De los sistemas CONVERSACIONALES a los robots parlantes

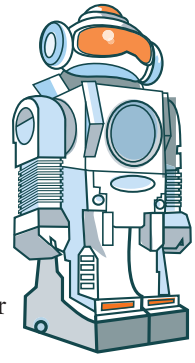
Luis Alberto Pineda Cortés

El desarrollo de sistemas multimodales es una empresa que se ha abordado desde el origen de la computación y que continúa hasta nuestros días. Aunque los sistemas capaces de sostener diálogos de manera autónoma son todavía limitados y dependen del avance de diversas tecnologías, su desarrollo ha sido constante y sistemático.

Introducción

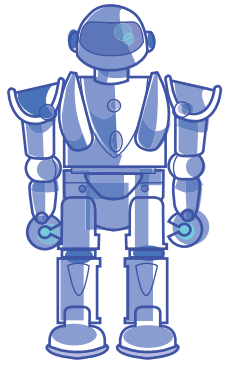
El desarrollo de máquinas con la capacidad de interactuar con los seres humanos a través del lenguaje hablado y la visión computacional es un tema que ha estado presente desde que aparecieron las computadoras.

La ciencia ficción nos ha brindado numerosas imágenes de estas máquinas, y los robots están ya en la conciencia pública. Pero la realidad científica y tecnológica es muy diferente de la que nos muestra el cine, y a pesar de los avances logrados desde los inicios de la computación, las máquinas con la capacidad de hablar, ver y moverse están aún en su infancia. Sin embargo, dado su amplio interés e impacto potencial, las diversas tecnologías involucradas, así como su integración, serán temas de investigación constante a lo largo del presente siglo.



Procesamiento computacional del lenguaje

La posibilidad de construir computadoras con la capacidad de pensar y hablar se planteó originalmente en el mundo académico y científico por Alan Turing con la publicación de su artículo "Computational machinery and intelligence" (Maquinaria computacional e inteligencia; Turing, 1950). En este artículo fundamental, Turing esbozó el programa de la inteligencia artificial (IA) y propuso dos tareas para esta disciplina: construir



máquinas capaces de jugar ajedrez, es decir, máquinas pensantes, y construir máquinas con la capacidad de "entender" el lenguaje natural, es decir, parlantes. La historia del ajedrez computacional es ampliamente conocida, y hoy contamos con máquinas prácticamente invencibles. La historia del procesamiento computacional del lenguaje es mucho menos sabida, pero es también muy rica, como trataremos de ilustrar a continuación.

El procesamiento del lenguaje humano se ha abordado desde varios frentes, y desde los diversos niveles de representación lingüística, como el fonético y fonológico, el prosódico y entonativo, el léxico y morfológico, el sintáctico y semántico, y también el nivel pragmático o contextual (Jurafsky y Martin, 2000). Este esfuerzo se inició con la formalización de la sintaxis y la teoría de la gramática transformacional introducida por Chomsky (1957), y se ha mantenido de manera continua hasta nuestros días, aunque la forma de ver a la estructura sintáctica ha evolucionado significativamente, y la visión actual es muy diferente al planeamiento original de Chomsky.

En esta tradición, la sintaxis se asume en gran medida como el nivel privilegiado de representación lingüística, ya que la estructura sintáctica colecta la información fonética y léxica y es, a su vez, la portadora del significado. Estas ideas dieron lugar a una inmensa actividad de investigación y se conformaron en un paradigma clásico del procesamiento del lenguaje centrado en la idea de analizar la oración para obtener estructura sintáctica, y de manera paralela producir una representación de su significado, ya que mediante esta última sería posible crear o actualizar representaciones en la memoria de la máquina, así como producir las conductas motoras consecuentes (hablar, moverse). Esto permitiría el enganche lingüístico de la máquina con su interlocutor humano.

