Tailoring output to the user: what does user modelling in generation mean?

Karen Sparck Jones
Computer Laboratory, University of Cambridge
New Museums Site, Pembroke Street, Cambridge CB2 3QG, UK
December 1989

1 Introduction

There is a general presumption that natural language output generated by interactive systems should be tailored to the user. Thus in question answering or dialogue, outputs should not only be linguistically fitting, e.g. be anaphorically coherent, and appropriate in content, e.g. give answers to questions; they should provide cooperative responses, e.g. by volunteering information, based on the inferred beliefs, goals etc of the user (Webber 1986). Tailoring output to the user, in other words, assumes a model of the user (Kobsa and Wahlster 1989).

But what does this imply? Does having an individual user model mean that the system should always be looking for, and address, the person behind the user's inputs, forming hypotheses about the person that influence both the content and form of outputs, whether these convey information to, or seek further information from, the user? Or is the user no more than his inputs, so generation should just fit these? In much work on interactive systems there is a tacit assumption that the more tailoring the better, which is taken to imply that the more intensive and extensive the modelling the better. This seems natural for the advisory or consultation paradigm within which many application systems fall; and it seems plausible by analogy with interaction between humans given the aim of making human-machine interaction as much like human-human interaction as possible (as tacitly assumed in e.g. Grosz and Sidner 1986).

I have argued in earlier papers that user modelling as a whole has many elements, and that this complex notion needs unpacking for a proper approach to modelling in any particular case. I have also argued that the difficulty of obtaining robust enough information means that modelling should be very conservative. In this paper I shall look specifically at what this means for generation. The conclusion is that, in a fundamental sense, the user is indeed his inputs. I am not claiming this as an original conclusion. My aim is rather to explore and make clear, using my previous analysis, what this interpretation of user modelling implies for output tailoring. The argument is thus that we should be doing what we are doing, but not for the reasons commonly taken for granted; and if the grounds for model-based generation are not those usually assumed, what constitute rational aims for the future are not those usually assumed either. My concerns are therefore with the basis for generation, and not with the specific mechanisms of generation.

Section 1 recapitulates my previous analysis: this somewhat lengthy preamble is needed to provide an adequate framework for the subsequent discussion. In Section 2 I consider some (hypothetical) example systems designed to illustrate alternative approaches to tailoring output. In Section 3 I look at the implications of these examples for the characterisation of

user models. Section 4 then considers what follows from this for generation, with a further example. Section 5 summarises the argument and conclusions. I shall try to flesh out the distinction between modelling the user in a strong sense as an autonomous human individual with characteristics which are not relevant to the functional task for which the system is designed, and modelling in a more restricted sense limited to those characteristics that are relevant to the system's task, and examine what follows from this distinction for output generation with natural language interfaces. In practice hitherto there has been no call for this distinction, mainly because our system tasks and modelling capabilities have been limited, but also because maintaining dialogue is deemed a system task. Maintaining dialogue is, however, an essentially subordinate task: it is possible to separate a system's primary decision task - diagnosing a disease, determining a flight booking - from dependent support tasks, like gathering information and communicating decisions. In modelling, therefore, the main question about how the user is viewed is in relation to the system's primary decision task. However any need to communicate a system decision depends on some view of the user which may presuppose some characteristics not needed for primary decision purposes, and modelling beyond decision data is to that extent justified. Indeed interacting comprehensibly and cooperatively in natural language may require a rather more substantial view, in his role as discourse participant, of the user. But these considerations still only imply a generically based and restricted approach to modelling, and hence a moderate approach to output tailoring.

2   Background

In previous papers (Sparck Jones 1984, 1989) I considered user modelling in a wide variety of contexts, to provide a general framework within which particular systems with specific tasks and domains, and modelling conditions, could be placed. My aim was to supply the base for a proper analysis in approaching modelling in any particular case. I identified both several classes of modelling factor and a range of system functions, implying a complex set of possibilities for modelling to serve functional requirements.

Thus modelling has to take into account the user's role in relation to the system: there are two possible roles, patient, referring to the subject of the system's primary decision making, and agent, referring to the system's driver. There may be systems with only one human role, as patient or as agent, or systems with both roles, which may be filled by the same or distinct people: the last in particular may imply a rich modelling situation. [I shall use "user" generically where the point applies to either patient or agent, and whether or not distinct human beings are involved: where any of these distinctions apply, and they are not made explicitly, they should be clear from the context.] Modelling has also to take the status of the user's properties into account: they may be decision properties for the system's primary task, or other, non-decision properties. The nature of the user's properties is relevant as well: they may be broadly categorised as objective (and hence relatively accessible and uncontroversial), or subjective (as in beliefs and goals, which are generally less accessible). and also categorised as static (unchanging through a session with the system) or dynamic (changing over the session). These category distinctions cut across one another, and also across the decision/non-decision distinction, and may of course hold in quite different ways when there are distinct people as patient and agent.

