAM.
- Braunstein, N. (1982). *El lenguaje y el inconsciente freudiano*. México; S. XXI
- Callison, B., Osborne, M. & Kohen, P. (2006). Re-evaluating the Role of Bleu in Machine Translation Research. Consultado el 7 de diciembre 2017 en <http://homepages.inf.ed.ac.uk/miles/papers/eacl06.pdf>
- De Vos, J. (2012). *Psychologisation in times of globalization*. NY: Routledge
- Gelbukh A. & Sidorov, G. (2006). *Procesamiento automático de español con enfoque en recursos léxicos grandes*. México: IPN.
- Hutchins, J. (1994). *The encyclopedia of languages and linguistics*, En R.E. Asher. (Ed.) Oxford: Pergamon Press, vol. 5, pp. 2322-2332

- Hutchins, J. (Ed.) (2000) Yenoshua Bar-Hillel, a philosopher's contribution to machine translation. *Early years in machine translation*. Amsterdam: John Benjamins, pp. 299-312. Consultado el 17 de agosto, 2017, en <http://www.hutchinsweb.me.uk/Bar-Hillel-2000.pdf>
- Hutchins, J. (2010) Machine translation: a concise history. Consultado el 15 de septiembre, 2017, en <http://www.hutchinsweb.me.uk/CUHK-2006.pdf>
- Johnson, M. (2017b) Stanford Seminar: Google's Multilingual Neural Machine Translation System. Consultado el 22 de octubre, 2017, en <https://www.youtube.com/watch?v=nR74lBO5M3s>
- Lewis, W. (2014) Microsoft Translator Hub - Discussion of BLEU Score. Consultado el 22 de octubre, 2017, en <https://www.youtube.com/watch?v=-UqDljMymMg>
- Olvera, C. (2016). *Sobre la mente*. México: UAA
- Pfeiffer, J. (1936). *La poesía*. México: FCE, 1986
- Papineni, K. & Roukos, S. & Ward, T. & Zhu, W. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311-318. Consultado el 5 de diciembre, 2017, en <http://www.aclweb.org/anthology/P02-1040>
- Parker, I (2003) Psicología Crítica: Conexiones Críticas en D. Villuendas y A. J. Gordo López (comps.) *Relaciones de Género en Psicología y Educación*, pp. 199-221. Madrid: Consejería de Educación.
- Parker, I. (2007). *La psicología como ideología, contra la disciplina*. Londres: Pluto press.
- Parker, I (2009). Psicología crítica: ¿Qué es y qué no es?. *Revista Venezolana de Psicología Clínica Comunitaria*, 8, pp. 139-159.
- Pavón-Cuéllar, D. (2009). Untying Real, Imaginary and Symbolic: A Lacanian Criticism of Behavioural, Cognitive and Discursive Psychologies. *Annual Review of Critical Psychology*, 7, 33-51.
- Pavón-Cuéllar, D. (2012). El año de publicación y la mala memoria de la psicología en. *Teoría y crítica de la psicología* 2, 184-188.
- Pavón-Cuéllar, D. (2015). El Ejército Zapatista de Liberación Nacional y su crítica de las psicologías conformista, despótica y desmemoriada. *Estudios de Psicología*, 32 (2), 557-568.
- Paz, O. (1956). *El arco y la lira*. México: FCE, 1972.
- Stranzy, P. (Ed.) (2005) *Encyclopedia of Linguistics*. NY: Fitzroy Dearborn, vol.1, 124-126

- Villa, J. (2009). *El dominio de la lingüística más allá de las ciencias naturales y exactas*. México: UNAM.
- Weaver, W. (1949) *Translation*, Fundación Rockefeller. Consultado el 6 de diciembre, 2017, en www.mt-archive.info/Weaver-1949.pdf
- Zhan, Y., Vogel, S. & Waibel, A. (2017) Interpreting BLEU/NIST scores: How much improvement do we need to have a better system? Consultado el 25 de noviembre, 2017, en <http://www.lrec-conf.org/proceedings/lrec2004/pdf/755.pdf>
-

Fecha de recepción: 9 de octubre 2017

Fecha de aceptación: 12 de mayo 2018