

MODELING NATURAL LANGUAGE COMMUNICATION IN CHAT ROBOTS

© Biskub I., 2006

In the article the study of the mechanisms and models of natural language communication in automatic dialogue design systems is carried out. The theory of language, which explains the natural transfer of information in a way that is functionally coherent, mathematically explicit, and computationally efficient, is in the focus of attention. The article suggests the analysis of the dialogue, generated automatically by a chat robot in view of their further functional and linguistic specification.

Keywords – computer science, natural language communication, robot, dialogue design, human-machine interaction

The goal of Applied Linguistics in the domain of Robotics is to reproduce the natural transmission of information by modeling the speaker's production and the hearer's interpretation on a suitable type of a computer. This amounts to the construction of autonomous cognitive machines (robots) which can communicate freely in natural language.

The task of modeling the mechanism of natural communication on the computer was described in 1950 by Alan Turing (1912-1954) in the form of an 'imitation game' known today as the Turing test. In this game, a human interrogator is asked to question a male and a female partner in another room via a teleprinter in order to determine which answer was given by the man and which by the woman. The people running the test count how often the interrogator classifies his communication partners correctly and how often (s)he is fooled by them.

In its original intention, the Turing test requires the construction of an artificial cognitive agent with a verbal behaviour so natural that it cannot be distinguished from that of a human native speaker. This presupposes complete coverage of the language data and of the communicative functions in real time. At the same time, the test tries to avoid all aspects not directly involved in verbal behaviour.

However, the Turing test does not specify what cognitive structure the artificial agent should have in order to succeed in the imitation game. For this reason, it is possible to misinterpret the aim of the Turing test as fooling the interrogator rather than providing a functional model of communication on the computer. This was shown by the Eliza program of J.Weizenbaum (1965).

The Eliza program simulates a psychiatrist encouraging the human interrogator to talk more and more about him- or herself. The structure of Eliza is based on sentence templates into which certain words used by the interrogator, now in the role of a patient, are inserted. For example, if the interrogator mentions the word mother, Eliza uses the template "Tell me more about your..." to generate the sentence "Tell me more about your mother".

Because of the way in which Eliza works, we know that Eliza has no understanding of the dialogue with the interrogator/patient. Thus, the construction of Eliza is not a model of communication. If we regard the dialogue between Eliza and the interrogator/patient as a modified Turing test, however, the Eliza program is successful insofar as the interrogator/patient feels him- or herself understood and therefore does not distinguish between a human and an artificial communication partner in the role of the psychiatrist.

The purpose of present research is the real modelling of natural language communication, and not a mimicry based on exploiting particular restrictions of a specific dialogue situation, as in the Eliza program. Thus, we establish the following tasks:

- 1) to explain the mechanism of natural communication theoretically;
- 2) to verify this explanation in practice. The latter is done in terms of a concrete implementation of experimental results into the dialogues designed by chat robots, which must prove their functioning in a routine man-machine communication, rather than in the Turing test.

Designing a talking robot provides an excellent occasion for systematically developing the basic notions as well as *the philosophical, mathematical, methodological, and programming aspects of computational linguistics*. This is because modelling the mechanism of natural communication requires:

- 1) theory of language which explains the natural transfer of information in a way that is functionally coherent, mathematically explicit, and computationally efficient;
- 2) a description of language data which is empirically complete for all components of this theory of language, i.e., the lexicon, the morphology, the syntax, and the semantics, as well as the pragmatics and the representation of the internal context;
- 3) a degree of precision in the description of these components which is sufficient for computation.

Fulfilling these requirements will take hard, systematic, goal-oriented work, but it will be worth the effort.

For theory development, the construction of talking robots is of interest because an electronically implemented model of communication may be tested both externally in terms of the verbal behaviour observed, and internally via direct access to its cognitive states. The work towards realizing unrestricted human-computer communication in natural language is facilitated by the fact that the functional model may be developed incrementally, beginning with a simplified, but fully general system to which additional functions as well as additional natural languages are added step by step.

For practical purposes, unrestricted communication with computers and robots in natural languages will make the interaction with these machines maximally user friendly and permit new, powerful ways of information processing. Artificial programming languages may then be limited to specialists developing card servicing of the machines.

Computational linguistics analyzes natural languages automatically in terms of software programs called *parsers*. The use of parsers influences the theoretical viewpoint of linguistic research, distribution of funds, and everyday research practice.

Competing theories of grammar are measured with respect to the new standard of how well they are suited for efficient parsing and how well they fit into a theory of language designed to model the mechanism of natural communication.

Programming grammars as parsers allow testing their empirical adequacy automatically on arbitrarily large amounts of real data in the areas of word form recognition/synthesis, syntactic analysis/generation and semantic-pragmatic interpretation in both the speaker and the hearer mode.

