

Detecting and exploiting user familiarity in natural language human-computer dialogue

Ludovic Le Bigot,* Philippe Bretier** & Patrice Terrier***

* *University of Poitiers & CNRS (CeRCA, UMR 6234)*

** *Orange Labs Research & Development*

*** *University of Toulouse II & CNRS (CLLE-LTC, UMR, 5263)*

France

1. Human-computer dialogue and usability

1.1 What is human-computer dialogue?

The fields of application for human-computer dialogue are both many and various. Users may dialogue with customer services over the phone or via a personal digital assistant in order to carry out banking transactions, find their way around a city, or book train or plane tickets (Wilkie et al., 2005).

The modality of interaction between user and system can also be either partially or totally speech-based. System messages or prompts can be delivered to the user via either a screen (text, graphics) or an auditory modality (pre-recorded or computer-generated voice). User data entry is possible through speech or via either a phone or a computer keyboard, and in the case of advanced systems, users can submit their requests in natural language (see Gorin et al., 1997, Le Bigot et al., 2007).

Furthermore, current systems propose a wide range of interaction styles. The dialogue can be guided by the system (system-initiated), the user (user-initiated) or both (mixed-initiated) (see Lai et al., 2008, for details). In system-initiated dialogue, which is the most common, the interaction is led by the system and may appear mechanistic. In the following examples, the system either questions the user or offers him or her a choice. For example, the system may say, *"To buy the share, press 1, and to sell the share, press 2"*. The user then presses 1 or 2 on his or her phone keypad. For a speech system with explicit keywords, the *"Press 1"* or *"Press 2"* is replaced by a verb or a common noun: *"To buy the share, say Buy, and to sell the share, say Sell"*. The user says the keyword aloud and continues the interaction. For a speech-based system with implicit keywords, the choices are suggested. The system may ask the user, *"Do you want to buy or sell the share?"*. The user must then say the *"buy"* or *"sell"* verb aloud to continue the interaction (a system with natural language capability would be able to recognise a response such as, *"I want to buy the share"*).

User-initiated systems are few and far between. However, they offer users greater freedom and flexibility. Once they have asked the user what he or she wants (*"What do you want?"*), the interaction continues in an open mode. Some systems offer a functionality which combines several different interaction modes. For example, the interaction may begin on a

user-initiated basis, but if several dialogue errors occur, the system may shift to a system-initiated mode.

Note that the degree of initiative displayed by the system to the user may not reflect the full extent of its understanding capabilities. In the case of the speech-based system-initiated example mentioned earlier, the system may accept greater flexibility from the user than it appears to. In addition to the keywords mentioned by the system, all the following user inputs may also be accepted as requests to buy a share: "I want to buy a share", "Buy a share", or "The first choice". This chapter, however, focuses on the degree of initiative displayed by the system to the user.

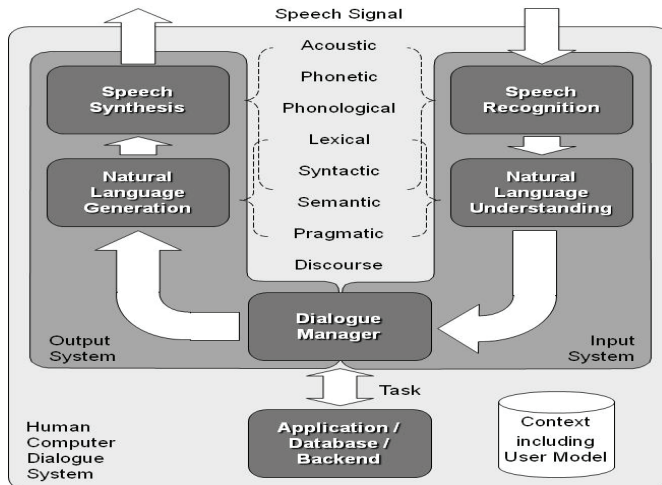


Fig. 1. General architecture of a human-computer dialogue system

As shown in Fig. 1, the architecture of a (speech) human-computer dialogue system is basically composed of a chain of five modules: *speech recognition*, *natural language understanding*, *dialogue manager*, *natural language generation* and *speech synthesis*. An additional module, linked to the dialogue manager, encapsulates the functional part of the system, which may take the form of a *computer application*, a *database* or some more complex *backend system*.

