PUC

Series: Monografias em Ciência da Computação, 31/89

SPEECH ACTS IN COMPUTER-HUMAN INTERFACES

Clarisse S. Souza

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

RUA MARQUÊS DE SÃO VICENTE, 225 – CEP 22453

RIO DE JANEIRO – BRASIL

PUC/RJ - DEPARTAMENTO DE INFORMÁTICA

Series: Monografias em Ciência da Computação, 31/89

Editor: Paulo Augusto Silva Veloso December, 1989

SPEECH ACTS IN COMPUTER-HUMAN INTERFACES

Clarisse S. Souza

This work has been partially sponsored by FINEP

In charge of publications:

Rosane Teles Lins Castilho
Assessoria de Biblioteca, Documentação e Informação
PUC RIO, Departamento de Informática
Rua Marquês de São Vicente, 225 - Gávea
22453 - Rio de Janeiro, RJ
BRASIL

Tel.: (021)529-9386 TELEX: 31078 FAX: (021)274-4546
BITNET: userrtlc@lncc.bitnet

SPEECH ACTS IN COMPUTER-HUMAN INTERFACES

Clarisse Sieckenius de Souza

Departamento de Informática - PUC/RJ
R. Marquês de São Vicente 225
22453 - Rio de Janeiro/RJ - Brazil
e-mail: usercsds@lncc.bitnet

ABSTRACT

Communication between users and intelligent systems involves a higher degree of complexity if compared to conventional computer-human interaction (CHI), specially in terms of the role **language** plays in the computations performed by systems. This paper claims that intelligent systems are essentially **linguistic systems** and that their actions are, therefore, intrinsicly related to the notion of **speech acts.** Within this frame, basic guidelines for interface design are discussed and an extension to menu-driven natural language understanding systems is proposed as a means to deal with speech acts in CHI.

Keywords: Computer-Human Interfaces, Natural Language Processing, Speech Acts.

Computer-human interaction (CHI) within the scope of intelligent systems faces the challenge of providing an adequate account of meaning which is neither that of artificial languages, nor that of natural language. On the one hand, the semantics of conventional programming languages, and apparently also that of the so-called fifth-generation languages, fails to capture subtle cognitive aspects involved in intelligent user/system dialogues, as is the case for the rendition of **explanations** in expert systems. On the other, the semantics of natural language has not yet been satisfactorily formalized to provide the basis for machine understanding and generation of linguistic material.

Although attractive at first sight, the idea of interacting with computers in Natural Language raises relevant questions [Kam88] among which those of cognitive nature are the most important ones. When humans communicate with each other by linguistic means, their language reflects the presupposition that participants share the same **communicative competence**, which is not merely that of mastering the same language, but also that of having reasoning and perceiving abilities that are common to the species. The problem is, then, one of endowing machines with human cognition and perception, or an equivalent form of these [Dre72].

The realization that language, cognition and perception are intrinsicly related - as evidenced by psycholinguistic research - shifts the focus of computer/human interaction from natural language to artificial agency, since it is the global **behavior** (or **acting**) of intelligent machines that calls for an adequate characterization in the field of artificial intelligence. Without this, attempts at natural language processing cannot be fully sound [C&L87].

This paper discusses relevant design issues for a system's interface in terms of its rational and perceptual abilities and explores how these can possibly interact with its linguistic abilities in conversing with potential users. The emphasis of the presentation lies in the rationale behind the guidelines it assumes for building adequate interfaces for intelligent systems. It is not our goal to discuss the fundamentals of rational agency, or to provide a full specification of an intelligent system's linguistic competence.

BASIC FEATURES OF INTELLIGENT SYSTEMS

Intelligent systems have been defined in a variety of ways. The common substratum of all definitions, however, seems to be that such systems show the ability to **reason** (i.e. to perform computations on knowledge) and to **explain** (in different degrees of adequacy and efficiency) their behavior. Reasoning implies an internal representation of rules and facts that are reasoned about, whereas explaining implies an internal representation of reasoning itself - a meta-level of rules and facts. It follows from this that intelligent systems require some sort of internal meta-language, and that the language they use to communicate with users has to account for at least these two levels of symbolic manipulations.

