Modelo computacional del diálogo basado en reglas aplicado a un robot guía móvil

Grigori Sidorov, Irina Kobozeva, Anton Zimmerling, Liliana Chanona-Hernández, Olga Kolesnikova

***Resumen*—En este artículo se presenta la descripción formal detallada del módulo de control del diálogo para un robot móvil que funciona como guía (en un museo). El módulo incluye el modelo proposicional del diálogo, la especificación de los actos de habla y los bloques del habla, así como el inventario de los patrones de habla correspondientes a todos los actos de habla en el modelo. El modelo del diálogo se implementa como una red de estados y transiciones entre estados condicionados por reglas que estipulan los factores verbales y visuales. La arquitectura del módulo no depende de un idioma particular y puede ser adaptado a cualquier lenguaje natural.**

***Palabras clave*—Modelo de diálogo, reglas, robot guía móvil, actos de habla, bloques de habla.**

Computational Model of Dialog Based on Rules Applied to a Robotic Mobile Guide

***Abstract*—This paper presents a formal detailed description of the dialogue management module for a mobile robot functioning as a guide. The module includes a propositional dialogue model, specification of speech acts and speech blocks as well as the inventory of speech patterns corresponding to all speech acts of the model. The dialogue model is implemented as a network of states and transitions between states conditioned by rules, which include verbal and visual factors. The architecture of the module is language independent and can be adapted to any natural language.**

***Index Terms*—Model of dialog, rules, mobile robotic guide, speech acts, speech blocks**.

Manuscrito recibido 6 de julio de 2014, aceptado para su publicación 3 de noviembre de 2014, publicado 15 de noviembre 2014.

Grigori Sidorov (autor correspondiente) trabaja en el Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), México, DF, México (correo electrónico: sidorov@cic.ipn.mx, página web: www.cic.ipn.mx/~sidorov).

Irina Kobozeva es catedrática de la Universidad Estatal Lomonosov de Moscú, facultad de filología, Departamento de Lingüística Teórica y Aplicada, Moscú, Rusia.

Anton Zimmerling trabaja en la Universidad Estatal para Humanidades Sholokhov de Moscú, siendo jefe del Laboratorio de Lingüística General y Computacional, Moscú, Rusia.

Liliana Chanona-Hernández trabaja en el Instituto Politécnico Nacional (IPN), Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME), México D.F., México.

Olga Kolesnikova trabaja en el Instituto Politécnico Nacional (IPN), Escuela Superior de Cómputo (ESCOM), México D.F., México.

I. Introducción

La posibilidad de manejar un robot móvil dándole comandos directamente por voz humana es muy atractiva, ya que es mucho menos natural teclear los comandos. Aún mejor se percibe la posibilidad de dialogar con este dispositivo móvil, cuando el robot no solamente entiende un comando, sino también puede contestar dentro de su ámbito de competencia, porque tal actividad se asemeja mucho a la interacción con un ser humano. Sin embargo, el modelado de diálogo es un problema muy difícil y a pesar de ser investigado por más de cincuenta años sigue siendo un problema abierto.

La dificultad de manejo de diálogo se relaciona con la información multimodal que el robot tiene que "entender", tomando en cuenta las diferencias fundamentales entre los lenguajes de representación del espacio en la "mente" del robot y en la mente humana. También a eso se anuda el problema de comprensión del lenguaje natural a un nivel tan profundo que permita interpretar correctamente las intenciones del interlocutor [10, 11, 13].

Durante los últimos veinte años se han realizado muchos trabajos de modelado de la comunicación humana con robots móviles, pero por lo general es de carácter exploratorio. Normalmente se analiza en profundidad alguno de los aspectos particulares de este problema bastante complejo, tales como la capacidad de influir en el éxito de la comunicación con gestos, la importancia de la dirección de la mirada y el conocimiento de la situación para la interacción efectiva [6, 25]. Por otro lado, cabe mencionar que existen algunos desarrollos de robots móviles que ya pueden realizar ciertas funciones y entablar un diálogo con el usuario en un nivel aceptable de naturalidad en unas áreas muy específicas, por ejemplo, [1, 2, 7, 17, 19]. Sin embargo, no existe un modelo estándar para el modelado de diálogo e interpretación de los actos de habla.

II. Enfoque algorítmico vs. aprendizaje automático por computadora

En este artículo se presenta un enfoque algorítmico para el problema del modelado de diálogo en lenguaje natural tomando como ejemplo un diálogo guiado por un robot móvil. Cabe señalar que en la lingüística computacional moderna, cuando hablan del enfoque algorítmico en general, a menudo se utiliza la metodología basada en el aprendizaje automático por computadoras. Por ejemplo, en el caso del diálogo, un método basado en el aprendizaje automático fue aplicado al

problema central de su procesamiento: detección de los actos de habla, como se describe en [21]. Sin embargo, allá no se presentaron las evidencias convincentes de que este aprendizaje permite una buena detección de los actos de habla en un diálogo.