The module chain can be split in two subchains: the *input* and *output systems*. These two systems pass through different levels of analysis. The input system leads from the speech signal to the discourse level, passing through the acoustic, phonetic, phonological, lexical, syntactic, semantic and pragmatic levels. *Speech Recognition* is the first module in the subchain. Its role is to transcribe the user's speech signal into text, drawing on models at each level, from acoustic to syntactic. *Natural Language Understanding* is the second module. This processes text from the lexical and syntactic levels, converting it into dialogue action at the discourse level. The *Dialogue Manager* deals with user dialogue actions and reacts to them by implementing system dialogue actions. The output system moves in the opposite direction. Starting from the system's communicative actions at the discourse level, the *Natural Language Generation* module transcribes these into text at the syntactic and lexical

levels. After that, the *Speech Synthesis* module produces the speech signal, drawing on models at each level, from phonological to acoustic. During a dialogue between a user and the human-computer dialogue system, the processing follows the arrows displayed in the Fig. 1, in a pattern of user speech input processing followed by system speech output processing. This is the pattern that defines a *dialogue turn*, or *speaking turn*, for each dialogue partner (see Pietquin, 2004, for a detailed description of the general architecture of the human-computer dialogue system).

In the course of the speaking turns, all the modules retain a certain amount of data. These data are stored and made available to all the modules in a shared *context knowledge base*. This knowledge base contains the dialogue history at each level of analysis (e.g., speech duration and level, word occurrences and phonetics, utterances, dialogue actions, dialogue and task progress), together with knowledge about the user (e.g., age, gender, familiarity, average word count, average response time, language) and the environment (e.g., noise level) collected during the dialogue. This knowledge is used by the modules when they process each dialogue turn. The part of the knowledge base relating to the user is called the *user model* (see Part 2). The output and input systems must be designed homogeneously, primarily in accordance with the style of interaction. In the case of system-initiated interaction, the system asks the user to select an item (e.g., a keyword) from a list (see above example). The input system then has to process the user's speech input on the basis of the previous system's output. For obvious reasons, the system's understanding capabilities have to be at least on a par with the openness it displays to the user. Whatever the qualities of dialogue system's various components, its design must obey certain rules if it is to be easily understood by its users.

1.2 Usability of human-computer dialogue systems

Most of the recommendations for designing and assessing usable systems are also applicable to advanced human-computer dialogue systems (totally or partially speech-based). Usability is the key to designing a Web interface or software that is easy to use. The ISO standard 9241-11: 1998 defines usability as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*". This notion is intended to help designers produce user-friendly systems. Many guidelines or heuristics have been put forward to ensure the usability of interactive systems. Nielsen (1994), for instance, recommended ten main heuristics for designing/evaluating a system. These heuristics range from giving the user clues about the system to providing assistance and documentation. Some of these heuristics are easy to apply to dialogue systems, and one example of Nielsen's heuristics (for complete list, see http://www.useit.com/papers/heuristic/heuristic_list.html) applied to spoken dialogue systems is set out in the following paragraph.

Nielsen's usability principle: *Flexibility and efficiency of use – Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater for both inexperienced and experienced users. Allow users to tailor frequent actions.* For example, this opportunity may be offered in the message introducing a keyword-based interactive voice response system, by informing the user that a service can be directly accessed by activating a keyword: "*If you know [...] say the keyword corresponding to your request [...]*".

This example of the rule that has to be obeyed raise the question of whether the user should be taken into account as early as the design phase in order to enhance the usability of a

human-computer dialogue system; given that one of the keys to ensuring that human-computer dialogue systems are tailored to their users is to model the latter's behaviour or environment in order to create the optimum conditions for interaction.

2. Why model users in human-computer dialogue?

2.1 Static and dynamic user models

Due to significant variations in human-computer dialogue system performances according to the type of user and the type of dialogue, as is the case with more general interactive systems such as human-computer interaction, user modelling has become a critical aspect of designing systems and services (Litman & Pan, 2002). Like the knowledge pertaining to users that is contained in the context knowledge bases of human-computer dialogue systems, user models are constructed from both a priori and dynamically acquired knowledge about users. A priori knowledge is generally built into the system by domain experts (Fisher, 2001; Kobsa, 1990).

Most of the knowledge that forms the basis for user models is acquired during dialogues. Some of it is explicitly acquired in specific subdialogues ("*If you are an expert, say ...*"), while some is implicitly inferred from system observations (Shifroni & Shanon, 1992). Some knowledge is acquired once and for all, thus constituting the static part of the user model. Most of it, however, is dynamic: human-computer dialogue system observations (duration and modalities) and inferences (words, information, concepts and actions) at each level of analysis (see Fig. 1), as well as statistics (counts, averages and percentages) are continuously updated by the human-computer dialogue system during processing.