The first and most typical difference between the language used in interactions with non-AI systems and that used with AI systems is precisely the need for a distinction between the two levels of operations AI systems should exhibit. Borrowing Jakobson's terminology [Jak60], we can say that communication with AI systems requires that the language being used be provided with a **metalinguistic** function, besides the referential and connative ones which characterize most linguistic codes used with non-AI systems.

Intelligent systems, unlike conventional ones, do not primarily operate on **data**, but on **concepts**. Data cannot be reasoned about, and users of conventional systems cannot have access to meanings, but only to results. AI systems, however, manipulate conceptual structures and this is their distinguishing feature. Therefore, users not only have access to (some sort of) meaning, but computations on meaning are actually the prime purpose of their interaction with such systems.

Thus, the second and most essential difference between the language people use to communicate with non-AI systems and with AI systems is that, in the latter case, language underlies the system's behavior itself. Since representation of concepts and manipulation of meaning structures is basically all AI systems do, such systems are themselves linguistic systems. So, not only is it the case that the code people use with intelligent systems must have additionally a metalinguistic function, but also - and more importantly - it must account for the fact that all actions performed by the system are linguistic. Therefore, this code must accommodate actions performed with and within language, which brings us to the idea **speech acts**.

SPEECH ACTS IN INTELLIGENT INTERFACES

The theory of speech acts has been originally proposed to account for the fact that some utterances in natural language are themselves **actions**. The limitations of meaning accounts in terms of formal semantic theories [Tar56] have been unveiled by the fact that **performative** utterances such as (a)

(a) I promise to help you memorize your script.

could not be adequately analysed in terms of truth values. In this case, the truth of the proposition **help you memorize your script** could not possibly be established at the time of the utterance, whereas the truth of **I promise** simply could not be challenged. All that could be said about this was that the speaker was manifesting his/her intention to make the state of affairs referred to by some proposition come true in the future [Bra87]. In other words, he/she was committing his/herself to bringing about that state of affairs. Other utterances such as (b)

(b) I declare the session open.

simply by being uttered in the appropriate situation and by the appropriate agent brought about the state of affairs referred to by the proposition(s) embedded in them. For these too, no truth value could be assigned in the traditional sense.

Austin [Aus62] proposed that utterances such as (a) and (b), and others, were **actions** in themselves. Searle [Sea69,75] revised and extended Austin's ideas and formed the basis of **speech act theory** as it is currently known. More recently, attempts have been made to unify both aspects of meaning in a single theoretical framework, ie. one in which formal semantics and speech act theory are conjoined to provide a more adequate account of human linguistic behavior [Van87].

One of the major difficulties encountered by speech act theorists is that of characterizing what Searle has called **indirect speech acts** [Sea75]. Austin's original proposal was essentially one in which speech acts and certain verbs or adverbs and phrases were mapped onto each other. This was clearly true for verbs such as **promise** and **declare** and **commissive** and **exercitive** speech acts, which Austin called illocutionary acts. However, as Searle pointed out in his criticism [Sea69], no one-to-one correspondence could be claimed to exist between illocutionary acts and illocutionary verbs. Austin's examples were just a particular case in the language. Not all speech acts could be claimed to be expressible by a particular verb or phrase, neither could a particular verb or phrase be claimed to express always one and only speech act. Indirect speech acts such as that in (c), below, illustrate the situation.

(c) Speaker 1: "I don't know what time it is now".
Speaker 2: "It's half past nine".

Despite the assertive shape of the utterance speaker 1 makes, it is clearly interpreted to be the case that he/she is making a request for speaker 2. In daily conversation, humans perform thousands of such indirect speech acts, which are all interpreted by their listeners according to principles of cooperativity [Gri75] and politeness [Lee83], among others. Therefore, the ideal situation suggested by early speech act research, in which some syntactic feature of the utterance could be expected to signal the speaker's perlocutionary goal proved to be only a particular case in the language, and not a general rule.