Cabe mencionar que los actos de habla constituyen claramente un buen material para el uso de este tipo de aprendizaje automático, pero el problema de utilizar las computadoras para esa tarea se dificulta dado las características de los actos de habla en un diálogo, que tienen la naturaleza semántica y pragmática, y en su etapa actual del desarrollo de la teoría de procesamiento de diálogo y de texto en general es muy difícil detectarlas confiablemente. Entonces, para el manejo computacional del diálogo normalmente se usan los enfoques y modelos algorítmicos más tradicionales, y no los métodos basados en el aprendizaje automático.

En este artículo nos basamos en un enfoque algorítmico para la descripción del módulo de manejo de diálogo que se usa para interactuar con un robot guía móvil siguiendo el enfoque general basado en reglas descrito en [15].

III. Módulos para implementación de un guía robótico autónomo

El esquema general de procesamiento de diálogo incluye dos módulos lingüísticos auxiliares: el módulo de reconocimiento de voz (en nuestro caso hemos usado las librerías de *Dragon Naturally Speaking*) y el módulo de análisis lingüístico (de niveles morfológico y sintáctico) utilizando el sistema Freeling [9]. En nuestro caso, el sistema que maneja el diálogo está siendo desarrollado para el idioma español, sin embargo, al cambiar los módulos lingüísticos de análisis y los módulos de procesamiento de voz, el sistema puede funcionar para cualquier idioma, es decir, el conjunto de actos de habla es universal: independiente de un idioma específico.

Ya que estamos hablando del diálogo con un robot, se necesita tener un robot físicamente. La "mente" del robot (donde se procesa toda la información) es nada más que una computadora estándar. En este sentido, todos los procedimientos consisten en el desarrollo y aplicación de programas de computadoras, siguiendo la metodología tradicional de desarrollo de software. Si usamos un robot móvil, es cómodo usar una computadora portátil (y no de escritorio), la cual puede ubicarse directamente sobre el robot.

En nuestro caso, el "cuerpo" del robot es un robot móvil Pioneer 3DX. El robot sabe navegar en un espacio conocido de un punto a otro evadiendo obstáculos que se encuentran en su camino. Es un robot autónomo, es decir, nadie lo está guiando paso por paso, sino el robot calcula su ruta por sí mismo. El robot está equipado con un sensor láser de gran precisión para medir distancias hasta posibles obstáculos en su camino: esta manera de conocer los posibles obstáculos sustituye la visión de los humanos. Digamos, de manera similar se mueven los murciélagos, usando principalmente su sonar biológico. Note que este tipo de robot es el estándar de facto para el desarrollo de aplicaciones de los robots móviles autónomos en el mundo real —sus dimensiones son comparables a los de un ser humano.

El robot se mueve utilizando un mapa de su entorno preparado con anterioridad. La preparación del mapa es un procedimiento muy sencillo y rápido, simplemente el usuario debe mover, en este caso manualmente, al robot en el espacio del mapa y después un programa recalcula el mapa correspondiente. En el mapa se puede marcar las áreas prohibidas, es decir, lugares donde el robot no debe pasar. También se puede marcar algunos puntos específicos como metas provisionales. Por ejemplo, en este artículo consideramos, dado que estaremos describiendo un posible escenario de uso del robot como guía de un museo, por decir algo, las obras maestras del museo de Louvre (Francia): la Mona Lisa (Mona_Lisa, M_L), la Venus de Milo (Venus_de_Milo, V_M), el Escriba Sentado (El_Escriba, E_S). Más adelante, estos nombres se utilizarán en nuestra descripción del modelo de diálogo, pero en cada aplicación particular de este modelo, deben ser sustituidos por otros objetos materiales que se encuentren en el entorno particular.

El robot se mueve con mucha precisión en el espacio representado en el mapa, dado que se conoce su propia posición en cada momento. Más aún, si aparece un obstáculo nuevo (ausente en el mapa), el robot puede entender esa situación y procesarla correctamente, es decir, evadir este obstáculo si se le obstaculiza su camino. Normalmente los obstáculos son los humanos que van y vienen, o los muebles que cambiaron de lugar.

En una situación muy particular, cuando los obstáculos (humanos), tapan todo el espacio disponible para el robot, el robot puede perderse. Cuando se despejara el ambiente, un módulo de localización tiene que volver a calcular su posición.

La idea de conocer su posición y la posición de otros objetos (posibles metas en el desplazamiento) en el mapa, permite al robot saber cuándo cada objeto aparece en su campo de "visión". Esto sucede si no hay ningún obstáculo entre el robot y el objeto (o la posición conocida del objeto) y la distancia entre ellos es la suficientemente pequeña. Proponemos considerar el umbral de 3 metros para considerar un objeto como "visible", es decir, dentro del campo de "visión" del robot. En nuestro modelo, cuando un objeto X se encuentra en el "campo de visión" (recordemos que no hay cámaras en el robot, sino se usa el sensor laser), eso se refleja como la proposición Visual_Act(X), donde X es la variable para un objeto.