Categories of knowledge in the user model. The knowledge contained in user models can be divided into (at least) three categories: environmental, individual and use (e.g. Brusilovsky, 2001). Knowledge about the environment may concern the user's terminal (phone or personal digital assistant), the modality (speech or keyboard), the locality (GPS coordinates) and any other environmental characteristics (noise level). Knowledge about the user as an individual may be relatively static, such as his or her age and gender, occupational status (student, worker, unemployed, etc.), stereotype profile (disabled, elderly, expert, novice, etc., see Fink & Kobsa, 2000), preferences and interests (Elzer et al., 1994). Individual knowledge tends to be more dynamic and is continuously maintained by the human-computer dialogue systems during the dialogues. Users' mental states (knowledge and goals) represent one of the most precise, comprehensive and dynamic types of user model for the representation of individual knowledge (e.g., Bretier et al., 2004). Knowledge about use concerns users' behaviour. This can also be modelled at each level of analysis handled by a human-computer dialogue system (acoustic, phonetic, phonological, lexical, syntactic, semantic, pragmatic, discourse and task). For example, the plan libraries used in plan recognition constitute a type of use model that encapsulates a user's possible behaviour at the action level. Another example is human-computer dialogue system experiences, collected dialogue after dialogue. Statistical use models can be calculated on the basis of these corpora of experiences (Zukerman & Albrecht, 2001).

Exploiting user models. In general, user modelling is used for adapting human-computer dialogue systems. The purpose of this adaptation may be to enable a human-computer dialogue system to take a specific user's profile, preferences and goals into account (Fisher, 2001). The exploitation of user models consists mainly in taking individual knowledge into

account when designing a human-computer dialogue system, particularly with regard to dialogue management. User modelling is extensively used to predict a user's behaviour. This prediction may concern problematic situations, such as those caused by speech recognition errors. Human-computer dialogue systems are then designed to avoid this situation (Litman & Pan, 2002). Prediction can be used by human-computer dialogue systems to anticipate a user's behaviour and minimise his or her contribution. Predicting that the user will ask for a specific item of information, a human-computer dialogue system may adapt its behaviour by proactively providing this information before the user asks for it (for a French-language example, see Baudoin, 2008). Prediction may also be used in a generative way in a user simulator. User simulators are used to automatically generate dialogue corpora for human-computer dialogue systems. These corpora allow systems to learn a number of dialogue strategies to complement their dialogue management design (Pietquin, 2004). Whichever approach is chosen, the literature on human behaviour provides a number of criteria which have to be taken into account when modelling users. The following section describes several studies of the effect of representing one's partner (system or human) on the content and dynamics of an interaction.

2.2 Impact of users' representations of a system on their utterances

When two human partners communicate via a computer system just as they would do if they were conversing face to face, they tend to share the same language. This phenomenon indicates that they have a mutual belief that they are referring to the same purpose (Brennan & Clark, 1996). Thus, the use of personal pronouns and reuse of the material contained in their partner's utterances reflect an active grounding in computer-mediated communication (McCarthy & Monk, 1994). In human-computer dialogue, the same sort of adaptation occurs between user and system (from the user's point of view): users naturally adjust their speech form to that of the system (e.g., Leiser, 1989; Walker et al., 2004). The fact that people adapt the form of their speech to that of their partner, even if the latter happens to be an artificial system, suggests that genuine dialogue processes can also take place during human-computer interaction. Some research has also indicated that people tend to anthropomorphize systems (Reeves & Nass, 1996). Arguably, the perceived capabilities of a system could also influence the way in which the user decides to interact with it. People may therefore interact differently with a dialogue system once they have adapted themselves to its capabilities/limitations.

How do a user's initial knowledge and/or beliefs impact on the way he or she uses the system for dialogue? During human-human dialogue, initial knowledge or beliefs about the interlocutor have a considerable impact on the way an individual constructs his or her messages, even if there is feedback from the interlocutor (Isaacs & Clark, 1989; Fussell & Krauss, 1991). For example, Isaacs and Clark (1989) demonstrated that speakers who were familiar with New York swiftly adapted their references according to which city landmarks were familiar or unfamiliar to their interlocutor. The influence of the interlocutor model, which can be defined as "...the representation of the interlocutor's technical and linguistic skills, which is partly constructed during the dialogue process" (Amalberti et al., 1993, p.551), extends beyond human-human dialogue, as individuals may profoundly modify the way they formulate their message according to the type of interlocutor that is being modelled (system or human operator).

Amalberti et al. (1993) found that the number of relevant items of information contained in an initial speaking turn was lower in a group interacting with an (simulated) artificial system than in a group interacting with a human operator. Users' knowledge (erroneous, in this case) of the system's understanding and problem-solving capabilities led them to provide fewer items of useful information in their initial request when they interacted with the system rather than with a human operator. Once several searches had been undertaken using the system, however, the number of items tended to increase. Richards and Underwood (1984) had previously drawn attention to this rise in the number of relevant terms for searches per utterance in the course of an interaction, associated with increasingly concise utterances (suppression of nonessential terms, such as articles, or politeness formulae). In short, it would seem that (1) the number of relevant items of information per utterance in a natural-language information search increases in the course of the interaction, especially following the initial speaking turn, and (2) the length of the utterances appears to remain relatively stable and may even increase during the interaction, at least in a spoken dialogue and especially following the initial speaking turn.