Sin embargo, en este enfoque el significado de las palabras y las relaciones léxicas se tomaba en cuenta sólo de manera muy limitada, y las estructuras sintácticas resultaban sumamente ambiguas, ya que cada oración podía tener muchas interpretaciones. Como respuesta a esta problemática, el énfasis pasó al estudio de la estructura del lexicón o diccionario mental, donde el procesamiento estadístico del lenguaje ha resultado muy productivo (Manning y Shütze, 1999).

Por otra parte, comprender el significado de la oración involucra procesos adicionales a los léxicos y sintácticos, como la determinación de las referencias de los pronombres. Si el interlocutor A dice: "La ventana está abierta; ¿la podrías cerrar por favor?", la referencia del pronombre proclítico "la", es decir la ventana, se puede determinar o resolver en términos del contexto creado por la primera oración. Sin embargo, si A dijera la segunda oración señalando a la puerta del cuarto, también abierta, el contexto de interpretación sería el espacio local (el cuarto donde ocurren los hechos) y la referencia del pronombre sería la puerta y no la ventana. El ejemplo ilustra que estos dos tipos de inferencias, que aquí llamamos *anafóricas* e *indexicales*, respectivamente, ya no dependen solamente del significado de las palabras ni de la estructura sintáctica, que es local a la oración. Estas inferencias han sido objeto de un esfuerzo de investigación muy considerable, ya que son indispensables en cualquier programa de cómputo con la capacidad de sostener una conversación, y aunque a primera vista su modelación pudiera parecer sencilla, el problema es realmente muy complejo.

Sistemas conversacionales

A pesar de toda esta complejidad, a finales de los años sesenta del siglo pasado, Terry Winograd presentó SHRDLU (Winograd, 1972), el primer programa computacional capaz de llevar a cabo una conversación en inglés con un ser humano. La conversación era acerca del llamado "mundo de los bloques", consistente en una mesa sobre la cual había varios bloques geométricos, como cubos y prismas de diversos colores y tamaños, y el usuario humano podía hacer preguntas al sistema acerca de sus propiedades y relaciones, y también ordenarle que los reacomodara. Un diálogo de ejemplo así como el escenario correspondiente se puede ver en http://hci.stanford.edu/~winograd/shrdlu/index.html. Sin embargo, a pesar de lo impresionante del sistema, su funcionamiento era muy frágil y no se podía escalar a dominios más complejos.

Independientemente de esto, SHRDLU mostró que las cosas no son todo o nada, y que es posible modelar computacionalmente la conversación. Pero también hizo evidente que el progreso en esta empresa requería entender mejor la naturaleza del lenguaje y la comunicación.

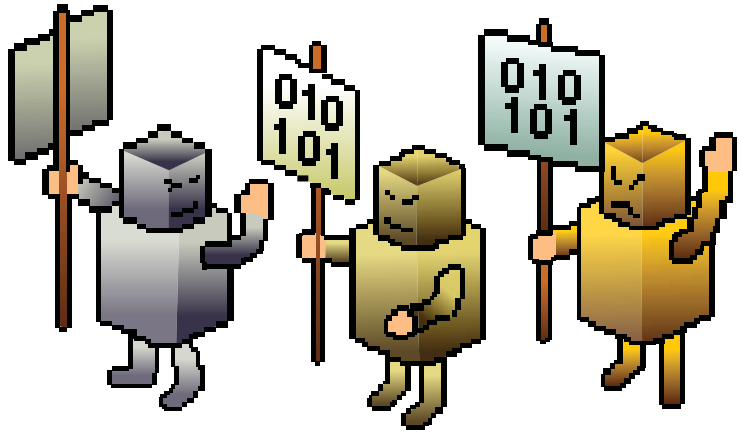
En un nivel más profundo, podemos decir que interpretar una oración consiste en comprender la intención que tiene el hablante cuando la expresa tomando en cuenta que la misma oración puede expresar diferentes intenciones en diferentes situaciones. Este enfoque corresponde con el nivel de representación "pragmático", y su estudio tiene sus antecedentes más directos en la filosofía del lenguaje, en los trabajos de Austin, Grice y Searle (Levinson, 1983), quienes introdujeron y desarrollaron el concepto de "acto del habla". De acuerdo con esta tradición, hablar es actuar, y las acciones lingüísticas son análogas a las acciones motoras, pero en vez de modificar el mundo físico, tienen sus efectos en el ámbito del conocimiento y las creencias de los interlocutores, y en el entorno social definido por las instituciones humanas.