At the same time, because modelling is done for a purpose, it is necessary to consider the different functions it can serve. The obvious one is effectiveness, the system's ability to

reach a 'correct' decision. But it is also possible to address efficiency, the system's ability to reach a decision economically, and acceptability, the way what the system does and how it does it looks to the user. Acceptability is a complex notion, referring both to the intrinsic decision making of the system and to how this is presented to the user. In general, information about any property of a user in either role can be exploited to serve every function, with the exceptions that agents do not have decision properties and that effectiveness depends directly only on the patient's decision properties.

Thus in the social security illustration of Sparck Jones 1984 and 1989, for example, the system used information about non- decision properties of the elderly patient, like her honesty and her ignorance of social security benefits, to control the acquisition of information about her income and to provide explanations making its announced decision more comprehensible and hence more acceptable to her, as well as information about her decision properties - age, disability, and so forth - to calculate the benefits themselves; and it used its agent model of the mediating clerk both to seek information using technical language and to bypass his supposed chauvinism. This and the other (hypothetical) systems considered in Sparck Jones 1984 show that though differences in the domain and task specifications cut very different ways on the distinctions made, these distinctions are always relevant, and modelling recognising them can influence system behaviour and performance. The examples showed in particular that the non-decision properties of patient and agent could in principle be significant, and that they should therefore be noted, checked, and used. These included, in the social security example for instance, the patient's honesty (not in itself a decision property for benefit calculations), and the agent's chauvinism.

Within the general framework of factors and functions, individual systems with different tasks and modes of use can impose quite different constraints on, or offer quite different opportunities for, modelling. Where there is a human patient, decision property modelling is of course required, but beyond that it has been tacitly assumed that modelling is always useful and hence desirable. Information about the user is obviously and also most naturally acquired for the patient role, and about decision properties (especially objective ones), to support modelling for effectiveness. But there are many difficulties in obtaining relevant and reliable information even here, and the value of modelling is much more problematic in relation to other user characteristics or system contexts. This is clearly seen where the user's implicit beliefs and plans are important as they are patient decision properties, but is also the case in situations where other roles and property categories are relevant.

I therefore suggested in Sparck Jones 1989 that, for the kinds of system with very limited power we can currently envisage, where the evidence for modelling hypotheses is likely to be poor in quantity and quality, it is necessary to be conservative in modelling both in relation to the system's primary decision making, as in an expert system, and in relation to the system's conduct of its interaction with the user. The system should not be quick to make assumptions about the user, whether in seeking or providing information, and should be slow in abandoning general for specific grounds either in interpreting the user's inputs or in determining the form and content of its own outputs. Thus it is generally safer to avoid tailoring outputs on the basis of selective but poorly supported hypotheses about the user, and to seek more information from the user if this would not be oppressive or to proffer information on a neutral basis, and perhaps somewhat beyond the bare minimum. This applies both to the use of decision properties, especially where it is not necessary to attempt exhaustive data gathering before offering a candidate decision, and more importantly, to the use of non- decision properties. Thus with a notional travel advice system, it is not necessary to hypothesise from an inquiry

about trains that the user is motivated by economy (price being a decision property and assuming trains are cheap), or alternatively by fear of flying (a non-decision property), to provide a perfectly adequate response, even if in principle adopting either of these specific hypotheses might lead to a more targetted one; in practice, with a poorly supported hypothesis about miserliness or fear (which would be only too likely), a targetted response could be very inappropriate, and it would be much safer to give straightforward train information. But even with a conservative approach it is important to recognise the various aspects of modelling I distinguished, for example that users can have different roles even where they are not different people, and that it may be feasible as well as useful to exploit modelling to serve several functions, as in output designed to promote both efficiency and effectiveness.

These discussions used expert (alias knowledge based) systems as examples, but with a wide definition, and were, moreover, intended to apply to a very broad range of situations involving computational systems and their human users, including those where the system's primary task is user independent, or where the human involved is more a subject than a user in the ordinary sense, or where interaction is slight, as well as those, like teaching systems, where the system exists for the sake of its user, and where it operates highly interactively. But even in cases where the user is more conspicuous there can be great variations, for example in the type of task, in the source of initiative in interaction and of direction in problem solving, and so forth.

