

Entrevista con Mikel Forcada

Presidente de la European Association of Machine

Translation



Amparo Alcina



Mikel Forcada
Universitat d'Alacant
mlf@dlsi.ua.es
ORCID:
[0000-0003-0843-6442](https://orcid.org/0000-0003-0843-6442)



Amparo Alcina
Universitat Jaume I
alcina@uji.es
ORCID:
[0000-0002-4931-564X](https://orcid.org/0000-0002-4931-564X)

Presentación

Mikel Forcada es Catedrático de Universidad en el Departamento de Lenguajes y sistemas informáticos de la Universidad de Alicante, donde desarrolla su investigación en el marco del grupo de investigación Transducens. Su investigación está marcada por su interés en relación con la traducción automática, la traducción asistida y el procesamiento del lenguaje natural. De todos conocido es el sistema de traducción automática Apertium, pero también han desarrollado otras aplicaciones como Bitextor, TagAligner, DocTrans, por nombrar algunas, todas ellas de código abierto. Mikel Forcada es presidente de la Asociación Europea de Traducción Automática (European Association of Machine Translation, EAMT). Destacaremos además que Mikel ha estado siempre predispuesto a colaborar e intercambiar puntos de vista entre el mundo de la informática y el de los lingüistas y traductores, y esta entrevista es también fruto de esta colaboración. Le hemos preguntado a Mikel sobre el presente y futuro de la traducción automática y sus distintos desarrollos, y sobre el papel de traductores y lingüistas en esta evolución.

Palabras clave: traducción automática.

Presentació

Mikel Forcada és Catedràtic d'Universitat al Departament de llenguatges i sistemes informàtics de la Universitat d'Alacant, on desenvolupa la seva investigació en el marc del grup de recerca Tansducens. La seva recerca està marcada pel seu interès en relació amb la traducció automàtica, la traducció assistida i el processament del llenguatge natural. Tothom coneix el sistema de traducció automàtica Apertium, però també han desenvolupat altres aplicacions com Bitextor, TagAlaginer, DocTrans, per citar-ne algunes, totes elles de codi obert. Mikel Forcada és president de l'Associació Europea de Traducció Automàtica (European Association of Machine Translation, EAMT). Destacarem a més que Mikel ha estat sempre predisposat a col·laborar i intercanviar punts de vista entre el món de la informàtica i el dels lingüistes i traductors, i aquesta

entrevista és també fruit d'aquesta col·laboració. Hem preguntat a Mikel sobre el present i el futur de la traducció automàtica i els seus diferents desenvolupaments, y sobre el paper de traductors i lingüistes en aquesta evolució.

Paraules clau: traducció automàtica.

Presentation

Mikel Forcada is a full professor in the Department of Languages and Computing Systems at the University of Alicante. His work is heavily influenced by his interest in machine translation, computer assisted translation and processing natural language and carried out within the context of the research group Transducens. Their most well-known machine translation application is Apertium, but they have developed others such as Bitextor, TagAligner, DocTrans, all of which are written in open code.

Forcada is the President of the European Association of Machine Translation, EAMT and has always been open to collaborating and exchanging points of view with those in the fields of computing, linguistics and translation. This interview is likewise the fruit of such collaboration, in which we quiz him about present and future trends in machine translation, its stages of development and the role of translators and linguists in its evolution.

Keywords machine translation.

Sinergias entre las tecnologías y la traducción ¿hablamos el mismo lenguaje? ¿nos podemos entender? ¿Conceptos como la calidad del texto traducido se perciben igual desde la perspectiva del traductor y del desarrollador de sistemas de TA?

Aunque hay gente que está en los dos campos, sí que es verdad que se hablan lenguajes distintos. Creo que desde el campo de la tecnología hemos tendido mucho a impulsar soluciones sin conocer bien los problemas, a tratar a la comunidad de traductores como si fuera despotismo ilustrado: "todo para el pueblo, pero sin el pueblo". Por otro lado, en la comunidad de la traducción se ve una gran desconfianza, probablemente por lo que acabo de decir, y también porque las tecnologías no han estado a la altura de lo que prometían. El trabajo de traducción es muy complejo y a veces se ve de manera excesivamente simplificada desde las tecnologías.

Pero nos podemos entender. Creo que es importante que en este campo, como en muchos otros, se aplique la directriz europea llamada "innovación e investigación responsable" (RRI: "responsible research and innovation", que justamente va en contra de ese despotismo ilustrado y recomienda que en la cadena de investigación e innovación deben estar implicados todos los agentes, incluidas las personas usuarias [1].