Aquí hay que tomar también en cuenta que en el lenguaje cotidiano distinguimos la modalidad de las *elocuciones*, y hablamos de las *declarativas* para hacer afirmaciones, las *interrogativas* para preguntar, las *imperativas* para ordenar y las *exclamativas* para expresar emociones, y cada una de estas formas tiene su entonación característica. Sin embargo, frecuentemente cambiamos la modalidad para lograr diferentes efectos retóricos, como cuando se da una orden mediante una pregunta en vez del imperativo, que podría sonar descortés: si la ventana está abierta y le pido a María que la cierre diciendo "¿podrías cerrar la ventana, por favor?" no espero que me conteste que sí, que efectivamente tiene la capacidad de cerrarla, sino que espero que la cierre. Este tipo de actos, muy frecuentes en la conversación espontánea, se conocen como actos del habla indirectos, en oposición a los actos directos que usan la entonación natural, y su interpretación presenta la problemática adicional de que la intención expresada no corresponde con el significado literal. El problema se complica aún más si consideramos expresiones metafóricas como: "¿está rico tu veranito, mi amor?".

A finales de los años sesenta del siglo pasado, Terry Winograd presentó SHRDLU, el primer programa computacional capaz de llevar a cabo una conversación en inglés con un ser humano

Más aún, como lo importante son las intenciones, los actos del habla se pueden expresar con gestos o signos que se perciben mediante la vista, como el imperativo "detente" marcado con la mano extendida en forma vertical por un policía de tránsito, o por medio de una señal de tránsito octagonal roja con la palabra "alto". Desde esta perspectiva, podemos decir que la intención es independiente no sólo de la modalidad de la expresión, sino también de la modalidad de la percepción, y cualquier mensaje que exprese una intención comunicativa se puede considerar como un acto del habla.

Esta visión pragmática de la interpretación del lenguaje fue abordada directamente a finales de los años setenta del siglo pasado, y dio lugar a una generación de programas de cómputo capaces de interpretar actos



del habla indirectos en dominios acotados. Uno de éstos, construido por James Allen (Allen, 1995), era capaz de emular a un vendedor de boletos del tren (S), y sostener conversaciones (en inglés) con un usuario humano (U) como la siguiente: U: "Uno a Toronto"; S: "A las siete o a las ocho"; U: "A las siete por favor"; S: "son veinte dólares"; U: "Ok, gracias"; S: "hasta luego". Este programa formalizaba en buena medida las intuiciones pragmáticas descritas arriba, especialmente de acuerdo con el programa de Grice (Levinson, 1983), donde la intención se infería asumiendo que estas conversaciones son cooperativas, y que siguen las llamadas "máximas de Grice": ser informativo, ser relevante y usar las maneras y el orden apropiado para decir las cosas.

El estudio del procesamiento del lenguaje desde el punto de vista pragmático se siguió desarrollando de manera empírica durante los años noventa del siglo pasado, mediante el análisis de corpus conversacionales, es decir, de transcripciones de diálogos entre interlocutores humanos tomados tanto de conversaciones espontáneas como de experimentos controlados. Uno de los aspectos que se hicieron evidentes y que se tienen que enfrentar directamente es que no todos los actos del habla tienen el propósito de comunicar información o de transferir contenido *conceptual*; muchos actos del habla están más bien dirigidos a administrar y mantener a la conversación, es decir, a verificar que los interlocutores están hablando de lo mismo, y que las intenciones que uno expresa las entiende el otro de la manera apropiada. Por ejemplo, una solicitud de acción muy simple del interlocutor A a B puede desarrollarse como sigue: A: "cierra la ventana"; B: "¿perdón?"; A: "que si cierras la ventana"; B: "¡no te oí!"; A: "que si la cierras, por favor"; B: "hum, ¿tienes mucho frío?"; A: "¡ajá!"; B: "¡pues ciérrala tú, mi amor!". En este nivel el interés se centra no tanto en transmitir contenidos, sino más bien en establecer "el acuerdo" entre los interlocutores. No debería sorprendernos que transmitir nuestras intenciones, creencias y deseos no sea tan fácil, y cada vez que detectamos una falla hay que iniciar un diálogo de "reparación" con el propósito de reestablecer el "piso común" o *common ground* (Clark y Schaefer, 1989), y que una buena parte de la habilidad lingüística consista en darse cuenta cuándo se pierde el piso común y cómo reestablecerlo. Además, hay que tomar en cuenta el ruido en el ambiente, y que muchas veces hay errores tanto en la articulación como en la percepción del habla, y cuando esto ocurre es necesario reparar también.