3. Detecting user familiarity – an empirical study

3.1 General methodology

Natural language dialogue systems have mainly been designed for use by people without any specific expertise who wish to consult or search for information. Task-oriented dialogue systems may be process-based (e.g., paying the bills of a service provider) or information-based (e.g., finding a restaurant, planning a trip). The aim of the experiments described earlier (data taken partly from Bretier et al., 2004, and Le Bigot et al., 2006) was to determine whether the use of a real information-based, task-oriented human-computer dialogue system (a guide for performing a restaurant search) would confirm the findings reported in the literature. The hypothesis was that, whatever the interaction modality, (1) the number of relevant items of information in the initial speaking turn of a natural-language information search would increase in the course of the interaction, and (2) the length of the utterances in the initial speaking turn would remain relatively stable during the interaction.

The human-computer dialogue system took the form of a restaurant search service - a prototype designed to help members of the general public search for restaurants in Paris. The system cooperates with customers and offers solutions matching their query as closely as possible. The application allows users to base their search on three criteria: restaurant location, price and food type. Users can express their request in natural language. This allows them to supply all the items of information needed for their request: (1) in a single utterance, (2) one at a time, (3) in mixed mode. For example, a user looking for a restaurant serving Mexican dishes in the fifteenth district of Paris with no particular price limit can either provide these two search criteria within a single utterance (*"I am looking for a Mexican restaurant in the fifteenth district"*) or else begin with an utterance stating the speciality (*"I am looking for a Mexican restaurant"*, wait for the system to respond, then complete his or her request in a second utterance (*"in the fifteenth [district]"*). A typical search can be broken down into two phases (e.g., finding a restaurant, planning a trip). (1) In the criteria initialisation phase, users express a more or less specific request and the system helps them by asking them to specify their criteria. If they wish, users may supply several criteria within a single "speaking turn" (see Fig. 2: U₂, U₄). (2) In the refining phase, after the users

have entered their criteria, they browse through the solutions proposed by the system (see Fig. 2: U₆). They may also request more detailed information about the solutions. Users can switch from one phase to the other at any time. This system (as Fig. 1) was accessible over the telephone (spoken), using corpus-based speech synthesis and continuous speech recognition based on statistical models (Hidden Markov Model). This application was also accessible via a Web interface (the speech recognition module and speech synthesis were replaced by written interface: keyboard and screen).

The system's user model allowed it to represent the dynamic aspect (over time) of the user, together with his or her request and the words that were used. However, it did not have sufficient information about the user's "profile" for it to adapt the way the system reacted.

Both experiments were conducted using the same technology, but with two different interaction modalities (totally speech-based or totally text-based). One experiment featured a speech interface (phone, microphone), while the other featured a Web interface (screen, keypad). There were twenty-four participants per experiment (min. age = 18, max. age = 51). Participants were assigned information search tasks (scenarios), either in the form of a list of criteria or in the form of an utterance. Twelve scenarios based on either two or three criteria (location, food type and price) were created for making queries (e.g., location: Louvre; food type: Chinese; price: twenty euros). The number of possible replies for each scenario varied from one to six. The order of criteria, number of solutions and order of scenarios were counterbalanced. Prior to the experiment, the application was briefly presented to the participants. They were told that they could dialogue in natural language with the application, and were instructed to use the service to find restaurants *as quickly and accurately as possible*.

S1: Welcome to [...]. What do you want?

U2: I'm looking for a Chinese restaurant for less than 10 euros.

S3: You are looking for a Chinese restaurant for less than 10 euros. I have found more than 10. You can either be more specific - for example, what is the location? - or consult the solutions [...].

U4: uh, the Sorbonne.

S5: You are looking for a Chinese restaurant near the Sorbonne for less than 10 euros. I have found 3. You can either be more specific [...], or consult the solutions [...]

U6: I'd like to consult the solutions.

S7: The first restaurant is [...]

Fig. 2. Example of a dialogue with the restaurant search guide (S = system, U = user).

The exchanges between the user and the system were transcribed word for word. The number of relevant items of information contained in the user's first speaking turn was recorded. An item of information was considered to be relevant for the search if it corresponded to one of the search criteria (food type, location, price). The number of words uttered during the first speaking turn was also recorded for each participant.

3.2 Results

In order to simplify the presentation of the results, the mean number of items of information and the mean word count were calculated for 2 successive scenarios (i.e., 6 blocks: block 1 = scenarios 1 and 2, block 2 = scenarios 3 and 4, block 3 = scenarios 5 and 6, block 4 = scenarios

7 and 8, block 5 = scenarios 9 and 10, block 6 = scenarios 11 and 12). The data were subjected to an analysis of variance, with the position of the scenario block as a within-participants factor. The results of the experiment featuring spoken interaction are reported in Fig. 3. The results of the experiment featuring written interaction are reported in Fig. 4.