In general, however, it was assumed that, within the framework of the system's global specification, referring to its task (e.g. diagnosis, dissemination), domain (machines, law) and user class(es) (children, technicians), modelling could in principle apply as much in less obvious cases, like process control, as in the more obvious ones like consultation or instruction; it was further evident not only that all the system's functions could involve modelling, but that to support both the system's internal effectiveness and efficiency functions and its external acceptability function, the system's interaction with the user could exploit all of the kinds of model data and hypotheses. [If effectiveness is in decisions about the user, outwardness is derived or secondary.] Modelling may therefore, as the illustrative examples in Sparck Jones 1984 and 1989 showed, be very complex even if in practice, playing for safety implies radical simplification.

These discussions did not presuppose communication in language, and especially natural language, though natural language is a particularly powerful means of communication and in general provides more support for modelling than other means. However even where natural language is used for communication it may not always be required for both input to and output from the system, though modelling can apply to either; [For input I am considering here only modelling derived from interpretation and not modelling required to achieve interpretation, though in general the points referring to the former also refer to the latter; but for output I am of course considering modelling influencing generation.] modelling may also apply whether inputs or outputs occur independently, i.e. are not paired in any material sense, are essentially paired as in 'one- shot' exchanges, or form elements of a more extensive dialogue; further, modelling may be relevant, especially for the system's output, whether the output content is an end in itself, as in announcing a system decision, or is a means to an end, as in seeking more data. Finally, modelling can be related to the form as well as the content both of the system's output and its input. In this paper I shall consider only communication in free natural language, taking it for granted that it will be required both for system input and system output, for a variety of specific purposes, whether or not it is also the sole means of communication.

What my previous analysis implied, therefore, was that within a general context presupposing cooperative behaviour on the system's part for those operations that are connected with its user(s), the system's outputs can and should be tailored to the user(s), i.e. in a given session to the one or more individual users (agents and/or patients) involved. [An industrial plant controller, for instance, will have other concerns as well.] This applies whether information is being supplied or sought and whatever function is being served. Of course the extent to which an output is motivated by the user will depend on the system task, on its global and local state, and on the communicative purpose of the output (reply versus signal) as well as on such features of the user as their role; but it is commonly assumed that output should be tailored to the user as an actual and not just representative individual. However as I suggested in Sparck Jones 1989, the fact that modelling is notionally feasible as well as apparently desirable is offset by the fact that it is difficult to get reliable modelling information, so caution is required in forming hypotheses and applying them to internal processes and external interaction.

Thus given the general framework outlined, I shall now consider the specific issue of tailoring output in more detail: what could tailoring output in principle, and what should it in practice, look like?

I shall try to answer these questions, and especially the second, through examples. The examples in the next section are designed to compare outputs pushing user modelling with alternatives adopting the more cautious approach advocated in Sparck Jones 1989, at particular points in a notional interaction. I have given a number of examples in order to show how the comparisons work across different contexts in terms of system tasks and user roles. The later and more elaborate illustration of Section 4 looks in more detail at the implications of realism for generating tailored output in successive system responses in an extended dialogue.

3   System examples

In many actual or planned systems, there are simplifications either because the number of factors affecting modelling is reduced or because the factors present are so only in a limited form. For example in the common advice case, patient and agent are identical, all the recognisable user properties are decision properties, and the system's task is to react to the user's inputs ad hoc; for example with a train inquiry system there may be no global system goal to achieve in the form of designing a whole trip to meet the user's requirements; the system may have to do no more than respond to user-set 'local' goals by supplying individual pieces of information like a train time. The same applies to systems like UC (Wilensky et al 1988). However even here we can have a mix of objective and subjective user properties, or of static and dynamic ones. We can also, even within this restricted consultative paradigm which has been a common base for work on user modelling, have more complex situations since we can distinguish, somewhat schematically, the cases where the user is talking about himself from those where he is talking about something else. In the first he is seeking advice about and for himself, for example about books which would suit him; in the second about something else, but (directly) for himself, for example about a failed engine. The contrast between the user talking about himself and about something 'out there' affects modelling because in the first case the user's attitudes are the paramount factor, and in the second case they are not so, though for some task systems they may be very important, for example in choosing college courses. In instruction systems the user talking as himself is the dominating factor. Expert

systems with a single user can be designed to occupy very different positions along the line from patient dominance to agent subservience.

