Indirect Speech Acts and Collaborativeness in Human-Machine Dialogue Systems

Frédéric Landragin

DRAFT

Abstract: In human-machine dialogue systems, indirect and composite speech acts have to be treated in a proper way. First because they appears frequently in human-human dialogue, and then constitute an important aspect of spontaneous communication. Second because they are linked to collaborative aspects. We describe some complex speech acts phenomena and some methods for a system to treat them, with the help of hypotheses on user's mental states.

1. Introduction

A lot of semantic and pragmatic research works (Searle, 1975; Perrault & Allen, 1980, etc.) deal with complex speech acts, particularly indirect speech acts—meaning something instead of something else, for instance the question "can you give me the salt?" instead of the request "please give me the salt"—, and composite speech acts—meaning several things simultaneously, for instance "how long does it take to go by this way which seems to be the shortest?" where a comment ("this way seems to be the shortest") is added to the question ("how long does it take to go by this way?"). Some other research works (Grice, 1975; Quignard, 2002, etc.) deal with collaborative dialogue and argumentation in dialogue, and are not always linked to speech acts theories.

On the other hand, there are only a few computer science works in the area of human-machine dialogue that exploit relevantly the previous research works in order to build on systems with high understanding abilities. Only a very few real systems are able to manage indirect and composite speech acts, and most of them (see for instance the dialogue systems that were designed during the MIAMM and OZONE European projects, http://www.miamm.org and http://www.hitech-projects.com/euprojects/ozone) rely on simple predefined rules that transform pre-identified types of utterances into types of speech acts.

Face to these discrepancies between theories and implemented systems, some key questions arise on the way to design more efficient dialogue systems:

- Does a human-machine dialogue system need to identify the surface and profound speech acts (for an indirect speech act), and the several speech acts (for a composite one)? Among these identified speech acts, to which must the system react? All of them or only a privileged or optimal one? What are the possible criteria that allow the system to choose between them?
- When interpreting a complex speech act, does the system need to make hypotheses on the speaker's mental states (beliefs, desires, intentions)? Does the system need to manage internal structures reflecting the dialogue state, for instance, following the terms from (Portner, 2004), the 'common ground', the 'questions stack', and the 'to-do lists'? How can all these aspects be confronted and managed together? What are the minimal requirements, i.e., the simplest internal structures for interpreting correctly the user's messages?
- What are the links between complex speech acts processing and collaborativeness in dialogue? When interpreting complex speech acts, does a system increase the dialogue

collaborative aspects? Which collaborative characteristics allow the system to resolve indirect speech acts?

In this paper we will not be able to answer to all of these questions. But, keeping them in mind, we want to clarify some aspects of the design of human-like understanding systems. In the second section we present a quick state of the art for research works dealing with speech acts. In the third section we address the problems that appear when applying theories to the implementation of systems. Then, we propose in the fourth section the foundations for a computational model to process complex speech acts. Some principles are drawn and illustrated, and we conclude with the future works to increase the communicative abilities of dialogue systems.

2. Complex speech acts

Two main manners to apprehend speech acts can be distinguished. Speech acts can first be viewed as semantic units. Types of sentences and utterances are then identified using semantic criteria. (Hamblin, 1987) is an example of such an approach, with the case of imperatives. But semantics is not sufficient for explaining everything, and some pragmatic aspects such as basic conversational principles—the maxims of (Grice, 1975), for instance—are required. Some semanticists try to integrate these aspects into semantic factors, but the resulting theories are not really convincing. In fact, the use of language especially in dialogue situations seems impossible to be included into the semantic content. Second, speech acts can be viewed as pragmatic units. Following Searle (and, for instance, Vanderveken), speech acts intervene at a different level than semantic content. An illocutionary force F is added to a propositional content P: "F (P)", see (Searle, 1979). This force is not linked to any semantic parameter, that can be seen as a lack. A link can sometimes be useful, for instance between the semantics and the illocutionary force of an expressive attributive adjective such as "damn" ("I have to mow the damn lawn", see Potts, 2003).