Para el desarrollo de aplicaciones del robot Pioneer 3DX, se usa la arquitectura Cliente-Servidor. Es decir, existe un servidor que procesa e interpreta los comandos de un cliente y los transforma a los comandos de bajo nivel para el robot, y existe un programa cliente que manda al servidor los comandos de alto nivel (tipo "ejecuta el bloque 1" o "desplázate hacia Venus_de_Milo"). Note que el programa cliente puede correr en la misma computadora que el servidor

o en alguna otra computadora. En este caso el cliente se comunica con el servidor, por ejemplo, a través del protocolo TCP-IP.

IV. Escenarios de uso de un robot móvil

Entre los escenarios de uso de un robot móvil (es decir, las funciones que puede ejercer un robot móvil) de manera muy general se puede mencionar los siguientes:

1) La función de un guía (por ejemplo, un guía en un museo, empresa o exposición). Es el escenario que estamos considerando en este artículo.
2) La función de un mecanismo transportador, como, por ejemplo, un usuario en silla de ruedas que está controlada con comandos de voz.
3) La función de un mensajero, para pasar un objeto a cierto lugar de manera autónoma o siguiendo comandos enviados por el usuario. En este caso hay dos posibilidades:
   a) el robot "conoce" el camino;
   b) el robot no conoce el camino y para pasar debe comunicarse por el acceso remoto con el usuario [2].
4) La función de un reportero que informa de lo que percibe en cada momento.
5) La función de un discípulo que está capacitado para comprender su entorno (véase [4, 5, 20]).
6) La función de un camarero en un restaurante.
7) La función de un ayudante en selección de objetos, por ejemplo, ayudar al comprador en una tienda.
8) La función de un manipulador que ejecuta los comandos del usuario para mover los objetos (por ejemplo, el primer robot manipulador de T. Winograd, SHRDLU).

Por supuesto, el robot puede configurarse para ejecutar una combinación de esas funciones. A menudo se combina las funciones de un mensajero con la función de un manipulador: el robot debe moverse hacia el usuario en la posición adecuada y hacer algo con unos objetos especificados, por ejemplo, un mesero manipularía los platillos. Las funciones del estudiante y periodista pueden combinarse entre sí y con la función de un mensajero, como por ejemplo, en [18], donde el robot navega recibiendo los comandos del usuario e informa qué él percibe y además aprende los nombres de los objetos percibidos, preguntándolo al usuario.

Cada función impone ciertos requisitos para el diálogo y un repertorio de actos comunicativos (actos de habla: réplicas y/o reacciones no verbales). Dependiendo de las características del diálogo, el iniciador y el controlador de desarrollo de un diálogo puede ser tanto el usuario como el robot. Las funciones del robot a veces lo obligan ser el iniciador y el controlador que ejecuta todo el diálogo, pero en otros casos el robot solo inicia el diálogo y es el usuario quien lo controla posteriormente. También ambos comunicantes (el robot y el usuario) pueden asumir la función de iniciar el diálogo.

Otros aspectos del diálogo que están fuera del alcance de este artículo son: el repertorio más amplio de los actos comunicativos que el robot debe producir y entender, el módulo conceptual intencional (información enciclopédica y modelos de comportamiento racional). Es decir, en nuestro caso, el módulo de diálogo está desarrollado para el sistema que no tiene el módulo de "conocimiento amplio" y por lo tanto modela solamente una función relativamente simple del robot guía, cuando el robot asume el control en el transcurso de diálogo. En este caso, se reduce el repertorio de los actos comunicativos necesarios (*speech acts*, SA).

V. Particularidades de diálogo de un robot guía con visitantes

El término de robot guía en este caso, se refiere no sólo a una determinada ocupación (profesión), sino a cualquier agente que en determinadas situaciones da a conocer a los visitantes un lugar concreto y realiza una visita guiada en el lugar y proporciona la información sobre aquellos objetos que se encuentran en este lugar. Ese lugar puede ser alguna localidad (por ejemplo, de la ciudad), la propiedad de una persona (por ejemplo, una casa) o el interior de un edificio o parte del mismo (por ejemplo, un museo, un piso de alguna escuela, o departamentos de las instituciones). En su caso, el guía puede ser el dueño de la casa, un agente inmobiliario, un empleado de una institución, un residente local, y en general cualquier persona que se compromete a ejecutar la función social indicada: dar a conocer.

En todo caso, la ejecución de funciones un guía comienza cuando el agente acepta este papel, que a su vez implica que la contraparte acepta la función de un "visitante de lugares desconocidos" y tiene la intención correspondiente. En el caso de un robot, es el ser humano el responsable de la conveniencia de su aceptación en el papel del guía, dado que es la persona que inicia el proceso en el momento adecuado, es decir, una persona puede, por ejemplo, llamar al programa apropiado y el robot comienza a ejecutar el diálogo.

El objetivo general de un guía —dar a conocer el lugar a los visitantes— se divide en una serie de subtareas, cada una de las cuales es traerlos al objeto específico (en el sentido más amplio) de interés para ellos y, posiblemente, contar una determinada cantidad de información sobre este objeto. Durante la ejecución de cada una de las subtareas, el guía puede, además, llamar atención a lo largo de la ruta de acceso al siguiente objetivo a los objetos "secundarios", por ejemplo, en un museo durante el camino hacia un obra de arte, mencionar donde está la cafetería.