Como se puede ver, los símbolos explícitos son, en última instancia, portadores de actos del habla, las unidades fundamentales de la comunicación. El problema se complica aún más si tomamos en cuenta que una elocución puede portar más de un acto del habla, y que éste puede tener la función tanto de comunicar contenido conceptual como de administrar y mantener la conversación. Como resultado de estas observaciones y con el fin de modelar estas conductas computacionalmente, fue necesario idear esquemas de transcripción para clasificar los actos del habla en los niveles del contenido, del acuerdo, y de la administración del canal de comunicación propiamente. Uno de estos esquemas fue el desarrollo, por el propio Allen, del esquema *Dialogue Act Marking Scheme Language* (DAMSL), y su extensión para diálogos prácticos multimodales en español, el esquema DIME-DAMSL, desarrollado en el Departamento de Ciencias de la Computación del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la Universidad Nacional Autónoma de México (UNAM) (Pineda y colaboradores, 2007).

El análisis de corpus de diálogos hablados resaltó nuevamente la complejidad del lenguaje humano, pero también mostró que los diálogos orientados a solucionar tareas simples de la vida cotidiana tienen una complejidad significativamente menor, y que su modelación computacional es posible en principio y además útil en las tareas en las que la asistencia computacional es deseable. Estas experiencias dieron pie a que se creara y fundamentara una nueva generación de sistemas conversacionales orientados a modelar este tipo de diálogos, a los que llamamos "diálogos prácticos". Sin embargo, nuevamente surgieron grandes dificultades, ya que una vez más sólo era posible construir sistemas de demostración muy difíciles de escalar a aplicaciones reales, que fueran lo suficientemente robustos para enfrentarse al gran público.

Hasta este punto se había asumido, en gran medida, que la interpretación de las elocuciones es un

proceso que se lleva a cabo independientemente del contexto. Sin embargo, los actos del habla no vienen por separado, sino que se dan en protocolos genéricos que se pueden utilizar en diferentes contextos, con diferente contenido conceptual, y que dan lugar a conductas esquemáticas que siguen los interlocutores de manera bastante fiel. Por ejemplo, un protocolo muy común para saludar es: A: "hola, ¿cómo estás?"; B: "muy bien, ¿y tú?"; A: "muy bien, gracias", y sólo cuando termina este intercambio pasamos a la siguiente fase de la conversación. Un protocolo de más alto nivel, definido en términos de un conjunto de "situaciones conversacionales", puede ser: saludar, presentar el problema, explorar las soluciones posibles, tomar una decisión, planear cómo se va a llevar a cabo, comunicar el plan y despedirse. Cada una de estas situaciones involucra un subdiálogo, y la conversación como un todo puede conceptualizarse en términos de un conjunto básico de esquemas que se pueden componer dinámicamente para formar diálogos más complejos.