More precisely about indirect speech acts, the semantic point of view is well illustrated by (Hamblin, 1987). Proper imperatives (commands, requests, demands, advices), whishes, permissives, and undertakings correspond to speech acts types. There are not linked to particular forms of utterances, but all of them have semantic properties, such as the presence of a conditional clause for permissives. In theory, the speech act type can be identified when analyzing the semantic characteristics of the utterance. But these characteristics are not sufficient and Hamblin emphasizes the importance of pragmatic factors. Concerning the pragmatic point of view, the main fundamental work is the one of (Searle, 1969) and (Searle, 1975). Following his idea of an illocutionary force added to a propositional content, Searle propose five main categories of speech acts: assertives, directives, commissives, expressives, and declarations. For each of them, the formula F (P) can be completed with more precisions on the adjustment direction (from the word to the world, or from the world to the word, or both) and on the sincerity condition (B = belief, D = desire, I = intention). If the adjustment direction has no interest for computational pragmatics and human-machine dialogue (Quignard, 2002), on the other hand taking the user's mental states (B, D, I) into account clarifies a lot the differences between the categories of speech acts (Searle, 1976):

- assertives = B (P),
- directives = D (listener does Q),
- commissives = I (speaker does Q),
- expressives = ∅ (?) (speaker/listener + property),
- declarations = ∅ (P).

For the treatment of indirect speech acts—so, for the identification of the correct speech act category—, (Searle, 1975) proposed a ten-steps method consisting of a chain of inferences. Among these inferences, one is related to the speaker's mental states. With the example of "can you give me the salt?", this inference corresponds to "the speaker probably knows that the answer is yes, so his utterance is perhaps something else than a question". Even if Searle considered that this inference is not essential, it seems to be of importance for the resolution of indirect speech acts. The role of mental states (four in the Searle's paper: belief, desire, intention, pleasure) is here emphasized for indirect speech acts processing. Searle exploited these mental states to provide a list of indirect directives due to some conventional principles.

Pragmatic aspects are then privileged, and more recent works such as (Potts, 2003) and (Portner, 2004) emphasize their importance. In his Ph.D. dissertation, Potts describes in detail conventional implicatures of two sorts: supplements and expressives. Parentheticals, appositives, or relatives like the "who"-relative in "I spent part of every summer until I was ten with my grandmother, *who lived in a working-class suburb of Boston*", are examples of supplementary expressions, as well as speaker-oriented adverbs such as "amazingly" in "After first agreeing to lend me a modem to test, Motorola changed its mind and said that, *amazingly*, it had none to spare" (Potts, 2003). Epithets, honorifics, or expressive attributive adjectives are examples of expressive expressions. These categories can be considered as typical cases of composite speech acts.

What about collaborative aspects? Are they included into pragmatics aspects, or do they intervene at a different level, illustrated by the following formula: "C (F (P))"? Following (Quignard, 2002) and other works dealing with collaboration and argumentation, the collaborative aspects consist of dialectic functions that intervene at another level than the utterance understanding. Speech acts are linked to an informative and communicative level, but dialectic functions are linked to an evaluative level. Nevertheless, we can consider that the use of an indirect speech act by the speaker is a way for him to be less incisive and more polite and more collaborative. As an illustration of the argumentative possibilities, the 'C' in the previous formula can be anyone of the dialectic functions proposed by (Quignard, 2002), for instance 'ARG-PRO-MT' (the speaker provides an argument in favor of his own thesis) or 'REQ-TDP-MT' (the speaker asks his opponent to take a position with respect to his thesis).

3. Speech acts processing in dialogue systems

Human-machine natural language dialogue systems cover an increasing number of phenomena, and exploit more and more the results from linguistic and pragmatic researches. For a long time there existed two main categories of dialogue systems, first the 'command systems' where the user uttered only a simple chain of orders, and second the 'information systems' where the user could only ask questions to the system, as he did with classical database systems using an artificial query language such as SQL. Then, there was no speech act processing. In fact, each dialogue system was designed for the treatment of one speech act (command or question). Whatever the form of an utterance (assertion, question, or order), its illocutionary force was systematically brought back to the expected one. It was the case for most of industrial systems, and also for research prototypes such as the multimodal dialogue system from the MIAMM European project.

With the objective of more natural and less unilateral dialogues, there is a need for different treatments for the various illocutionary forces, and then for a module dedicated to the identification of the most probable illocutionary force of each speaker's utterance. The

simplest way to proceed is to specify during the system design phase a 'hashtable' linking each possible utterance form to its supposed force. But all possibilities have to be thought by advance, and it is difficult to make the system evolve. The main advantage of this method is to avoid the implementation of a complex algorithm for indirect speech acts resolution. That is why it is used for the design of a lot of dialogue systems, such as the INRIA dialogue system demonstrator of OZONE European project.