A su vez, la ejecución de cada tarea implica la ejecución de actos de comunicación (verbales y/o no verbales) y el desplazamiento al objeto de interés preestablecido por el camino conocido de antemano para el guía.

VI. Los actos de habla y los bloques del habla

En el caso del diálogo del robot guía con el visitante vamos a distinguir los actos de habla y los bloques del habla. En

nuestro caso, el modelo se desarrolla para el diálogo de un robot guía en un museo y el repertorio de las acciones del robot incluye tres tipos de actos: los actos de habla, los bloques del habla y los bloques multimodales.

A. *Actos de habla*

Los actos de habla entendemos de la manera estándar como en la teoría de los actos de habla: como actos de pronunciar una expresión lingüística, generalmente simple, que tienen una cierta fuerza ilocutiva, es decir, expresa (directa o indirectamente) la intensión del hablante, junto con varios otros componentes del significado pragmático [13].

El conjunto de actos de habla para el robot en esta situación es:

- Pregunta sobre la aceptación de la sugerencias de visita (Tour-Question),
- Sugerencia para seleccionar el primer objeto de la visita de un conjunto de alternativas, en este caso, Venus_de_Milos, la_Mona_Lisa, el_Escriba_Senta-do, etc. (Question_Altern(V_M, E_S, M_L)),
- Finalización del diálogo en caso del rechazo de la propuesta (Closure) o cuando se acaban los objetos de interés.

Cada acto de habla corresponde a un conjunto de réplicas que permiten a evitar que se repita la misma frase. Se dan ejemplos de algunas réplicas a continuación.

B. *Bloques de habla*

Los bloques del habla son actos de habla compuestos que están organizados como una secuencia en la cual los actos se interconectan a través de la misma intención comunicativa general de tal manera que la respuesta se espera a toda la secuencia y no a los actos de habla individuales.

En el caso que estamos considerando se utilizan tres tipos de bloques del habla: a) los bloques fronterizos (Borderline Blocks, BB); b) los bloques narrativos (Narrative Blocks, NB) y c) los bloques secundarios (Secondary Blocks, SB).

1) *Bloques fronterizos*

Los bloques fronterizos incluyen el bloque introductorio (Introductory Block, IB) y el bloque terminal (Terminal Block, TB). Cada bloque consta de una secuencia de actos de habla elementales.

El bloque introductorio (IB) es una secuencia de la siguiente forma (en el orden dado):

Saludo (Greeting) + Presentación de la persona (Introduction) + Propuesta de la visita (Tour-Proposal).

El bloque introductorio comienza su operación en la fase inicial, la cual pertenece al tipo de fases de habla (los tipos de fases se considerarán en adelante).

El bloque terminal (TB) es una secuencia de la siguiente forma:

El anuncio del fin de la inspección (Announce_End) + Autoevaluación (Evaluation) + Despedida (Farewell).

El bloque terminal se inicia en la fase final, la cual se determina con base en el historial del diálogo. Luego, en la representación formal del diálogo, la transición a la fase de pronunciación del bloque terminal depende de la limitación contextual del tipo NB(X) ∈ HD, donde HD es el historial del diálogo en transcurso (*history of the dialogue*), y X obtiene uno de los tres valores —V_M, M_L o E_S— es decir, se identifica el hecho de que los objetos planeados para la inspección ya han sido vistos todos.

2) *Bloques narrativos*

Los bloques narrativos (NB) son textos o presentaciones que contienen una cierta cantidad de información acerca de los objetos de demostración (en nuestro caso, los objetos son la_Venus_de_Milos, la_Mona_Lisa, el_Escriba_Sentado). Los bloques narrativos se ponen en operación al alcanzar el robot el objeto de demostración, o diciendo con más precisión, después de que el robot completa el bloque multimodal del tipo MMB2 (se considerará en adelante). A los bloques narrativos es posible subir la información necesaria con anticipación.

3) *Bloques secundarios*

El bloque secundario (Secondary Block, SB) es pronunciado por el robot en una situación relacionada con la posibilidad durante la inspección de las localidades utilizar los servicios (baño y fuente de agua fría), o visitar la cafetería y funciona en conjunto con el bloque multimodal secundario SMB (se considerará a continuación).

El bloque secundario es una desviación del objetivo principal: es una declaración de la intención de esperar hasta que el usuario termine su uso del servicio *(Promise_Wait) + la petición dirigida al usuario de informar de su regreso (Request_Return)*.

El bloque secundario se inicia cuando el usuario responde positivamente a la propuesta de utilizar el servicio.

C. *Bloques multimodales*

Los bloques multimodales son secuencias de actos de habla (SA) y acciones no verbales realizados en un orden fijo; se clasifican en los principales y los secundarios.

1) *Bloques multimodales principales*

El sistema incluye dos tipos de bloques multimodales principales (Main Multimodal Blocks, MMB).