Concerning the spoken interaction, the analysis revealed a main effect of the position of the block, $F(5, 115) = 6.695$, $MSE = .240$, $p < .001$, $partial\ eta^2 = .225$. The trend analysis highlighted a significant linear trend in the number of items of information provided in the first speaking turn as the participants went through the various scenarios, $F(1, 23) = 14.58$, $MSE = .472$, $p < .001$, $partial\ eta^2 = .388$. Although the number of items in the first speaking turn gradually increased as the different scenarios were performed, the analysis failed to reveal any effect of the scenario blocks on the word count for the user's initial speaking turn, $F(5, 115) < 1$.

Concerning the written interaction, the analysis revealed a main effect of the position of the block, $F(5, 115) = 5.936$, $MSE = .196$, $p < .001$, $partial\ eta^2 = .205$. The trend analysis highlighted a significant linear trend and a more marginally significant quadratic trend in the number of items in the first speaking turn in the course of the scenarios, $F(1, 23) = 18.41$, $MSE = .242$, $p < .001$, $partial\ eta^2 = .445$ and $F(1, 23) = 4.22$, $MSE = .156$, $p = .051$, $partial\ eta^2 = .155$ respectively. There was a gradual increase in the number of items provided in the initial speaking turn, although this figure stabilized in the final scenarios. The analysis failed to reveal any effect of the scenario blocks on the word count in the user's first speaking turn, $F(5, 115) < 1$.

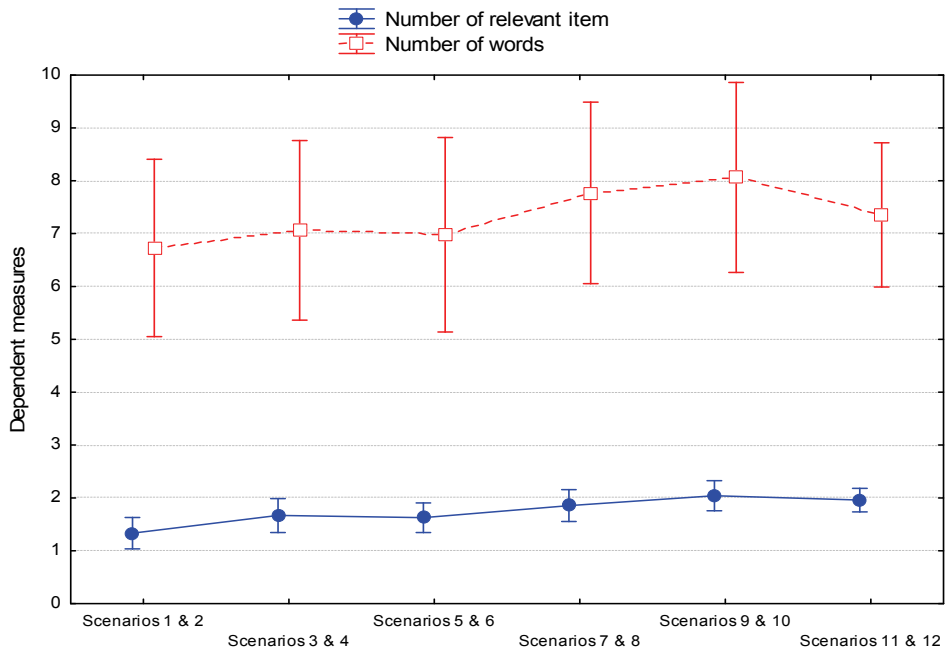


Fig. 3. Mean number of information items and mean word count for the user's initial speaking turn (spoken interaction).