La idea de representar el conocimiento a través de esquemas, que determinan en parte el contexto de cada acto de interpretación y acción, tiene también una larga historia en la inteligencia artificial, y a este capítulo pertenecen, por ejemplo, los marcos (*frames*) de Minsky (1985) y los esquemas de Shank (1975), desarrollados de manera muy fructífera desde los años sesenta hasta los ochenta del siglo pasado (aunque no era completamente claro si eran esquemas de lenguaje o esquemas de pensamiento, o ambos, lo que oscurecía un poco su estatus representacional). Por otra parte, en el campo de la lingüística, la investigación pragmática ha identificado también que el lenguaje humano usa hasta cierto punto protocolos conversacionales que se utilizan de manera recurrente (Levinson, 1983). Estas observaciones abren la posibilidad de definir y construir esquemas y protocolos conversacionales genéricos, formando un contexto de interacción *a priori*, que sitúan al programa conversacional y le permiten interpretar las elocuciones proferidas por el interlocutor humano en relación con dicho contexto y actuar de manera relevante. Aunque un modelo de este tipo es sin duda limitado para capturar la conversación humana sin restricciones, es probablemente adecuado para modelar diálogos prácticos de manera robusta, que es el tipo de aplicaciones al que podemos aspirar en el corto y mediano plazos mediante el uso de la tecnología computacional.

Una ventaja adicional de esta perspectiva es que los protocolos de interacción definidos en términos de los actos del habla permiten representar de manera integrada las intenciones expresadas tanto a través del lenguaje como de la visión, e incluso los eventos del mundo que podemos concebir como expectativas ante las cuales hay que actuar de manera intencional. Esta última observación abre la puerta para la creación de sistemas con lenguaje y visión, tanto para plataformas fijas como en robots móviles, centrados en la especificación e interpretación de protocolos intencionales, definidos en términos de los actos del habla.

Una consideración adicional es que los programas conversacionales como los que se han descrito, deben estar integrados o embebidos en un conjunto de programas y estructuras computacionales adicionales que modelen la percepción, el pensamiento y la acción, al cual nos referimos como "arquitectura cognitiva". En ésta, los conceptos particulares del dominio de aplicación (por ejemplo, venta de boletos del tren o avión, ordenar comida en un restaurante) pueden variar, así como el contenido específico de cada conversación, pero las estructuras y procesos computacionales son invariantes y representan un modelo de computación

genérico común a los sistemas conversacionales "de la misma especie".

Una arquitectura cognitiva requiere también contar con una memoria semántica para almacenar los conceptos generales y particulares utilizados en la conversación, así como una memoria de "hechos" donde se almacene la información factual con la que cuenta el programa conversacional. Esta distinción, ampliamente utilizada en la inteligencia artificial, es análoga, hasta cierto punto, a la distinción entre la *memoria semántica* y la *episódica* (Tulving, 1972), que ha sido sumamente productiva en psicología cognitiva y neuropsicología. Adicionalmente, es necesario contar con un "diccionario mental", es decir, con una memoria que relacione los conceptos léxicos (los significados de las palabras) con las imágenes acústicas, visuales y de otras modalidades asociadas a dichos conceptos. Lo importante es que tanto los eventos de interpretación como de generación del lenguaje requieren consultar y actualizar estas memorias, disponibles en la arquitectura cognitiva.

Figura 1. Robot Golem, primer robot mexicano capaz de sostener una conversación en español hablado.

El proyecto *Golem*

Las ideas esbozadas hasta este punto se han retomado y desarrollado en el proyecto Diálogos Inteligentes Multimodales en Español (DIME), en el Departamento de Ciencias de la Computación del IIMAS (DCC-IIMAS) de la UNAM, del cual surgieron el robot conversacional Golem (Figura 1), el sistema "Adivina la carta: Golem en *Universum*" (Figura 2), y más recientemente el robot Golem-II+ (Figuras 3 y 4). Estos sistemas son capaces de sostener una conversación en español hablado, apoyada con otras modalidades como la visión computacional, y el despliegue de textos, imágenes y videos. En particular, el robot Golem, ya retirado, era capaz de fungir como el guía de una sesión de carteles acerca de los proyectos de investigación del DCC-IIMAS; por su parte, el sistema "Adivina la carta: Golem en *Universum*", exhibido en un módulo permanente en dicho museo, cuenta con un conjunto de cartas con motivos astronómicos y "piensa" una, la cual debe ser descubierta por el usuario humano a través de una conversación en español hablado apoyada por otras modalidades. Por su parte,

Figura 2. Módulo "Adivina la carta: Golem en *Universum*", instalado en el Museo *Universum* de la UNAM.