MMB1: *la petición (comando) de seguir al robot pronunciada por él (Imper_Move) + el desplazamiento del robot al objetivo (Move(X))*, donde X ∈ {V_M, M_L, E_S}.

Cabe mencionar que estamos suponiendo que el usuario seguirá el robot, es decir, realizará el comando (Imper_Move), porque él estuvo de acuerdo con realizar la visita. Si es necesario verificar que el usuario está siguiendo al robot, es

posible introducir la confirmación adicional en el modelo: el robot puede pedir la confirmación para la continuación de la visita cada 5-10 minutos, y en el caso de ausencia de la respuesta o de la respuesta negativa finalizar la visita.

Se supone que el robot guía comienza su operación siempre desde el punto inicial fijo L (en nuestra aplicación particular, este punto es la entrada al museo pasando las escaleras) y se mueve hacia el primer objetivo aceptado por la ruta más corta, luego se desplaza al segundo objetivo por la ruta más corta, etc.; aquí recordamos que el robot es capaz de evitar los obstáculos que no están en su mapa. También es importante recordar que el robot del cual estamos hablando no puede ni subir ni bajar las escaleras. Aún más, él no puede detectar la escalera hacia abajo, porque su sensor laser solo funciona en un plano. Por lo que es necesario marcar las zonas con escaleras como áreas prohibidas para el robot.

Hay dos bloques de este tipo en nuestro modelo de diálogo. En el transcurso del diálogo, el primer bloque se inicia a través de la respuesta del usuario al acto de habla (SA) del robot del tipo Question_Altern(V_M, M_L, E_S) (véase más arriba). El segundo bloque se pone en ejecución al obtener la respuesta negativa del usuario a la pregunta del robot acerca del deseo de usar un servicio y se pronuncia al finalizar el primer bloque narrativo.

MMB2: *la detención del robot junto al objetivo (Stop_H) + el comando de detención (Imper_Stop)*. Este bloque se inicia mediante un acto visual representado en el modelo como Visual_Act(X), donde X ∈ {V_M, M_L, E_S}.

2) *Bloques multimodales secundarios*

Los bloques multimodales secundarios (Secondary Multimodal Blocks, SMB) son la secuencia de acciones relacionada con los objetos secundarios en la ruta (en nuestro caso son los servicios del baño, la cafetería y fuente de agua fría representados en el modelo como WC, Café y Cooler respectivamente).

SMB: *la detención junto al objeto secundario (Stop(X)) + la petición de llamar la atención al objeto secundario Announce_Utility(X) + la pregunta sobre el deseo de usar el servicio X, Question_Utility(X),* donde X ∈ {Cooler, Café, WC}. SMB, igual como MMB2, se inicia mediante el acto visual representado en el modelo como Visual_Act (X), donde X ∈ {Cooler, Café, WC}, es decir, el robot "ve" el objeto que está marcado en el mapa y se encuentra suficientemente cerca de él.

VII. Actos de habla y su fuerza ilocutiva

Según la descripción anterior, el robot por un lado es capaz de realizar las acciones ilocutivas independientes, y por otro lado las acciones ilocutivas dependientes en el sentido de Baranov y Kreidlin (1992) [23].

Las acciones ilocutivas independientes, es decir, iniciados por la "iniciativa" del robot, son bloques del habla de los tipos BB, NB, los bloques multimodales MMB2 (Stop_X + Imper_Stop) y SMB.

Las acciones ilocutivas dependientes, es decir, iniciados con la condición de que el usuario realizó algún acto de habla (SA), son el bloque del habla SB, los actos de habla Question_Altern y Ask_Confirm, y el bloque MMB1 (Imper_Move + Move_X).

Cada tipo de los actos de habla o los bloques del habla, así como el componente de habla del bloque multimodal, se relaciona en el módulo de síntesis de texto con un conjunto de los patrones de habla (réplicas) de los cuales el robot hace la selección memorizando cual opción ya fue seleccionada para elegir en la siguiente iteración otro patrón del mismo tipo y guardarlo en la memoria, etc. de elegir otra plantilla del mismo tipo y recordarlo, etc. Con esta estrategia el comportamiento del robot es más natural.

Cada acto de habla ilocutivo independiente y el bloque del habla del robot en un diálogo del tipo que estamos modelando, impone cierto compromiso comunicativo (*committment*) para el usuario, bajo la condición que por lo general él debe respetar las máximas de comunicación de Grice [3], es decir, él adopta un propósito común en el diálogo y respeta la dirección del desarrollo del diálogo y no tiene la intención de confundir al robot, etc. Todos los actos de habla del usuario en este tipo de diálogo son dependientes desde el punto de vista ilocutivo: por las limitaciones consideradas anteriormente no se supone que el usuario tome la iniciativa en el diálogo, por ejemplo, hacer una pregunta al robot aun acerca de la algo relacionado con la visita. El usuario debe responder de acuerdo con la finalidad ilocutiva de las acciones del robot, pero si él toma "libertades", el robot lo tratará como un error e intentará de "obtener" la reacción esperada. En el futuro, se planea eliminar esta restricción para el diálogo.