El concepto de calidad que se usa normalmente en el campo de la traducción automática se corresponde con el de la similitud entre la salida del sistema y una, tal vez un par, de traducciones de referencia, idealmente profesionales. Este concepto choca con el concepto de calidad como utilidad, que creo que sería mucho más importante para las personas que deben usar la traducción automática, por ejemplo

para posteditarla. La utilidad se podría medir en este último caso como ahorro de tiempo o de esfuerzo, y esto puede ir acompañado o no con la similitud entre la salida y alguna referencia. Debemos acercar las medidas de utilidad que se usan para entrenar los sistemas de traducción automática al resultado de mediciones objetivas de la utilidad de la salida en aplicaciones profesionales si queremos que la traducción automática sea verdaderamente útil.

Primero basada en reglas, luego estadística, ahora neuronal... ¿Qué recorrido le queda a la traducción basada en reglas y a la estadística?

La traducción basada en reglas sigue siendo una opción ideal cuando no hay suficiente texto paralelo para entrenar un sistema estadístico o uno neural, o cuando las lenguas son tan parecidas que los resultados son excelentes. A la estadística parece que le queda cuerda para rato y las comparaciones realizadas, por ejemplo, midiendo el esfuerzo de potedición, con la neural, no son aún totalmente concluyentes.

¿Qué recorrido le queda a la traducción automática neuronal?

Bastante. Aunque nació en el País Valenciano y en 1997 (de mano de Asun Castaño en Castellón o de Ramón Ñeco y mía en Alicante), durmió hasta que explotó en 2013. Acaba de empezar y en estos cuatro años ya han aparecido varias arquitecturas diferentes: codificador–descodificador, codificador–descodificador con atención, convolucional con atención, y sólo usando atención, así que seguro que nos esperan grandes cambios. Una de las virtudes de la traducción automática neural es su capacidad de integrarse sin solución de continuidad con otros sistemas como los de procesamiento de la imagen o del habla que usan técnicas similares. Creo que aún veremos muchas cosas interesantes.

¿Se intuye la aparición de sistemas de traducción automática de naturaleza diferente?

Ahora mismo no se me ocurre nada, pero no me hagan mucho caso: después de veinte años haciendo traducción automática y habiendo prácticamente inventado la traducción automática neural, no me imaginaba nunca que vendría esta eclosión neural, y eso que estuve hasta en el tribunal de una tesis que ya apuntaba en esa dirección.

¿Cuál es el techo de la traducción automática?

No le veo techo a la tecnología en sí misma, pero creo que la evaluación está aún en mantillas y podría ser que no estuviéramos sacando todo el partido de los sistemas al no saber cómo mejorarlos.

En sistemas que no se basan en reglas, ¿qué lugar queda para la lingüística? ¿Y para la terminología?

Los sistemas neurales pueden integrar información lingüística y terminológica siempre que la podamos representar en los términos en los que se representan las tareas de traducción en los sistemas de TA neural. Mucho camino que recorrer, aunque es verdad que los sistemas estadísticos y neurales tratan más bien con la información que se encuentra implícita en traducciones existentes en lugar de explícitamente en el conocimiento codificado por las personas expertas que las han hecho.

Cuáles son las prioridades de la TA ¿dónde encajan los traductores? ¿Seguirá habiendo lugar para la traducción asistida por ordenador?

La traducción asistida por ordenador será aún el modo básico de funcionamiento de la profesión en ámbitos en los que la precisión de la traducción es especialmente crucial: siempre hará falta que una persona profesional que conozca en profundidad la situación comunicativa y el propósito del texto para que cumpla la misión prevista. La profesión de traducción deberá adaptarse ese nicho de trabajo, porque en otros, la traducción automática será suficiente para el propósito previsto.

¿Qué aportación a la traducción automática se puede hacer desde la investigación sobre traducción e interpretación?

La traducción automática aún ve los textos como sacos de oraciones sueltas. Hay muchos aspectos (el discurso, la intención, la coherencia o la diversidad terminológica, etc) que aún están poco explorados.

¿Y desde la lingüística computacional?

La lingüística computacional nos hace falta, porque hay información y conocimiento lingüísticos excelentes sobre la tarea de traducción que debemos poder incluir en los sistemas que se vayan creando, particularmente en los neurales, aunque no debemos olvidar esos pares de lenguas que nunca podrán gozar de corpus de millones de oraciones traducidas y que seguirán necesitando sistemas basadas en reglas de naturaleza lingüística.