Figura 3. Robot Golem-II+, capaz de guiar una sesión de carteles en español hablado e interpretar gestos (como señalar, por ejemplo) expresados por el interlocutor humano.

Figura 4. Robot Golem-II+.

el robot Golem-II+ es también capaz de fungir como guía de la sesión de carteles, pero adicionalmente puede interpretar gestos (como señalar, por ejemplo) expresados por el usuario humano, ilustrando la coordinación del habla, la visión y la conducta motora. Varios videos de estos sistemas están disponibles en la dirección web http://leibniz.iimas.unam.mx/~luis/.

El núcleo central de estos sistemas es una estructura representacional a la que llamamos "modelo de diálogo" (Pineda y colaboradores, 2010). Estos modelos representan protocolos conversacionales al nivel de los actos del habla que se pueden combinar dinámicamente, de forma que la definición de un conjunto básico de modelos permite articular diálogos prácticos orientados a la solución de tareas, incluyendo protocolos para reparar fallas conversacionales. Cada uno de estos protocolos se define en términos de un conjunto de situaciones conversacionales que se tienen que "visitar" para realizar una tarea, y cada situación se define de manera abstracta en términos de un conjunto de expectativas, un conjunto de acciones y un conjunto de "siguientes situaciones".

Las expectativas corresponden a las intenciones que pueden ser expresadas por el interlocutor humano o a eventos que ocurran en el mundo, pero que sean esperados por el sistema en la situación. Cada expectativa tiene asociada una acción que el sistema tiene que realizar si ésta se cumple, así como la situación a la que se llega si la expectativa se satisface y se realiza la acción correspondiente, pero tomando en cuenta que las expectativas, acciones y situaciones se determinan dinámicamente en relación con el contexto conversacional.

Los modelos de diálogos, así como su intérprete, son el centro de una arquitectura cognitiva que se ha desarrollado también en el contexto del proyecto Golem, la cual se muestra en la figura 5. Esta arquitectura está orientada a la interacción, y cada ciclo corresponde al reconocimiento e interpretación de una intención expresada por el usuario humano, la determinación del acto del habla apropiado para atender dicha intención, y la especificación y realización de dicho acto por parte del sistema. Como los actos del habla son en última instancia acciones, los actos del sistema se pueden realizar como conductas lingüísticas, como acciones motoras (por ejemplo, el movimiento