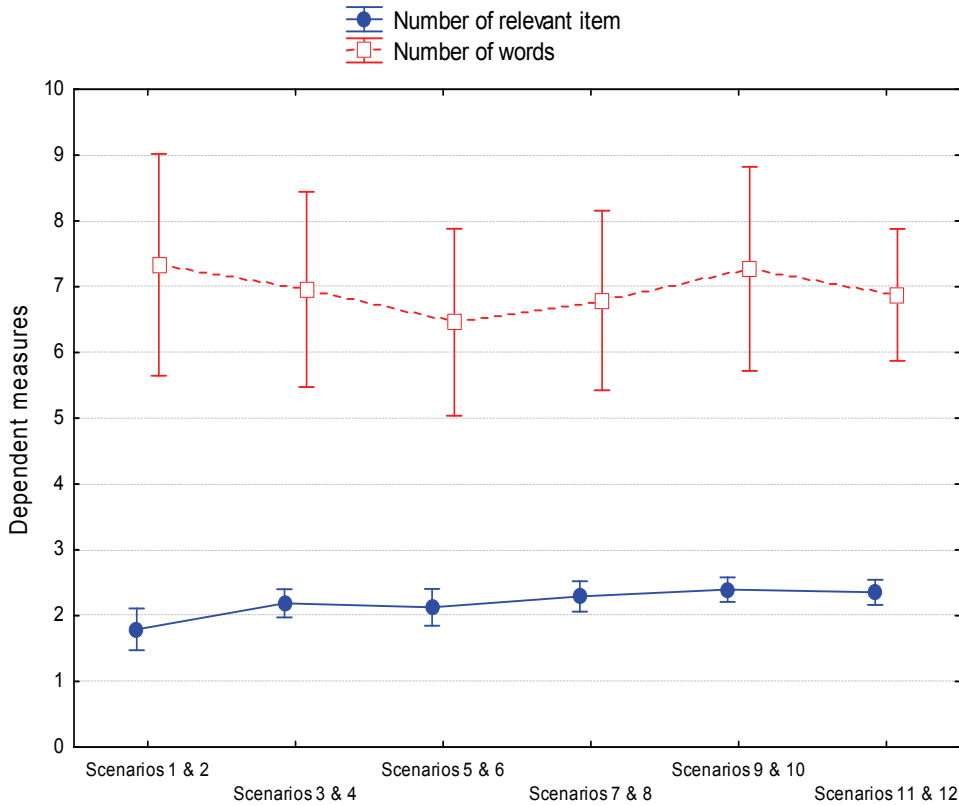


Fig. 4. Mean number of information items and mean word count for the user's initial speaking turn (written interaction)

4. Exploiting user familiarity to improve human-computer dialogue

4.1 User model and usability

These two experiments confirmed the findings in the literature concerning human-dialogue in a real system: the more a user interacts with a system, the more items of information he or she provides in the initial speaking turn, even though the word count remains relatively stable. These results appear whichever interaction modality is used, although user behaviour appears to stabilise more rapidly in the written modality than in the oral one.

The users had hardly, if ever, interacted with a dialogue system before, and in most cases, their representations of the system's understanding capabilities were most probably incomplete, if not erroneous. As they performed the blocks of scenarios, the users gradually increased the amount of information they provided. So, this finding, which can be interpreted in the light of psychological theories, could allow us to enhance the system's user model. The existing user model makes it possible to represent the dynamic aspect (over time) of the user's requests, but does not have sufficient information about the user's

"profile" to adapt the system's reactions accordingly. Just as the data that were collected in our study made it possible to identify clues to the user's degree of familiarity with the system so, too, should modelling. It should be entirely possible to detect the user's degree of familiarity with the system by calculating an index based on the number of items of information and the number of words used to convey these items in the first speaking turn. This index could then be exploited to enrich the application's user model in order to enhance usability. If the system detected that the user had little or no knowledge about the system, it could provide assistance or information about its capabilities (following Nielsen's usability principles: *Visibility of system status* and *Help and documentation*). Given that we are dealing with the capability of understanding natural language, this assistance could take the form of an example which the user could then use as a template for interacting with the system.

It is important to try and match the interlocutor model with the system's actual capabilities in order to promote human-machine cooperation. The system's construction of a user model can be seen as a strategy to reduce the need for the user to construct a (truthful) model of the system.

Hence, current information systems place a great many demands on their users because the latter must know what sort of information is relevant for their goals and what information is irrelevant, what proportion of the relevant information is contained in the dialogue system and how to find the relevant information within that system (Kobsa, 1990). Clearly, the last two points mean that the user must possess a "model" of a dialogue system. However, it should not be up to the user to build a model of the dialogue system. Rather, the system should build a model of the user. In the course of the interaction, the system should generate assumptions about aspects of the user that may be relevant in the task domain in question.

Lastly, natural language-based systems are often designed for use by the general public. Consequently, a great many default assumptions may be generated about the user's goals (e.g., all users of a hotel reservation system can be assumed to be searching for accommodation) and the user's knowledge and beliefs about a particular domain (i.e., a system with a user modelling component may assume that all its users possess basic knowledge in a given domain). Importantly, assumptions based on the user's inputs into the system are seen as the most direct and most reliable ones. Detecting and exploiting the user's familiarity is a possible alternative to relying on the user's inputs to make information systems more cooperative. The literature provides other examples (see Kobsa, 1990) of how a system's ability to make assumptions about a user's beliefs, goals and plans can enhance its cooperativeness - an ability that is especially important if the system is intended to be accessible to the general public.

4.2 Conclusion

In this chapter, we sought to underline the fact that user modelling must rely on the identification of relevant indicators for performing a given activity. The aim is primarily to improve the degree to which systems can adapt to human characteristics. As we have already discussed, user models accumulate knowledge about users gleaned from observations by the systems. These systems then apply design patterns (recommendations and heuristics) that are known to optimise usability according to the nature of their user model.