Concerning the elaboration of a module for indirect speech acts understanding, a lot of computational models have been proposed. Following the idea of conversational postulates from (Gordon & Lakoff, 1975), the hashtable previously mentioned can be replaced by a set of rules that allow the system to identify the profound speech act from the surface one. This method is more flexible, but it presupposes that it is always possible to extract from one utterance form one profound speech act. The problem is that the same utterance may have both interpretations (the surface act interpretation and the profound act one), i.e., some rules must not apply in some situations. A recent work (Xuereb & Caelen, 2004) introduces statistics in order to identify the most probable profound speech act in a dynamic and flexible manner. As a last example of dialogue system, the one from France Telecom R&D (Sadek *et al.*, 1997) follows Searle's theory by implementing an automatic identification of some mental states of the speaker. The system is then able to make inferences exploiting the speaker's beliefs and intentions. One problem is here the complexity of the algorithm, which needs to manage logical forms related to propositional contents and to mental states. Imagine for instance something like "Believes (Speaker, (Knows (System, P)))".

4. Managing mental states and dialogue structures to interpret complex speech acts

4.1. Segmenting speech acts

How can we design a computational model for speech acts processing? The first point to address is the segmentation of speech acts. Even if an utterance from the user corresponds to one grammatical sentence, several speech acts can be identified. This is the case with examples like "how long does it take to go by this way which seems to be the shortest?". In fact, it depends on the way we consider discourse structure and relations between discourse segments. With the previous example, a 'comment' relation can be identified between the two following discourse segments: "how long does it take to go by this way" and "which seems to be the shortest". Then, following a theory considering both discourse structure and speech acts, such as (Asher & Lascarides, 2003), this will lead to have two separated speech acts. To the contrary, following an approach where one utterance corresponds to one discourse segment (due to significant acoustic blanks before and after the utterance), this will lead to have one composite speech act. Thus, there is no immediate answer to the segmentation problem. Since composite speech acts might appear even with discourse structure considerations, we can follow the acoustic-based approach and consider that one speech act (simple or composite) is attributed to each dialogue turn.

4.2. Interpreting indirect speech acts

What are the prerequisites for a computational model of indirect speech acts processing? Considering that the profound speech act may be at the origin of the answer content, and that the surface act may be exploited for the answer form, the system must identify both of them. Considering that the same form may lead to different interpretations (due to contextual

factors) in terms of profound acts (Asher & Lascarides, 2001), the identification may rely on the following factors and resources:

1. *Linguistic and semantic characteristics of the utterance* (following Hamblin). For instance: imperative, interrogative, or indicative? what is the semantic category of the verb?
2. *The dialogue history*, i.e., the previous utterances and their interpretations represented with logical forms. A particular linguistic form can be used frequently by the user with a particular aim that implies the use of an indirect speech act (we can imagine machine-learning techniques for the management of such a phenomenon).
3. *A lexicon of dialogue pairs*, with the associated profound speech acts and possible reactions. For instance, the system may know that a proposition has to be answered to by an acceptation, a reject, or a counterproposition.
4. *Classical conventional uses* and associated set expressions. This is typically the case for "can you give me the salt?" and similar constructions. NB: This item can be seen as a part of the previous one.
5. *The list of the system abilities and all task constraints*. For instance, if the dialogue system helps the user finding a restaurant, a question like "can you list me the Chinese restaurants near Palaiseau" is of course a request.
6. *Hypotheses on the speaker's mental states*. For instance, the hypothesis that he already knows the answer to his question. That was the case for the salt, but also for a lot of less conventional situations: when the user asks the system "can you open this file?", he may know (or believe) that the system is able to open the file.

Then, a computational model for indirect speech acts processing may take these parameters into account. We can define a priority order that corresponds to the order used for the previous list. One point in this list is the importance of the dialogue history. The system reaction is based on the nature of the user's utterance as well as on the current state of the dialogue. The notions of 'common ground' (CG), 'questions set' or 'questions under discussion' (QUD), and 'to-do lists' (TDL) are here of importance. Following (Portner, 2004), to each of these three notions corresponds a stack of propositions. An assertion from the user increments the CG with the corresponding proposition, a question increments the QUD with the corresponding propositional function, and an order increments the TDL. CG, QUD and TDL are built on during the dialogue, right after each semantic and pragmatic analysis. Following its illocutionary force, an utterance is saved in the right stack with a logical form corresponding to the result of its semantic analysis. Taken together, the three stacks constitute the major part of the dialogue history. In task-oriented human-machine dialogues, the task constraints may be pregnant so that a 'default' TDL can be imagined.