VIII. Actos de habla del usuario

En virtud de las restricciones anteriores, el repertorio de actos de habla que se esperan del usuario, consta de los siguientes SA, a las siglas de las cuales se agrega el prefijo uSA (user Speech Act):

- Consentimiento (uSA(Yes)) o Denegación (uSA(No)) como reacciones a la propuesta de inspección, que forma parte del bloque IB, y corresponde a la propuesta de usar un objeto secundario, que está incluida en el bloque SB1.
- La respuesta con la alternativa seleccionada a la pregunta del robot acerca de con cual objeto se empieza la visita (véase más arriba Question_Altern(X)), en nuestro caso particular).
- El mensaje del regreso a su lugar después de usar el objeto secundario uSA(Ready).
- Todo SA no previsto en el sistema para este tipo del diálogo con el usuario se interpreta como error: Inesperado (no interpretable) acto de habla del usuario uSA(Error).

Si el sistema no puede asociar un acto de habla del usuario con alguno de esos cinco tipos ilocutivos, el robot vuelve a la fase inmediata anterior al acto de habla reconocido como

uSA(Error), y repite aquel acto de habla después del cual el usuario generó el acto de habla uSA (Error).

Si en relación con el tipo ilocutivo de enunciados el usuario se limita al ámbito de las obligaciones de comunicación impuestas por los actos de habla del robot, entonces en relación con la expresión de los actos de habla permitidos, el usuario tiene cierta libertad. El analizador de voz identifica unidades marcadas en el diccionario del sistema en correspondencia con su contenido proposicional o función ilocutiva. El módulo del control del diálogo asigna a la réplica del usuario en un momento dado del diálogo aquel tipo ilocutivo del conjunto de los posibles tipos, que corresponde a los marcadores de las unidades reconocidas. Por ejemplo, las lexemas *Venus_de_Milos* etc. en la réplica generada por el usuario en respuesta al acto de habla del robot del tipo "Propuesta de seleccionar el primer objeto de inspección del conjunto de las alternativas Question_Altern(V_M, M_L, E_S)" serán el marcador del tipo ilocutivo "Selección del objeto Venus_de_Milos uSA(V_M)". En el futuro, es posible realizar un análisis más complejo, porque el analizador sintáctico permite trabajar con el árbol de análisis sintáctico de cada frase del diálogo.

IX. Modelo del diálogo

El modelo del diálogo se puede representar como una red de transiciones del robot de un estado del diálogo al otro dependiendo de la información verbal o visual recibida, el historial del diálogo y las intenciones del usuario. Los estados se representan como reglas, otra posibilidad sería construir un autómata finito [11].

En nuestro modelo los estados del diálogo son de los siguientes tipos:

- el estado de habla (talk, talking state),
- el estado de percepción (perc., perception state),
- el estado de movimiento (move, moving state),
- el estado final (Final, Final State).

La transición de una fase a otra se realiza después de ejecutar las acciones (de habla y/o de movimiento) del robot. Las acciones se ponen en ejecución como reacción al acto de habla del usuario uSA(X), donde X toma un valor del conjunto de los tipos ilocutivos SA del usuario, o un acto de la percepción visual del robot de un objeto Y, Visual_Act(Y), donde Y toma valores del conjunto de los objetos marcados en el mapa interno del robot.

En consecuencia, la acción del robot dependiendo de su estado actual está dada por el par "(el acto de habla de usuario como entrada, el estado del robot) → (Acción del robot, el estado del robot como salida)".

Por lo tanto, el modelo del diálogo se puede representar en la forma de un conjunto de fases y acciones las cuales llevan el diálogo de una fase a otra. También se puede utilizar la recursividad, es decir, invocar una función dentro de sí misma, pero evitar ciclos infinitos el posible nivel de recursividad debe ser limitado.

En la tabla 1 se presenta la lista de reglas que describen la gran parte de nuestro modelo. Las reglas tienen la forma:

el estado del robot & condición (si se cumple), la acción (qué hacer) → un nuevo estado del robot

Entonces, para aplicar una regla, el robot debe estar en cierto estado y la condición se debe cumplir. En este caso, el robot realiza la acción correspondiente y pasa a un nuevo estado. Después de cada uso de una regla, esta regla se agrega al historial del diálogo HD (*history of dialogue*). A veces, la condición puede estar ausente: esto significa que la condición es verdadera, es decir, la acción debe ser ejecutada en todo caso. A veces la acción no es necesaria, es decir, no existe la necesidad para que el robot haga algo. Tales casos se simbolizan con ∅. Cabe mencionar que después de la aplicación de una regla, el robot siempre pasa a un nuevo estado. En la tabla 1 se dan ejemplos de las reglas. En esa tabla no se representan los actos de habla correspondientes a objetos M_L y CAFÉ, pero se hacen de manera similar. En el futuro, se planea usar actos de habla con parámetros, para no depender de la lista concreta de objetos.