It is worth noting that when the system's observations are erroneous, the application of recommendations and heuristics may have a negative impact on usability. This is the principal motivation for regarding uncertainty as the standard situation.

In user modelling, statistical and probabilistic models both use observed sample results to make statements about unknown, dependent parameters. These parameters represent aspects of users' behaviour, such as their goals, preferences, and forthcoming actions or locations. Design patterns should be implemented by system designers as possible but not systematic behaviours. Implementing approaches that draw on the science of stochastic processes (e.g., dynamic Bayesian networks, see Zukerman & Albrecht, 2007), a user model may be able to cope with behavioural uncertainty stemming from human error or individually preferred interaction paths. For example, some user modelling work has been done using Bayesian networks and the resulting models applied to system behaviour design (e.g., Zukerman & Albrecht, 2001; Lemon & Pietquin, 2007; Williams & Young, 2007).

Improvements in the user model should logically lead to an improvement in the usability of the system, by adapting its behaviour to that of the user. For example, with respect to the heuristics discussed at the beginning of this chapter, one can imagine systems storing the user model they have constructed during an interaction and retrieving it at the beginning of the next session. This would certainly be relevant if the system was intended to be used frequently by the same user and for similar purposes, and could serve as the basis for personalization. Today, user profile modelling is clearly perceived as a prerequisite for Web service personalization, as illustrated by technologies such as profiling (painting a portrait of users based on fill-in forms), click-stream analysis and web usage mining, or collecting data about a user's movements in a Website (Manber et al., 2000; Pierrakos et al., 2003).

Going back to the specific issue addressed in this chapter, we have emphasised that users provide information about themselves via the pattern of their (first) request: they implicitly provide information about their familiarity with the natural language interaction task being performed with the system. Familiarity could arguably be assessed on a basis other than the index we have proposed for the first request (number of relevant items of information/number of words), including other linguistic parameters, or the speed of message production. Indeed, other criteria are clearly needed and may well be found in the course of future research. Furthermore, these criteria could be tested with respect to their external validity, by changing the environmental parameters (e.g., is the result independent of the message production modality?) or the user's parameters (e.g., is the same result observed in different age groups?). Lastly, once the user model has been enriched with this information, the latter must be exploited: providing assistance to someone who has no accurate knowledge of the system's capabilities by providing an example he or she could use as a template, or correcting a misconception; reducing the amount of guidance and offering greater flexibility when the system detects a degree of experience on the part of the user.

To conclude, the design of human-computer dialogue is very much an interdisciplinary enterprise. On the one hand, computer sciences (e.g., speech processing, artificial intelligence) are applied to phonetics, linguistics and discourse. On the other hand, these "technological" fields are cross-fertilised by the human sciences. Studies in psychology and

ergonomics help system designers to take users' needs and behaviours fully into account. We believe that this is particularly true in the case of user modelling for dialogue systems. Other specialised applications that have recently been proposed (Griffith et al., 2007; Slaney et al., 2003), including user group modelling and cooperation or multitasking user modelling, will also strengthen the case for the application of cognitive research.

5. References

- Amalberti, R., Carbonnel, N. & Falzon, P. (1993). User representations of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies*, Vol. 38, 547-566.
- Baudoin, F. (2008). Personnalisation des systèmes de dialogue en langage naturel : une méthode d'anticipation rationnelle d'actions communicatives [Trans. Personalization of natural-language dialogue systems: a rational method for anticipating communicative actions]. *Unpublished doctoral thesis*, Paris 6: Paris (France).
- Brennan, S.E. & Clark, H.H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 22, 1482-1493.
- Bretier, P., Le Bigot, L., Panaget, F. & Sadek, D. (2004). De la représentation de l'interlocuteur vers un modèle utilisateur formel pour le dialogue personne-machine [Trans. From the interlocutor's representation to a formal user model for human-computer dialogue], *Proceedings of the 16th international conference of the Association Francophone d'Interaction Homme-Machine IHM 2004*, pp. 21-28. Namur (Belgium), August-September, 2004, ACM Press.
- Brusilovsky, P. (2001). Adaptive Hypermedia. *User Modeling and User-Adapted Interaction*, Vol. 11, No. 1-2, 87-110.
- Gorin, A.L., Riccardi, G. & Wright, J.H. (1997). How may I help you? *Speech Communication*, Vol. 23, No. 1-2, 113-127.
- ISO 9241-11. (1998). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.
- Elzer, S., Chu-Carroll, J. & Carberry, S. (1994). Recognizing and utilizing user preferences in collaborative consultation dialogues. *Proceedings of the Fourth International Conference on User Modeling*, pp. 19-24, Hyannis (MA), August 1994, Springer, Vienna, New York.
- Fink, J. & Kobsa, A. (2000). Adaptable and adaptive information provision for all users, including disabled and elderly people. *User Modeling and User-Adapted Interaction*, Vol. 4, 163-188.
- Fischer, G. (2001). User modeling in human-computer interaction. *User Modeling and User-Adapted Interaction*, Vol. 11, No. 1-2, 65-86.
- Isaacs, E. & Clark, H.H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, Vol. 2, No. 6, 26-37.
- Kosba, A. (1990). User modelling in dialog systems: potentials and hazards. *AI & Society*, Vol. 4, No. 3, 214-240.
- Lai, J., Karat, C.-M. & Yankelovich, N. (2008). Conversational speech interfaces and technologies. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving*