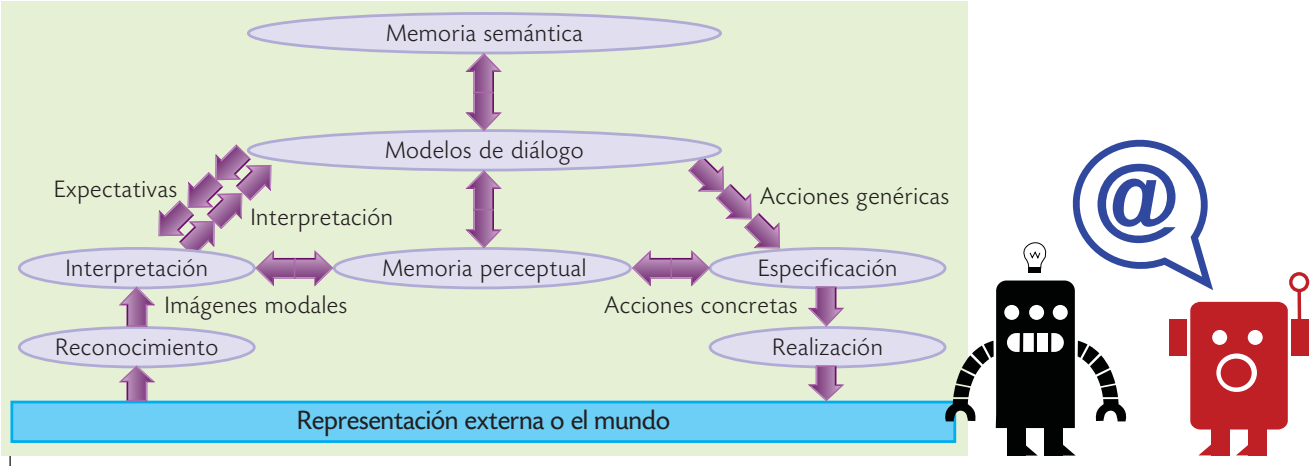


Figura 5. Arquitectura cognitiva desarrollada en el proyecto Golem.

de un robot móvil) y también como conductas internas, como razonar o planear, que sólo afectan a las estructuras representacionales del propio sistema.

El módulo "reconocimiento" contiene los sistemas de reconocimiento de voz y reconocimiento visual que traducen la señal hablada y las imágenes visuales a un código modal en el que se guardan las señales externas "en tanto imágenes", independientemente de su significado, por lo que las llamamos "imágenes internas". Por su parte, el módulo de interpretación asigna significados a dichas imágenes internas en términos de las expectativas del sistema en cada situación conversacional, y también en términos de la memoria perceptual en la que se guardan interpretaciones o significados asociados a imágenes codificadas de antemano, pero en los mismos códigos modales utilizados por los sistemas de reconocimiento de voz y de visión computacional. Por su parte, las acciones genéricas definidas en los modelos de diálogo se especifican dinámicamente de acuerdo con los contenidos de cada conversación, y el módulo de realización produce la conducta concreta del sistema, como sintetizar un mensaje de voz, desplegar un texto, una imagen o un video, o el movimiento del robot.

Finalmente, los procesos de interpretación y acción son relativos al contexto, el cual tiene dos componentes principales: 1) el *contexto a priori*, especificado en los modelos de diálogo, así como los contenidos de la memoria semántica y perceptual, que también se especifican de antemano, y 2) el *contexto dinámico* o historia de la interacción, que contiene las interpretaciones y acciones del sistema en cada conversación. En la arquitectura cognitiva, ambos contextos se toman en cuenta en cada nuevo acto de interpretación y acción.

Hacia el futuro

El desarrollo de sistemas de diálogo multimodales es una empresa de largo aliento, que se ha abordado desde los orígenes de la computación hasta nuestros días, y con toda probabilidad continuará su desarrollo en el mediano y largo plazos. Actualmente existe una actividad de investigación y desarrollo muy intensa en el entorno internacional, y se cuenta con una gran variedad de sistemas que operan en plataformas fijas y robots móviles. Esta empresa está también acotada por el avance de las tecnologías de reconocimiento de voz, de imágenes, así como de realización de conductas motoras autónomas, cada una con sus historias particulares, y depende también de los avances del *hardware*, como los procesadores y memorias de gran capacidad, tamaño reducido y bajo consumo de energía. Aunque el progreso ha sido lento, y los sistemas capaces de sostener diálogos de manera autónoma son todavía limitados, el desarrollo ha sido constante y sistemático, por lo que es probable que estos sistemas se integren poco a poco a la vida cotidiana.

El reto actual más importante es hacerlos robustos en ambientes de interacción reales. Por otra parte, la maduración de esta tecnología irá a la par de la reducción de costos y tamaños de los dispositivos computacionales. Estos sistemas, más que desplegar una

conducta inteligente de carácter general, como en las películas, muy probablemente vendrán integrados a artefactos y dispositivos de la vida cotidiana, como automóviles, teléfonos celulares, sistemas de audio y video o artefactos domésticos, para llevar a cabo diálogos prácticos útiles para los propósitos de cada dispositivo, haciendo que la tecnología sea "invisible". Más que un salto tecnológico, estos sistemas se irán haciendo presentes poco a poco, de manera muy discreta, y el día menos pensado, en un futuro tal vez todavía remoto, van a formar parte integral de nuestras vidas. Por lo mismo, su impacto económico y social será de grandes dimensiones.