Modern Pragmatics has the goal of accounting for psychological and sociological components of meaning, which have been systematically left outside the scope of formal semantic approaches. Human interactions provide numerous challenges to pragmaticists, and reveal the limitations of many existing proposals [Lee83]. However, since our problem in this paper is not exactly the same as that faced by pragmaticists, our aim is to examine current "inadequate" accounts of human-human interactions and see if they can possibly improve accounts of human-computer interactions.

Our point is that the complexity of speech acts in human linguistic behavior is seemingly due to the complexity of human experience itself, which is rightfully assumed to be shared by all participants of whatever conversation there is between people. It doesn't seem to be the case, though, that humans have the same assumptions in conversing with systems. Experimental research [Ken88] shows that users manifest different linguistic behavior when interacting with what they think to be a system and whom they know to be a person, at a remote terminal. The interesting aspect of this research is that it shows that one of the things humans suppose computers are not equally able to do is to deal with **discourse** phenomena, something that is central to pragmatics. What consequences does this have to the applicability of current speech act postulates to human-computer interactions?

SEARLE'S TAXONOMY OF ILLOCUTIONARY ACTS IN CHI

One of the main criteria used by Searle to outline the set of possible direct speech acts is the **direction of fit** between word and world. When applied to the computer-human interaction environment some modification of his original proposal is called for. First, the distinction between the artificial **world** and its symbolic representation (the **word** level, say) is virtually none. As we argued before, intelligent systems are basically linguistic systems. Thus, their reality is only that of a language. To illustrate what we mean by this, let us focus on the distinction Searle makes between **declarative** and **assertive** acts. Declarative acts bring about a certain state of affairs in the world by simply being uttered. The direction of fit is dual and can be sketched as world <-> world. Assertive acts, however, manifest a certain existing state of affairs in the world and the direction of fit is world ->word. If we think of a knowledge-based system, when a user says (d)

(d) Birds normally fly.

to express some piece of knowledge the system is expected to acquire, although it is essentially an assertive speech act, in Searle's sense, the effect of such utterance on the system is the same as that of a declarative act. In other words, assertions made by users can modify the **world** of an intelligent system, since this world is only a symbolic representation of an "unreachable real world" only humans have access to. It could be argued, at this point, that intelligent systems can interpret declarative acts as distinct from assertive acts, by evaluating the truth of the latter against a fixed model of the world. Our reply to this is that, at the first (interface) level, the distinction is still not made, and that all that is done is to create one model and to compare this to another, at a second level of interpretation.

If we carry this reasoning a little bit further, directive acts also pose a problem in CHI environments. Searle's definition of directive illocutionary acts includes the specification of word -> world fit. In other words, the word expresses a state of affairs in the world that is not true at the present moment, but that is expected to come true as a result of some action the listener should perform after he/she interprets the utterance. Following the arguments we have presented above, we could say that all utterances of a user are interpreted by an intelligent system as a **command** for some action to be performed, since the mere **interpretation**, or **evaluation**, of these involves an attempt to bring about, at some symbolic level of representation, the state of affairs he/she has mentioned.

It all goes to show, we think, that intelligent systems are essentially linguistic processors, and therefore all actions performed by them are bound to involve speech acts. It also points at some sort of sollipsism [Fod80] on the part of the system, since there is no other world it can refer to but its internal representation of a global state of affairs. So, although CHI could be said to revolve about the idea of speech acts all the time, some re-assessment of what speech act types there are must necessarily be done.

Another aspect which must be analysed in the light of computer-human interfaces is the fact that there is no **natural language** understanding or generation in real terms. All intelligent systems that are claimed to process natural language actually process a pseudo-natural language, which is artificial in **all** possible ways. Psychological reality is something that is completely out of question in artificial intelligence as it is commonly viewed by computer scientists [Win77]. Therefore, there is an indisputable limitation in a system's ability to process natural language which relativizes claims of inadequacy in speech act theory when it fails to account for **all** types of actions humans can perform with language. The issue seems to come down to the following question. Since speech acts must be dealt with in computer-human interfaces, can the current theoretic proposals account for the pragmatic requirements of (pseudo) natural language processing systems? In other words, can speech acts be partly incorporated by artificial languages?