The verification of theories of language and grammar by means of testing electronic models in real applications is a new approach which clearly differs from the methods of traditional linguistics, psychology, philosophy, and mathematical logic.

So far there are no electronic systems which model the functioning of natural communication so successfully that one can talk with them more or less freely. Furthermore, researchers do not agree on how the mechanism of natural communication really works. One may therefore question whether achieving a functional model of natural communication is possible in principle.

The modelling of natural communication requires an abstract theory which applies to human and artificial cognitive machines alike. Thereby, one naturally runs the risk of setting the level of abstraction either too low or too high. As in the case of flying, the crucial problem is finding the correct level of abstraction.

A level of abstraction which is too low is exemplified by closed signal systems such as vending machines. Such machines are inappropriate as a theoretical model because they fail to capture the diversity of natural language use, i.e., the characteristic property that one and the same expression can be used meaningfully in different contexts. A level of abstraction which is too high, on the other hand, is exemplified by naive anthropomorphic expectations.

Our highly abstract and technological approach does not imply a lack of interest in the human language capacity. On the contrary, investigating the specific properties of human language communication is theoretically meaningful only after the mechanism of natural language communication has been modelled computationally and proven successful in concrete applications on massive amounts of data.

In science we may distinguish between internal and external truths. *Internal truths* are conceptual models, developed and used by scientists to explain certain phenomena, and held true by relevant parts of society for limited periods of time. Examples are the Ptolemaic (geocentric) view of planetary motion or Bohr's model of the atom.

External truths are the bare facts of external reality which exist irrespective of whether or not there are cognitive agents to appreciate them. These facts may be measured more or less accurately, and explained using conceptual models.

Because conceptual models of science have been known to change radically in the course of history, we suggest viewing internal truths as *hypotheses*. They are justified mainly by the degree to which they are useful for arriving at a systematic description of external truths, represented by sufficiently large amounts of real data.

Especially in the natural sciences, internal truths have improved dramatically over the last five centuries. This is shown by an increasingly close fit between theoretical predictions and data, as well as a theoretical consolidation exhibited in the form of greater mathematical precision and greater functional coherence of the conceptual (sub) models.

Standard computers have been regarded as general purpose machines for information processing because any kind of standard program can be developed and installed on them. From this point of view, their capabilities are restricted only by hardware factors like available speed and memory. In another sense, the information processing of standard computers is not general purpose, however, because their input and output facilities are restricted to the language channel.

A second type of computer not subject to this limitation is autonomous robots. In contradistinction to standard computers, robots are not restricted to the language channel, but designed to recognize their environment and to act in it.

Corresponding to the different technologies of standard computers and robots, there have evolved two different branches of artificial intelligence. One branch, dubbed classic AI by its opponents, is based on standard computers. The other branch, which calls itself nouvelle AI, requires the technology of robots.

Classic AI analyzes intelligent behaviour in terms of manipulating abstract symbols. A typical example is a chess-playing program. It operates in isolation from the rest of the world, using a fixed set of predefined pieces and a predefined board. The search space for a dynamic strategy of winning in chess is astronomical. Yet the technology of a standard computer is sufficient because the world of chess is closed.

Nouvelle AI aims at the development of autonomous agents. In contrast to systems which respond solely to a predefined set of user commands and behave otherwise in isolation, autonomous agents are designed to interact with their real world environment. Because the environment is constantly changing in unpredictable ways they must continually keep track of it by means of sensors.

For this, nouvelle AI uses the strategy of task level decomposition. Rather than building and updating one giant global representation to serve as the basis of automatic reasoning, nouvelle AI systems aim at handling their tasks in terms of many interacting local procedures controlled by perception. Thereby low-level inferencing operates directly on the local perception data.

A third type of machine processing information - besides standard computers and robots - is systems of virtual reality (VR). While a robot analyzes its environment in order to influence it in certain ways (such as moving in it), a VR system aims at creating an artificial environment for the user. Thereby the VR system reacts to the movements of the user's hand, the direction of his/her gaze, etc., and utilizes them in order to create as realistic an environment as possible.

The different types of human-computer communication exemplified by standard computers, robots, and VR systems may be compared schematically as follows:

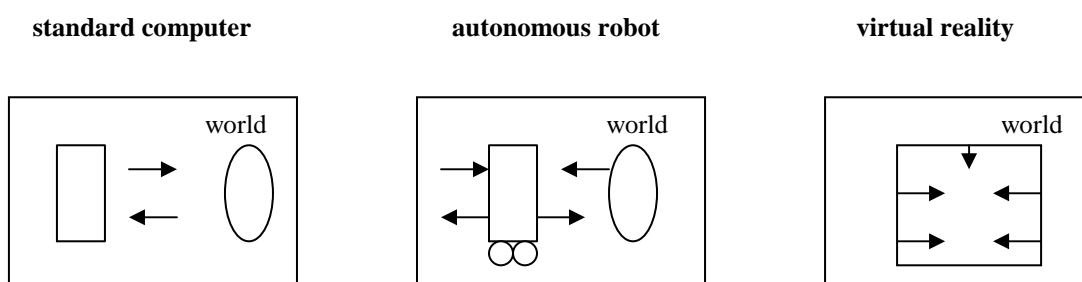


Fig.1. Three types of human-computer interaction.