Por ejemplo, las dos primeras reglas se interpretan de la siguiente manera: mientras el robot se encuentra en la fase inicial de habla, él realiza dos actos de habla por su "iniciativa", es decir, estos actos no se evocan por las causas externas: primero, el robot ejecuta el "bloque introductorio (IB)", luego pronuncia la pregunta acerca de la intención de hacer el recorrido Tour-Question, y después pasa a la fase de percepción (de la información de entrada) $perc_1$, etc.

X. Conclusiones

En este artículo, se presentó el módulo de control del diálogo en la comunicación con un robot guía móvil. El módulo consta del modelo del diálogo, la descripción de los actos de habla y los bloques del habla, a través de los cuales se puede construir cada componente del modelo. Todo el modelo, salvo los patrones de respuestas del robot, es independiente del idioma.

Aunque el artículo describe sólo una aplicación limitada del modelo, el modelo propuesto es universal en su arquitectura y proporciona los medios para el modelado formal del diálogo con el robot móvil autónomo usando el material lingüístico diverso, ya que el inventario de los actos de habla y las estrategias comunicativas del hablante, a diferencia de las restricciones que imponen diferentes idiomas a la estructura sintáctica [22], son comunes para todas las lenguas naturales.

Agradecimientos

Trabajo realizado con el apoyo de gobierno de la Ciudad de México (proyecto ICYT-DF PICCO10-120), el apoyo parcial del gobierno de México (CONACYT, SNI) e Instituto Politécnico Nacional, México (proyectos SIP 20144274, 20144534; COFAA), proyecto FP7-PEOPLE-2010-IRSES: "Web Information Quality – Evaluation Initiative (WIQ-EI)" European Commission project 269180, y Ministry of

Tabla 1. Lista de reglas de manejo de diálogo para dos objetos.

Estado previo		Actos de habla	Acción		Estado nuevo
$talk_1$	&	∅	IB	→	$talk_2$
$talk_2$	&	∅	Tour-Question	→	$perc_1$
$perc_1$	&	uSA(Yes)	Question_Altern(V_M, E_S, M_L)	→	$perc_2$
$perc_1$	&	uSA(No)	Closure	→	Final
$perc_1$	&	uSA(Error)	∅	→	$talk_2$
$perc_2$	&	uSA(V_M)	MMB1(V_M)	→	$move_1$
$perc_2$	&	uSA(E_S)	MMB1(E_S)	→	$move_2$
$perc_2$	&	uSA(E_S)	MMB1(M_L)	→	$move_3$
$perc_2$	&	uSA(Error)	∅	→	$talk_3$
$talk_3$	&	∅	Question_Altern(V_M, E_S, M_L)	→	$perc_2$
$move_1$	&	visual_act(V_M)	MMB2(V_M)	→	$talk_4$
$move_1$	&	visual_act(Cooler)	SMB(Cooler)	→	$perc_3$
$perc_3$	&	uSA(Yes)	SB	→	$perc_4$
$perc_3$	&	uSA(No)	MMB1(V_M)	→	$move_1$
$perc_4$	&	uSA(Ready)	MMB1(V_M)	→	$move_1$
$talk_4$	&	∅	NB(V_M)	→	$talk_5$
$talk_5$	&	NB(E_S) ∉ HD	MMB1(E_S)	→	$move_2$
$talk_5$	&	NB(E_S) ∈ HD	TB	→	Final
$move_2$	&	Visual_Act(E_S)	MMB2(E_S)	→	$talk_6$
$move_2$	&	Visual_Act(WC)	SMB(WC)	→	$perc_5$
$perc_5$	&	uSA(Yes)	SB	→	$perc_6$
$perc_5$	&	uSA(No)	MMB1(E_S)	→	$move_2$
$perc_6$	&	uSA(Ready)	MMB1(E_S)	→	$move_2$
$talk_6$	&	∅	NB(E_S)	→	$talk_7$
$talk_7$	&	NB(V_M) ∉ HD	MMB1(V_M)	→	$talk_7$
$talk_7$	&	NB(V_M) ∈ HD	TB	→	Final

education and Science of Russian Federation, federal project 2685 “Parametric description of grammar systems”.

REFERENCIAS

[1] “Dialog with Robots,” in *AAAI 2010, Fall Symposium*, November 2010, Arlington VA, 2010.

[2] S. A. Green, M. Billinghurst, C. X. Qi, G. J. Chase, “Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design,” *International Journal of Advanced Robotic Systems*, vol. 5, pp. 1–18, 2007.

[3] P. Grice, “Logic and conversation,” in *Syntax and Semantics, 3: Speech Acts*, P. Cole & J. Morgan (eds.), New York: Academic Press, 1975 (Reprinted in: *Studies in the Way of Words*, ed. H. P. Grice, Cambridge, MA: Harvard University Press, 1989, pp. 22–40).

[4] G. J. M. Kruijff, H. Zender, P. Jensfelt, H. I. Christensen, “Clarification dialogues in human-augmented mapping,” in *Human Robot Interaction’06*, Salt Lake City, Utah, USA, 2006.