When producing a message reacting to an indirect speech act, the system first takes into account the profound act in order to determine an answering content in 'coherence' with the task and the dialogue history. Then, the system has to choose between ignoring the surface act, or taking it into account for the linguistic form of the answer and its 'cohesion' within the dialogue. The only parameter for making this choice seems to be the maintenance of a certain linguistic cohesion. As an example, consider the classical example "do you have time?". The form is a question but the speaker will of course not be satisfied with a "yes" response. If she has a watch, the hearer can react with "five o'clock" or "yes, five o'clock". Including a "yes", the second answer has a better cohesion with the speaker's utterance than the first one. Then the system will favor this second answer.

4.3. Interpreting composite speech acts

Concerning composite speech acts, the system has to identify the primary act and the secondary one(s), and to classify them using a salience hierarchy. The identification factors are the same than for indirect speech acts. Moreover, the classification of the different acts relies on the same parameters than their identification:

1. *Linguistic and semantic characteristics of the utterance*. In particular: epithets, evaluative adverbs, appositions, subordinate clauses, etc. (following Potts). In "how long does it take to go by this way which seems to be the shortest?", the subordinate clause constitutes a criterion for identifying a secondary act. NB: As we will see with the other items of this list, the parameters categories for indirect speech acts processing and composite speech acts processing are the same, but the criteria that are exploited are not.
2. *The dialogue history*. When the same composite act is produced again by the speaker, the same classification has to be made by the system if the first one was a success (we can also imagine here machine-learning techniques, not only for the identification of the potential primary act, but also for the determination of the speech act category—simple or composite).
3. *A lexicon of dialogue pairs*, with the associated primary and secondary speech acts and all possible reactions. For instance, the system may know that the association of a question and a comment has to be answered to by a response to the question, a reaction to the comment (confirming, infirming), or both.
4. *Classical conventional uses* and associated set expressions. This is for instance the case for "who *the hell* did that?" and similar constructions. In French, this is above all the case for utterances like "qui a *bien* pu faire ça ?", where the presence of "bien"—a very used adverb with various significations—leads in this particular construction to the identification of a composite speech act.
5. *Task constraints*. When determining the primary act, the more relevant to the task the act is, the better it is classified.
6. *Hypotheses on the speaker's mental states*. When determining the primary act, the best hypothesis is the one that has the most important contextual effects to the mental states, see (Sperber & Wilson, 1995). With the example "how long does it take to go by this way which seems to be the shortest?", there are several possibilities. First possibility: the comment is true, i.e., the way the user is pointing out is really the shortest one. Then, confirming this belief will have a limited effect on the speaker's mental states, whereas uttering the response to the question will have a greater effect (adding a new knowledge to the user's mind). Second possibility: the comment is false. Then, infirming this belief will have an important effect on the speaker's mental states. The problem here is that it is difficult to compare this effect to the effect corresponding to the new knowledge. Thus, in this case, the system might react to both acts.

A computational model for composite speech acts processing may take these parameters into account, in the same order than presented in the list. The result of this process is an ordered list of speech acts, beginning with the primary act that was identified.

When determining the system reaction to a composite speech act, only this primary act may be answered to. Comments can be added for secondary acts. It depends on the level of collaboration that is expected from the system. For instance, with "how long does it take with

this travel, which seems to be the shortest?", the most salient act is the one of the main proposition (linguistic factor), and the secondary act is the comment. A simple answer from the system can be "this travel will take you 20 minutes" (reaction to the only primary act, corresponding to the expected reaction). A more collaborative answer can be "this travel will take you 20 minutes, and is actually the shortest" (reaction to all acts). If the comment were computed as little more important than the question, the answer should have been "you're right, this travel is actually the shortest, and it will take you 20 minutes". And if the comment were considered as the primary act, the question should have been ignored. More precisely, this case can appear when the comment is false. In this case, the system may consider that correcting the comment is more important than answering the question. Thus, the answer may be: "this is not the shortest travel, the shortest is that one".