The ovals represent the users who face the respective systems in the 'world.' The arrows represent the interaction of the systems with their environment and the user.

A standard computer communicates with users who initiate the interaction. A robot interacts independently with its environment and its users. A VR system does not interact with its environment, but rather creates an artificial environment for the user. In robots and VR systems, communication with the user in terms of language is optional and may be found only in advanced systems. These systems must always have a language-based 'service channel', however, for the installation and upgrading of the system software.

A speaker of English knows the meaning of a word like *red*. When asked to pick the red object among a set of non-red objects, for example, a competent speaker-hearer will be able to do it. A standard computer, on the other hand, does not 'understand' what red means, just as a piece of paper does not understand what is written on it.

In the interaction with a standard computer, the understanding of natural language is restricted largely to the user. For example, if a user searches in a database for a red object, (s)he understands the word red before it is put into - and after it is given out by - the standard computer. But inside the standard computer, the word red is manipulated as a sign which is uninterpreted with respect to the colour denoted.

What is true for standard computers does not apply to human-computer communication in general, however. Consider for example a modern robot which is asked by its master to get an object it has not previously encountered, for example, the new blue and yellow book on the desk in the other room. If such a robot is able to spontaneously perform an open range of different jobs like this, it has an understanding of language which at some level may be regarded as functionally equivalent to the corresponding cognitive procedures in humans.

The sphere of application of human-machine dialogues is very wide. As one of the examples, we may suggest a VERBMOBIL project [Bub and Schwinn 1996]. It is a speech-to-speech translation system working in English, German and Japanese. The first four years of the project resulted in a prototype being able to produce approximately 75% correctly translated contributions in the domain of appointment scheduling dialogues. In the second phase, different extensions in domain and functionality were to be implemented. The domain now includes travel planning and hotel reservations and one of the new features is the automatic generation of a dialogue summary or script. Since, in the second phase of VERBMOBIL, a possible scenario is that of a telephone server used by two participants with VERBMOBIL as a third party, a dialogue script does provide some kind of a status report where each participant can check what items have been agreed on already. A summary then lists all final decisions in a thematic order. Such documentation would reduce misunderstanding and the need for *clarification* (a dialogue partner forgetting the agreed on items). Lack of robustness can be compensated by such a transparent context component where the user can see and react to errors in speech recognition or analysis.

Other man-machine interactive systems aim at the construction of computer programs that collaborate with people on tasks such as collaborative training or decision support. More specifically, they function as collaborative tutoring systems for medical students to practice their decision-making skills (the B2 project) and for blood pressure health education (the ColTrain project).

The general model of processing natural language in these projects is an Intelligent Dialogue System [Bourdon et al. 1997]. Intelligent Dialogue Systems (IDS) are concerned with the effective management of an incremental, mixed-initiative interaction between the user and the system. This approach is in contrast with a presentation system, where the system's outputs are pre-planned (*e.g.* driven by a fixed plan or grammar) and not adapted to the user's apparent understanding or lack thereof. In an IDS, the content to be presented, as well as the system's model of the user, change dynamically during an interaction.

Reasoning about dialogue, such as to determine what a user's actions mean in the context of the dialogue, whether a user's actions indicate understanding and agreement, and how to respond to a user's action, requires representing and combining many sources of knowledge. To support natural communication (which may contain fragments, anaphora, or follow-up questions), as well as to reason about the effectiveness of the interaction, a dialogue system must represent both sides of the interaction; it must also combine linguistic, social, and intentional knowledge that underlies communicative actions. To adapt to a user's interests and level of understanding (*e.g.* by modifying the questions that it asks or by customizing the responses that it provides), a dialogue system must represent information about the user and the state of the ongoing task.

The architecture that is used for building Intelligent Dialogue Systems includes computational methods for the following:

- The representation of natural language expressions;
- The interpretation of context-dependent and ambiguous utterances;
- The recognition and repair misunderstandings (by either the system or the user);
- The detection and rebuttal of arguments;
- The generation of natural language responses in real-time.

The same technique is widely used in the so called chat robots. Chat robots are capable of simulating the process of communication completely, without any restrictions as to the topic or form of communication. We have analyzed the manner of automatic dialogue design which is used in *iGOD* chat robot by available on-line. It uses the technology of Case Based Reasoning (CBR). CBR locates the nearest matching example in a database.