My examples are designed to cover differences of task, user role, executive dominance, i.e. whether the task execution is system or user driven (it may be a mixture of both), communicative direction, i.e. whether communication is system or user initiated (or a mix), and interactive mode, i.e. whether we have independent, one-shot or dialogue interaction. The examples are primarily of consultation situations as these are obvious ones for user modelling, but of varying kinds, and I have included some non-consultation cases as well, to emphasise the wider relevance of the points being made. But I have deliberately, in spite of the importance of user modelling for instruction, excluded instruction examples because they may often involve non-linguistic inputs (for instance worked sums) or need more space than is convenient here. The points made nevertheless apply to them too. In referring to advisory and consultation systems, on the other hand, I am referring to more than what Morik 1989 calls 'information-seeking' as opposed to 'action- achieving' systems. A travel system that only recommended holidays would be an information-seeking system for Morik, one that made bookings as well would be an action-achieving system. But an action-achieving system can allow for consultation by the user, as in the preliminaries to booking, so the points made apply to advice-seeking interaction in general. Consultation or inquiry may also of course be an allowed form of interaction with a system primarily intended for independent action as in, say, a police traffic controller or as illustrated in a more limited way by XCALIBUR, for example (Carbonell et al 1983).

To simplify I shall consider only the effectiveness and acceptability functions. Effectiveness is an internal system function determined by the system's task though, depending on the nature of this task, it may have more or less outward manifestation for the user: more as in instruction, less as in plant monitoring. Acceptability is an external function with a wholly outward manifestation. Note here that giving individuals what they like, say in an advisory context driven by personal preferences (as in choosing dresses), is still a matter of effectiveness because the system's decisions are based on the user's wants; acceptability has to do with the way things are presented to the user. However whether a particular feature of the system output in the advisory case is motivated by effectiveness, i.e. giving the system's decision, or by acceptability, giving it in one way rather than another, is a complex matter and one with no simple relationship to the user's decision or non-decision properties. While in an advisory situation communicating the system's decisions might be deemed a necessary component of effectiveness, communication raises many presentational issues and it is hard to adhere to any 'neutral' notion of communication which might be taken as associated with the decision taking itself. I shall nevertheless assume that 'neutral' information exists and has to be conveyed as the decision advice through the system's output. (I shall return to this point in Section 4.)

It may also be noted that even in preferentially-based advisory situations, if the system's decision making is complex, or the relation between the way the user couches his preferences and the way the system characterises the objects it deals with (as in choosing library books), whether the system's specific decisions in themselves are acceptable may itself be a concern 1for the user, as in other types of application. Presenting decisions comprehensibly merges gradually into justifying them convincingly. I shall assume here that confidence in the system as such, and therefore its acceptability as a whole as a device for carrying out its generic task, is not in question, so acceptability is primarily a presentational matter relating to individual decisions. [This is quite compatible with having a limited system as long as the user is aware of the limitations.] This clearly includes not merely the 'outward trappings' of response, but

rendering the system's decisions explicable in terms of its knowledge resources. Whether or not in an advisory situation the system's decisions are in fact accepted is a separate question: with the travel advisor for instance, the user may recognise that the recommended trip is wholly reasonable, i.e. is well-founded and the best on offer, but still not take the advice.

The examples below are all of hypothetical systems, with idealised language processing capabilities. Note that the pieces of discourse given are not necessarily opening ones for the session, and that the system could already have some model information, derived from the current or a previous session. In these examples U represents user inputs, S system outputs. System outputs marked with a *, as in *S3, are key ones to be discussed; the immediately following outputs, for example S3a, are alternatives to these. These key outputs and their alternatives represent different ways and degrees of tailoring to the user.

Example 1

This is an equipment repair advisor, so with only a human agent. Imagine we have the following portion of dialogue.

```
    U1  : The freezer seems to be on all the time, making a lot of noise,
          and its only got a pound of coffee beans in it.
    S1  : How old is it?
    U2  : About two years.
    S2  : Have you checked the controller?
    U3  : Seems to be OK: the thermostat is OK.
 *  S3  : But did you look at the thermometer?
    S3a : What is the state of the thermometer?
```

I am assuming here that the controller relating thermostat and thermometer is designed so the state of the thermometer shows whether the thermostat is working without any independent indicator on the thermostat itself, i.e. that it is the state of the thermometer rather than that of the thermostat which shows whether the thermostat is working. Then U3 is ambiguous because it could indicate either that the user does understand the mechanism and having looked correctly at the thermometer is indicating his conclusion about the thermostat, or that he doesn't understand and is looking at the wrong thing. *S3 would be a natural system response to the second hypothesis.

But it is not necessary to be so user-oriented. The system could combine its knowledge of the structure of the controller with the amount of information explicitly supplied by the user, without bothering about the user as the man behind U1, U2 and U3 and about the reasons for his inputs, to decide on the alternative output S3a. [If the user had already input information about the thermometer as well as about the thermostat, the situation would of course be quite different.] S3a, without being tailored to the user as someone with beliefs etc, i.e. as being more than his inputs, is formulated in a manner which would make it acceptable to the user even if he had looked at the thermometer: i.e. it is not couched to imply he is incompetent.