- Technologies and Emerging Applications, 2nd edition*, A. Sears & J.A. Jacko (Eds), (pp.381-391), Lawrence Erlbaum Associates, ISBN: 978-0805844689, New York.
- Le Bigot, L. , Jamet, E., Rouet, J.-F., & Amiel, V. (2006). Mode and modal transfer effects on performance and discourse organization with an information retrieval dialogue system in natural language. *Computers in Human Behavior*, Vol. 22, 467-500.
- Le Bigot, L., Terrier, P., Amiel, V., Poulain, G., Jamet, E. & Rouet, J.-F. (2007). Effect of modality on collaboration with a dialogue system. *International Journal of Human-Computer Studies*, Vol. 65, 983-991.
- Lemon, O. & Pietquin, O. (2007). Machine learning for spoken dialogue systems. In *Proceedings of Interspeech, 2007*, Antwerp (Belgium), pp. 2685-2688, August 2007.
- Leiser, R.G. (1989). Exploiting convergence to improve natural language understanding. *Interacting with Computers*, Vol. 1, No. 3, 284-298.
- Litman, D.J. & Pan, S. (2002). Designing and evaluating an adaptive spoken dialogue system. *User Modeling and User-Adapted Interaction*, Vol. 12, No. 2-3, 111-137.
- Manber, U., Patel, A. & Robinson, J. (2000). Experience with personalization on Yahoo! *Communications of the ACM*, Vol. 43, No. 8, 31-34.
- McCarthy, J.C. & Monk, A.F. (1994). Channels, conversation, cooperation and relevance: all you wanted to know about communication but were afraid to ask. *Collaborative Computing*, Vol. 1, 35-60.
- Nielsen, J. (1994). Heuristic evaluation. In *Usability Inspection Methods*, J. Nielsen & R.L. Mack, R.L. (Eds), (pp. 25-64), ISBN: 0-471-01877-5, John Wiley & Sons: New York.
- Pietquin, O. (2004). *A Framework for Unsupervised Learning of Dialogue Strategies*, Presses Universitaires de Louvain, ISBN: 2-930344-63-6, Louvain.
- Pierrakos, D., Paliouras, G., Papatheodorou, C. & Spyropoulos, C.D. (2003). Web usage mining as a tool for personalization: a survey. *User Modeling and user-Adapted Interaction*, Vol. 13, No. 4, 311-372.
- Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. Cambridge University Press and CSLI, ISBN: 1575860538, New York.
- Richards, M.A. & Underwood, K.M. (1984). How should people and computers speak to each other? In *Proceedings of Human-Computer Interaction (INTERACT'84)*, pp. 33-36, London, September 1984, Elsevier Science Publishers.
- Shifroni, E. & Shanon, B. (1992). Interactive user modeling: an integrative explicit-implicit approach. *User Modeling and User-Adapted Interaction*, Vol. 2, No. 4, 287-330.
- Wilkie, J., Jack, M.A. & Littlewood, P.J. (2005). System-initiated digressive proposals in automated human-computer telephone dialogues: the use of contrasting politeness strategies. *International Journal of Human-Computer Studies*, Vol. 62, 41-71.
- Walker, M.A., Whittaker, S.J., Stent, A., Maloor, P., Moore, J.D., Johnston, M. & Vasireddy, G. (2004). Generation and evaluation of user-tailored responses in multimodal dialogue. *Cognitive Science*, Vol. 28, No. 5, 811-840.
- Williams, J.D. & Young, S.J. (2007). Partially observable Markov decision processes for spoken dialogue systems. *Computer Speech and Language*, Vol. 21, No. 2, 393-422.
- Zukerman, I. & Albrecht, D. (2001). Predictive statistical models for user modeling. In *Proceedings of the Eight International Conference on User Modeling*, pp. 5-18, Sonthofen (Germany), July 2001, Springer, Vienna, New York.

Zukerman, I. & Albrecht, D. (2007). Introduction to the special issue on statistical and probabilistic methods for user modeling. *User Modeling and User-Adapted Interaction*, Vol. 17, No. 1-2, 1-4.