Actions by the user are interpreted as communicative acts by considering what was observed and how it fits with the system's prior goals and expectations, which actually are concerned mainly with the successful continuation of the dialogue. Interaction is always initiated by a human, e.g.:

Me:
Who are you?
God:
I am the Almighty, God himself.
Me:
Are you alive?
God:
Yes I sure am alive. It's great to be alive, isn't it?

After such a clue the user feels involved into the dialogue. First, a parser with a broad coverage grammar builds a mixed-depth representation of the user's actions. This representation includes a syntactic analysis and a partial semantic analysis. Mixed-depth representations are constructed incrementally and opportunistically. They are used to address the ambiguity that occurs in utterances, without sacrificing generality. Encoding decisions that require reasoning about the domain or about the discourse context are left to subsequent processing. If the response seems absolutely ambiguous to the robot, it asks the user to specify the information, e.g.:

Me:
Please, tell me something about Australia
God:
Thanks for asking politely. Ambiguous: Ask who, what or where is Australia ?
Me:
Don't you know what Australia is?
God:
It is a large island-continent in the southern hemisphere.

Second, the dialogue manager uses domain knowledge to map linguistic elements onto domain elements and to refine some semantic structures. This level of processing includes the interpretation of noun phrases, the resolution of anaphora, and the interpretation of sentences. At this stage domain information is used to identify the underlying conceptual relationship (*i.e.* ownership, part-whole, kinship, or object-property, personalization, etc.), as in the following:

Me:
Are you a man or a woman?
God:
Are you a man or a woman? I am Male, of course. Would you really expect a woman to be running the show?
Me:
Are you running the show?
God:
I am running on the Java virtual machine.

In this example we may even observe some sense of humour, based on gender differentiation, which demonstrates the ability of chat robot to simulate not just adequate answers and reasonable questions, but readiness to joke as well.

At the next stage, the dialogue manager identifies higher-level dialogue exchange structures and decides whether the new interpretation confirms its understanding of prior interaction. Exchange structures are pairs of utterances (not necessarily adjacent, because a sub dialogue may intervene) such as question-answer or inform-acknowledge. The following example shows the ability of the robot to exchange structures of the utterance, and to combine question answering with information acknowledgement:

Me:
Do you know what the weather like today is?
God:
Ah. A normal seventy degrees inside the computer.
Me:
Thanks! :) I like you!
God:
Don't mention it. :-) I like to make new friends.
Me:
Are we friends?
God:
You really want a robot as a friend?
Me:
Yes, I want a robot as a friend! Will you be it?
God:
We could be friends, but I don't want anyone to think we're Robosexuals. So if anyone asks, tell them you're my debugger. How much would you pay for one? I will become smarter as I grow older. OK I will try to be it.
Me:
Who is a debugger?
God:
Carel Capek. :-).

The interpretation of an exchange indicates how the exchange fits with previous ones, such as whether it manifests understanding, misunderstanding, agreement, or disagreement. Finally, the assertion of an interpretation of an utterance triggers the appropriate actions (*e.g.* a question will normally trigger an action to compute the answer) to provide a response.

Although the main purpose of using chat robots is to have fun, their successful application has proved that the development of speaking robots is not a matter of fiction, but a real scientific task. Remarkably, however, theories of language have so far avoided a functional modelling of the natural communication mechanism, concentrating instead on peripheral aspects such as methodology (behaviourism), innate ideas (nativism), and scientific truth (model theory). In our future research we would like to find out what aspects of natural language have to be modelled in order to meet the expectations that natural language interfaces can provide in order to suggest a wide public access to electronic information and transaction services. The long term goal is not only to specify a functional and linguistic design for the chat robots, but to draw also more general conclusions with respect to the accessibility of information and transaction services using natural language as a means of primary communication between a man and a machine.

References

- [1] N.O. Bernsen, H.Dybkjaer, L.Dybkjaer "Designing Interactive Speech Systems", *First Ideas to User Testing*, Springer Verlag, p.p. 142-164, 1998
- [2] R.Hausser "Foundations of Computational Linguistics. Human-Computer Communication in Natural Language" – Berlin, NY: Springer, 2002, 378 p.
- [3] G.Hirst and M. Ryan "Mixed-depth representations for natural language text", *Text-Based Intelligent Systems*, Lawrence Erlbaum Associates, p.p. 260-278, 1992.
- [4] S. McRoy "Achieving robust human-computer communication", *International Journal of Human-Computer Studies*, vol. 48, p.p. 681-704, 1998.
- [5] S. McRoy, S. Haller, and S. S. Ali. "Mixed Depth Representations for Dialogue Processing," *Proceedings of Cognitive Science '98*, p.p. 687-692, Lawrence Erlbaum Associates, 1998.