5. Conclusion and future work

With the aim to design more collaborative and more natural speech-based dialogue systems, indirect and composite speech acts must be taken into account. The most relevant speech act has then to be identified, and this is an important issue for systems with deep understanding abilities. Criteria such as salience or relevance may be exploited during this identification process. Moreover, some of the user's mental states, and particularly intentions, must also be taken into account. The intention behind an utterance appears to be the main parameter for determining an adequate reaction or answer to this utterance. When this intention is pregnant, the form of the message has sometimes no importance. Dialogue systems must also be able to identify what the user already knows and what he is susceptible to want to know. The purpose is to never repeat what is already known and to focus the dialogue on what might be known by the user. Giving such a capacity to dialogue systems constitutes for the most part a future work.

Other future works can be identified concerning the determination and the exploitation of stronger links between complex speech acts and collaborativeness. When he is not familiar with the task, the user often produces utterances where several leads are left opened. This is the case with "I want to go to Paris, no problem?" or "if possible, I would like to buy a train ticket", where the system may choose between ignoring and reacting to the question and the "if possible" (since they belong to phatic aspects of oral communication). Strict assertions and orders are not frequent in spontaneous communication. They are often accompanied with phatic questions or expressions. Several questions can also be produced together inside one long utterance with a quick rhythm, like "can I have a taxi—uh is it possible?—to go to Palaiseau? uh is it here that I can asked for a taxi? is it possible?". Face to such an utterance, a collaborative system may be able to identify the main questioning of the user and to answer to it (and to calm him). That can be done with answers like "yes, I'm going to call the taxi company" or "I cannot do that but you can ask to my colleague". With such a spontaneous example, it really seems that collaborative behaviours rely on a fine treatment of complex speech acts.

References

Asher, N. & Lascarides, A. (2001). Indirect Speech Acts. *Synthese*, 128 (1-2).
Asher, N. & Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.
Ginzburg, J. & Fernández, R. (2005). Conversational Acts and Non Sentential Utterances in Multilogue. In: *Sixth International Workshop on Computational Semantics*, Tilburg.
Gordon, D. & Lakoff, G. (1975). Conversational Postulates. In: P. Cole & J. Morgan (eds), *Syntax and Semantics, Vol. 3: Speech Acts*. Academic Press.

Grice, H.P. (1975). Logic and Conversation. In: P. Cole & J. Morgan (eds), *Syntax and Semantics, Vol. 3: Speech Acts*. Academic Press.
Hamblin, C.L. (1987). *Imperatives*. Blackwell.
Landragin, F. (2004). Dialogue History Modelling for Multimodal Human-Computer Interaction. In: *Eighth Workshop on the Semantics and Pragmatics of Dialogue (CATALOG'04)*.
Perrault, C. & Allen, J. (1980). A Plan-Based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics*, 6 (3-4).
Portner, P. (2004). The Semantics of Imperatives within a Theory of Clause Types. In: K. Watanabe & R.B. Young (eds), *Proceedings of Semantics and Linguistic Theory 14*. CLC Publications.
Potts, C. (2003). *The Logic of Conventional Implicatures*. PhD dissertation. University of California.
Quignard, M. (2002). A Collaborative Model of Argumentation in Dyadic Problem-Solving Interactions. In: F. van Eemeren, J.A. Blair & C.A. Willard (eds), *Proceedings of the Fifth International Conference of the International Society for the Study of Argumentation (ISSA'02)*. Amsterdam.
Sadek, D., Bretier, P. & Panaget, F. (1997). Artimis: Natural Dialogue Meets Rational Agency. *IJCAI*.
Searle, J. (1969). *Speech Acts*. Cambridge University Press.
Searle, J. (1975). Indirect Speech Acts. In: P. Cole & J. Morgan (eds), *Syntax and Semantics, Vol. 3: Speech Acts*. Academic Press.
Searle, J. (1976). A Taxonomy of Illocutionary Acts. In: Gunderson (ed), *Language, Mind, and Knowledge*. University of Minnesota Press.
Sperber, D. & Wilson, D. (1995). *Relevance. Communication and Cognition*. Blackwell.
Xuereb, A. & Caelen, J. (2004). L'interprétation pragmatique en dialogue homme-machine finalisé. *Journées scientifiques Sémantique et Modélisation*, Lyon.