CAPTURING SPEECH ACTS IN COMPUTER-HUMAN INTERFACES

Adequate treatment of speech acts presents different challenges for understanding and generation of natural language utterances on the part of the system. Generation, as we have discussed elsewhere [S&S89], has as its crucial activity the selection of appropriate discourse structures to convey the desired speech act. Understanding, however, has the goal of recognizing the reason for a selection the user has made, based on hypotheses the system entertains about the user's perlocutionary aims throughout the whole history of the interaction. In sum, generation allows the system to have full control on **what is said**, but little control on **what is understood** by the user, whereas understanding allows for very little control on either. This is why, we will center the discussion of the present session on understanding, which seems to us to be more critical.

Because systems can only perform a limited set of intelligent actions, we can suppose that the collection of speech acts a system must deal with is similarly limited if compared with human. A crude example of the point we are trying to make is what Searle calls **expressive** speech acts. These are typically the reflection of the speaker's emotional attitude towards the topic of conversation. An appropriate response to them on the part of listeners must present some equally emotional quality in acting - something we cannot expect from artificial systems. Another type of speech act that does not make sense in CHI is the commissive act, such as that of a promise. The meaning of promises is deeply embedded in our social life, and artificial systems are not likely to share this part of human behavior [W&F86].

By examining the behavioral pattern of the interactions between users and knowledge-based systems [Gai88], we notice that users have a reason to be using the system, and a very precise one. It is not the case that user and system are in some sort of non-purposeful social interaction. The user wants to obtain from the system some response he/she believes the functionality of the system is able to handle. Therefore, **intentions and plans** are the very basis of computer-human exchanges. This is essentially what is conveyed in all utterances the user makes, and unless intentions and plans can be analyzed by the interface, all communication is at a risk [A&P80]. So, the issue now is how to incorporate speech act processing in CHI.

One of the first attempts to develop a natural language understanding system that incorporates the notion of speech acts as discussed above was Robinson's **DIAGRAM** [Rob82]. The author proposes that adequate speech act processing, in the light of what Allen and Perrault [A&P80] claimed intention and plan recognition should/could be, could be obtained by a classification of utterances into four different utterance types: declaratives, imperatives, polar (Y/N) interrogatives and WH- interrogatives. Furthermore, these types have been mapped onto surface linguistic structures, in the spirit of what Austin initially did with illocutionary verbs. Thus, declarative utterances were mapped onto declarative indicative sentences, and so on. The author, without explicitly saying so, subscribes the first and most disputable forms of speech act theory in pragmatics. But, should it be otherwise?

First of all, it has apparently been accepted by Robinson and others that natural language is implied when we speak of speech acts in CHI. It certainly seems natural that it should be so, since the very origins of speech acts are embedded in human use of language. However, two points should be made in this respect. One is that, if speech acts are to be redefined and adapted for use in CHI, it remains to be demonstrated that there is no artificial language that can be designed to account for this notion. The other is that, as we said before, there is no such thing as **natural language processing** in literal terms. All that is done in AI is to specify artificial languages that look very much like natural languages. So, in a way, these two points collapse into one an only big issue: how should pseudo-natural language incorporate the notion of speech acts in CHI?

Natural language is not universally believed to be the best way to communicate with intelligent systems. Sanford and Roach [San88] point out that, because real natural language processing is impossible in AI and only artificial codes can be designed to look like the natural means for human communication, it is very difficult for users to be aware of the differences between the two. Such differences should be clear, however, for adequate computer-human interaction. Otherwise, users naturally transfer to systems all of their own cognitive capacities. By doing so, they introduce a high potential for breakdown [W&F86] in communication, since the assumptions about a shared cognitive and social background between agents simply do not apply.

It is not the case, though, that **all** utterances are equally prone to causing communicative breakdown. There is a subset of utterances which inequivocally convey the desired meaning in the situation they are uttered. The problem is, then, one of controlling the user so that he/she does not go beyond the limits of automatic interpretability [San88]. Free NL dialogue with machines, one in which users can formulate their questions or comments in any way they want to, is certainly not easy to control. The menu-driven natural language interaction of the type proposed by Tennant

[Ten84] is a solution to this. However, the NLMenu system he has developed does not tackle the speech act problem explicitly, neither does it approach discourse phenomena in a theoretically motivated way.