El desarrollo de ese tipo de sistemas representa una oportunidad para la investigación y el desarrollo tecnológico. Desafortunadamente, los grupos centrados en estos temas en México se cuentan con los dedos de las manos, y estas tecnologías, como la investigación y desarrollo tecnológico en computación en México, son muy poco comprendidos no sólo por la sociedad en general, sino incluso por científicos de otras áreas que la ven con recelo y desconfianza.

La historia de la computación en México ha sido hasta nuestros días la historia de los equipos y sistemas que se han comprado, y cómo se han aplicado en los diversos sectores productivos o en la investigaciones científica, que no pueden prescindir de esta herramienta. Pero la ciencia y la tecnología computacionales desarrolladas en México están todavía por tomar su sitio.

Sin embargo, por su interés filosófico y científico, por el reto tecnológico que representan y por su impacto económico potencial, que realmente puede ser muy grande, el desarrollo de sistemas de diálogos prácticos multimodales representa una oportunidad, que estará abierta por varios años todavía, para subirnos al tren de la tecnología. Ojalá podamos aprovecharla.

Luis Alberto Pineda Cortés es ingeniero electrónico por la Universidad Anáhuac, maestro en ciencias computacionales por el Instituto Tecnológico de Estudios Superiores de Monterrey y doctor por el Centre for Cognitive Science de la Universidad de Edimburgo. Es investigador en el Instituto de Investigación en Matemáticas Aplicadas y en Sistemas (IIMAS) de la Universidad Nacional Autónoma de México (UNAM). Ha publicado extensamente en lingüística computacional e inteligencia artificial, y ha participado como ponente en numerosas ocasiones, tanto en México como en el extranjero. Es miembro del Sistema Nacional de Investigadores y miembro regular de la Academia Mexicana de Ciencias y de la Academia Mexicana de Informática (AMIAC). Actualmente es el coordinador de la Red Mexicana de Investigación y Desarrollo en Computación (Remidec).

lpineda@unam.mx

http://leibniz.iimas.unam.mx/~luis/

Bibliografía

Allen, J. F. (1994), *Natural language understanding*, Redwood City, California, Benjamin/Cummings.

Chomsky, N. (1957), *Syntactic structures*, La Haya, Mouton.

Clark, H. y E. F. Schaefer (1989), "Contributing to discourse", *Cognitive science*; 13: 259-294.

Jurafsky, D. y J. Martin (2000), *Speech and language processing*, New Jersey, Prentice-Hall.

Levinson, S. C. (1983), *Pragmatics*, Cambridge, Cambridge University Press.

Manning, C. D. y H. Schütze (1999), *Foundations of statistical natural language processing*, Cambridge, EUA, The MIT Press.

Minsky, M. (1985), "A framework for representing knowledge", en Ronald Brachman y Hector Levesque (eds.), *Readings in knowledge representation*, Los Altos, California, Morgan and Kaufmann, pp. 245-262.

Pineda, L. A., V. Estrada, S. Coria y J. Allen (2007), "The obligations and common ground structure of practical dialogues", *Revista Iberoamericana de Inteligencia Artificial*; 11(36): 9-17.

Pineda, L., I. Meza y L. Salinas (2010), "Dialogue Model Specification and Interpretation for Intelligent Multimodal HCI", en A Kuri-Morales y G. Simari (eds.), "IBERAMIA 2010", *Lectures notes in artificial intelligence*, Berlin, Springer-Verlag, 6433: 20-29.

Shank, Roger C. (1975), "The structure of episodes in memory", en George F. Luger (ed.), *Computation and intelligence: collected readings*, Menlo Park/Cambridge, MA: AAAI Press/The MIT Press, pp. 236-259.

Tulving, E. (1972), "Memory systems: episodic and semantic memory", en E. Tulving y W. Donaldson (eds.), *Organization of memory*, Nueva York, Academic Press, pp. 381-403.

Turing, Alan (1950), "Computing machinery and intelligence", *Mind*; 59: 433-460.

Winograd, T. (1972), *Understanding natural language*, Nueva York, Academic Press.