[5] G. J. M. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, H. Zender, I. Kruijff-Korbayova, N. Hawes, “Situated dialogue processing for human-robot interaction,” *Cognitive Systems*, pp. 311–364, 2010.

[6] H. Kuzuoka, K. Yamazaki, A. Yamazaki, J. Kosaka, Y. Suga, C. Heath, “Dual ecologies of robot as communication media: Thoughts on coordinating orientations and projectability” in *CHI’04 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2004, pp 183–190.

[7] S. Lemaignan, R. Ros, R. Alami, “Dialogue in situated environments: A symbolic approach to perspective-aware grounding, clarification and reasoning for robot,” in *Proc. Robotics, Science and Systems, Grounding Human-Robot Dialog for Spatial Tasks workshop*, 2011.

[8] M. Marge, A. Pappu, B. Frisch, Th. K. Harris, A. Rudnicky, “Exploring Spoken Dialog Interaction in Human-Robot Teams,” in *Proceedings of Robots, Games, and Research: Success stories in USA, RSim IROS Workshop*, St. Louis, MO, USA, 2009.

[9] L. Padró, M. Collado, S. Reese, M. Lloberes, I. Castellón, “FreeLing 2.1: Five Years of Open-Source Language Processing Tools,” in *Proceedings of 7th Language Resources and Evaluation Conference (LREC 2010)*, ELRA, La Valletta, Malta, 2010.

[10] P. Pakray, U. Barman, S. Bandyopadhyay, A. Gelbukh, "A Statistics-Based Semantic Textual Entailment System," in MICAI 2011, *Lecture Notes in Artificial Intelligence*, vol. 7094, Springer, pp. 267–276, 2011.

[11] L. A. Pineda, V. M. Estrada, S. R. Coria, J. F. Allen, "The obligations and common ground structure of practical dialogues," *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, vol. 11, no. 36, pp. 9–17, 2007.

[12] G. Salton, A. Wong, C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[13] J. Searle, "Indirect speech acts," in *Syntax and Semantics, 3:Speech Acts*, P. Cole & J.L. Morgan (eds.), New York: Academic Press, 1975. pp. 59–82. (Reprinted in: *Pragmatics: A Reader*, S. Davis (Ed.), Oxford: Oxford University Press, pp. 265–277, 1991).

[14] G. Sidorov, *Non-linear construction of n-grams in computational linguistics: syntactic, filtered, and generalized n-grams*, 2013, 166 pp.

[15] G. Sidorov, "Desarrollo de una aplicación para el diálogo en lenguaje natural con un robot móvil," *Research in computing science*, vol. 62, pp. 95–108, 2013.

[16] G. Sidorov, F. Velasquez, E. Stamatatos, A. Gelbukh, L. Chanona-Hernández, "Syntactic N-grams as Machine Learning Features for Natural Language Processing," *Expert Systems with Applications*, vol. 41, no. 3, pp. 853–860, 2014.

[17] E. A. Sisbot, R. Ros, R. Alami, "Situation assessment for human-robot interaction," in *Proc. of 20th IEEE International Symposium in Robot and Human Interactive Communication*, 2011.

[18] M. Skubic, D. Perzanowski, A. Schultz, W. Adams, "Using Spatial Language in a Human-Robot Dialog," in *IEEE International Conference on Robotics and Automation*, Washington, D.C., 2002.

[19] *Systems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vienna, Austria, Association for Computing Machinery, 2004.

[20] E. A. Topp, H. Huttenrauch, H. I. Christensen, E. K. Severinson, "Bringing Together Human and Robotic Environment Representations—A Pilot Study," in *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, 2006.

[21] Y. Wilks, N. Webb, A. Setzer, M. Hepple, R. Catizone, "Machine learning approaches to human dialogue modeling," in J.C.J. van Kuppevelt et al. (eds.), *Advances in Natural Multimodal Dialogue Systems*, pp. 355–370, 2005.

[22] P. M. Arkadiev, N. V. Serdobolskaya, A. V. Zimmerling, "Project of a typological database of syntactic constraints on movement," in *Computational linguistics and intellectual technologies, proceedings of the international conference "Dialogue. 2010"*, Moscow, vol. 9, no. 16. pp. 437–441, 2010 (in Russian).

[23] A. N. Baranov, G. E. Kreidlin, "Illocutive forcement in dialogue structure," *Problems of Linguistics [Voprosy jazykoznania]*, vol. 2. pp. 84–99, 1992 (in Russian).

[24] I. M. Kobozeva, N. I. Laufer, I. G. Saburova, "Modelling conversationtion in man-machine systems," in *Linguistic support of information systems [Lingvisticheskoe obespechenie informacionnyh system]*, INION of Academy of Sciences of USSR, Moscow, 1987 (in Russian).

[25] M. V. Prischepa, "Development of the user profile based on the psychological aspects of human interaction with the informational mobile robot," *SPIIRAS proceedings [Trudy SPIIRAN]*, vol. 2, no. 21, pp. 56–70, 2012 (in Russian).