We think of an extension to Tennant's approach. The menu-driven natural language interpretive component of the interface seems to provide a good alternative to the control problem Sanford and Roach have discussed. The user is, then, quite aware of the artificiality of the language he/she is using and simply cannot go beyond the interpretive abilities of the processor, since the opportunity for forming unanticipated utterances will never arise. The grammar we believe should be driving the presentation of menus, however, unlike NLMenu's grammar, does not concentrate only on sentential syntax and semantics, but rather on general discourse structure in which relevant pragmatic phenomena can be explicitly reasoned about. Syntactic and semantic levels of analysis are, thus, bound to overall communicative bahavior and allow for a much more intelligent dialogue between users and system.

Although Kennedy et al. [Ken88] have shown that discourse phenomena such as anaphora are less frequent in users' utterances when they communicate with what they think to be a system than when they communicate with a person via terminal, users have not given signs of **non-pragmatic** linguistic behavior. Quite contrarily, there are numerous reports in current natural language processing literature that indicate the need for a pragmatic component in an efficient processor.

We believe that the processing of speech acts, just as that of anaphoric references or rhetorical structures, must be part of a NL front-end's linguistic capacity. Therefore, the new challenge for menu-driven NL interfaces is how to incorporate an adequate pragmatic knowledge to account for these aspects of human language use. In that which concerns speech acts, specifically, we believe Robinson's proposal can be taken as the basis for the system's pragmatic competence. However, the treatment of speech acts does not have to be projected onto the syntactic level only, as part of the general grammar. It can be dealt with at a separate level, and reasoned about independently, even if we subscribe, in the end, to the correspondence between speech acts and illocutionary verbs or phrases.

CONCLUDING REMARKS

In this paper we have discussed relevant design issues we think must be taken into account in computer-human interface systems. Essentially, the baseline of our argument is that current intelligent systems are **linguistic systems** and all actions performed by them or their users are **linguistic actions**, or **speech acts**. Pragmatics provides intelligent systems interface designers with a variety of proposals. There are many controversial and/or unresolved issues involving human speech acts, but not many attempts have been made to investigate if current theoretic results apply to computer-human interactions and how. We emphasize that such an investigation necessarily involves a characterization of artificial rational agency and consequently a re-definition of speech acts themselves in CHI.

One of the points we have explored is the fact that early illocutionary verb-based theories, although inadequate in human communication environments, can be particularly helpful in computational environments, since the artificiality of the language being used there is inescapable and control over it highly desirable. The rigidity and unnaturalness of menu-driven NL interfaces, for example, should not be compared to human natural interaction in social life (subscribing to all current pragmatic debates), but rather to the rigidity and limited expressiveness of current programming languages (subscribing only to part of the debates, and reviewing basic concepts and taxonomies).

The idea of extending current menu-driven approaches to natural language understanding, so that pragmatic phenomena can be adequately dealt with, indicates some new avenues for research. One is that of the internal architecture of the front-end component. Should the analysis proceed sequentially or should it be done in an integrated way? Should pragmatic features be independently reasoned about or should they be considered together with semantic and syntactic features? What subset of discourse phenomena should be incorporated by the interface grammars?

There is much experimental research to be done, so that the current interactive pattern between users and intelligent systems can be clearly characterized. There is also much theoretic research to be done, specially in terms of defining what artificial rational agency is. Unless we can have such knowledge available to us, natural language processing investigations that center around grammatical aspects of a system's competence **per** se (and to a certain extent those that center around some semantic aspects as well) are bound to be brittle.

The next step in our personal investigation in the area is to specify a menu-driven natural language front-end to a knowledge-based system. We will use input from speech act theory as it is proposed in he field of pragmatics and provide a first approximation of some **CHI Speech Act Theory** based on experimental results reported in the natural language processing literature. The generation component of the interface is currently being developed using Rhetorical Structure Theory [M&T86] to account for pragmatic phenomena and assumes the interpretation component will handle users' intentions and plans in an adequate way. In both components, a computational model is assumed for which psychological reality is not critical. Instead, what we are looking for is optimal interpretation of utterances by the system (in analysis) and by the user (in synthesis), so that real communication can take place.

REFERENCES

[A&P80] Allen,J.F. and Perrault,C.R. (1980) **Analyzing Intention in Utterances.** Artificial Intelligence 15 (1980) 143-178.

[Aus62] Austin,J.L. (1962) **How to do Things with Words.** Cambridge, Mass., Harvard University Press.

[Bra87] Bratman,M.E. (1987) **What is Intention?** in P.R.Cohen and C.R.Perrault (eds) Formal Theories of Communication. CSLI Report no. CSLI-87-69. 3-22.

[C&L87] Cohen,P.R. and Levesque,H.J. (1987) **Persistence, Intention and Commitment.** in P.R.Cohen and C.R.Perrault (eds) Formal Theories of Communication. CSLI Report no. CSLI-87-69. 21-54.

[Dre72] Dreyfus,H. (1972) **What Computers Can't Do: A Critique of Artificial Reason.** New York, Harper and Row.

[Fod80] Fodor,J.A. (1980) **Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology.** Behavioral and Brain Sciences 3.1, 63-73.

[Gai88] Gaines,B.R. (1988) **A Conceptual Framework for Person-Computer Interaction in Complex Systems.** IEEE Transactions on Systems, Man and Cybernetics 18(4), July/August 1988. 532-541.

[Gri75] Grice,H.P. (1975) **Logic and Conversation.** in Cole,P. and Morgan,J.L. (eds) Syntax and Semantics: Speech Acts (vol.3). New York, Academic Press.

[Jak60] Jakobson,R. (1960) Linguistics and Poetics. in T.Sebeok (ed) Style in Language, Cambridge, Mass., MIT Press, 350-377.

[Kam88] Kammersgaard,J. (1988) **Four Different Perspectives on Computer-Human Interactions**, 28, 343-362.

[Ken88] Kennedy,A.,Wilkes,A.,Elder,L.,Murray,W.S. (1988) **Dialogue with Machines.** Cognition, 30 (1988), 37-72.

[Lee83] Leech,G. (1983) **The Principles of Pragmatics.** London, Longman.

[M&T87] Man,W.C. and Thompson,S.A. (1987) **Rhetorical Structure Theory: A Theory of Text Organization.** in L.Polanyi (ed) The Structure of Discourse, Norwood NJ. Ablex.

[Rob82] Robinson,J.J. (1982) **DIAGRAM: A Grammar for Dialogues.** Coomunications of the ACM, 25 (1), 27-47.

[S&R88] Sanford,D.L. and Roach,J.W. (1988) **A Theory of Dialogue Structures to Help Manage Human-Computer Interaction.** IEEE Transactions on Systems, Man and Cybernetics 18(4), July/August 1988. 567-574.

[S&S89] Scott,D.R. and Souza,C.S. (1989) **Getting the Message Across: a Computational Grammar of Text.** To Appear. Also available as a Philips Research Labs. Technical Report, no. 2885.

[Sea69] Searle,J.R. (1969) **Speech Acts: An Essay in the Philosophy of Language.** Cambridge, Cambridge University Press.

[Sea75] Searle,J.R. (1975) **A Taxonomy of Illocutionary Acts.** in Cole,P. and Morgan,J.L. (eds) Syntax and Semantics: Speech Acts (vol.3). New York, Academic Press.

[Tar56] Tarski,A. (1956) **Logic, Semantics and Metamathematics.** Oxford University Press.

[Ten84] Tennant,H. (1984) **Menu-Based Natural Language Understanding.** in AFIPS Proceedings, vol 53. 1984.

[Van87] Vanderveken,D. (1987) **On the Unification of Speech Act Theory and Formal Semantics.** in P.R.Cohen and C.R.Perrault (eds) Formal Theories of Communication. CSLI Report no. CSLI-87-69. 129-170.

[Win77] Winston,P.H. (1977) **Artificial Intelligence.** Reading, Mass. Addison-Wesley (2nd Edition 1884)

[W&F86] Winograd,T. and Flores,F. (1986) **Understanding Computers and Cognition: A New Foundation for Design.** Norwood,NJ, Ablex.