Example 2

This is a horticultural designer where because, as in the travel agent case, the system is designed to work from the user's attitudes, we have a human patient, who we will suppose is also the agent. (As noted earlier, if there is an actual garden involved, the extent to which its characteristics are treated as (objective) properties of the user or as defining a non-human task subject is determined by the global system task specification, i.e. whether it is to assist or determine; clearly there are complex borderline cases, especially with systems with a range of behaviour options.)

Suppose, then, we have: [I am assuming the system already knows the user's garden is in Western Scotland.]

```
    U1  : The soil is rather heavy, so what's good for that? I believe roses
          are OK but I don't like them.
    S1  : Well there's peonies and ranunculas.
    U2  : Do they need a lot of watering?
 *  S2  : Only occasionally when its very dry.
    S2a : Not over the year.
    S2b : They never need a lot at once.
    S2c : They will only need watering if it is very dry, which does not
          often happen in Western Scotland, and even then will need only a
          modest sprinkle each day.
```

U2 could suggest various hypotheses about the user: that he doesn't like watering, that he can't afford to buy water, or that he can't carry heavy cans. *S2 is essentially motivated by the first hypothesis and so is tailored to this view of the user; the alternative S2a is motivated by by the second, and S2b by the third. But it is surely adequate and indeed preferable, as each of these replies though natural in form is not wholly unambiguous, to generate the straight and full response of S2c, i.e. a response to the simple question U2 amplified using system domain knowledge about different aspects of watering relevant to gardening. Carefully generating an unambiguous response for each hypothesis, e.g. 'Even in a dry season they never need much water each day' instead of S2b would, after all, lead to nearly as much, and as slightly plonking, an output as S2c, with the added danger of missing the point. In fact, in this case as in the previous one, the system needs generic task knowledge as well as domain knowledge: it depends, implicitly if not explicitly, on some notion of the kind of task for which it is designed, here advising, as well as on knowledge about the topic on which advice is sought and being offered. The task and domain knowledge together constitute the system's generic application model, which is instantiated in a specific session as a particular application model. (Note that the assumption in the example is that while the user's attitudes to flowers are part of the system's decision model, his attitudes to watering are not; however information about the watering needs of plants is part of the domain knowledge base used for decisions.)

Example 3

Here we assume a libel advisor operated by a lawyer with a client who wishes to sue. The publicity typically associated with libel suits means that the patient's attitudes are important; but as the separate agent has to interpret the system's advice, his properties may also be

significant. Such a legal advice system could be a full decision maker or merely responsive to queries: here I assume the latter.

```
      S1  : Your client wants to sue Megajunk News about an article calling him
            a bigamist?
      U1  : Yes.
      S2  : And is he?
      U2  : He says not.
   *  S3  : How resilient do you think he is? Megajunk aren't exactly scrupulous.
      S3a : Ask him if he believes he could withstand a relentless and unscrupulous
            newspaper campaign with harrying reporters and numerous damaging
            stories about him.
```

In *S3 the system is seeking to exploit a model of the lawyer (which we will assume is in fact a personal one already built up, and not just a type one) as an experienced judge of the real qualities of clients. But if the system has to rely for its client assessment on a model of the lawyer which is more than a stereotype this implies that it should have a high quality model which would be very difficult to establish because it would depend, in cases like the present, on some means of evaluating the accuracy of the lawyer's judgments of his client. But even then, it is not clear that this would be superior to using the agent merely as a (deemed reliable) transmitter, and using the previous dialogue plus task and domain information to obtain the information required about the client via the alternative system output S3a.

Example 4

Just to emphasise that these considerations also apply to cases where the user is not seeking advice, and where the system takes independent, perhaps concurrent, actions, imagine an automated warehouse shipping system translating messages into truck loads and journey instructions.

```
      U1  : Four hundred of B700s to Smiths at Reading.
      S1  : OK.
      U2  : They'll need some P20s to go with them.
      S2  : Two hundred? It would need two trucks.
      U3  : No, three hundred.
   *  S3  : There's one P20 per pair of B700s, so you don't need that
            many P20s.
      S3a : You said four hundred B700s.
      S3b : That's a P20 for each pair of B700s plus 100 P20s.
```

The system's *S3 response is natural for the hypothesis derived from U3 that the user does not know what the precise relation between B700s and P20s is, i.e. that it is two for one. S3a is motivated by the alternative that the user has forgotten by now, or incorrectly remembers,

how many B700s he said. But S3b, which is simply descriptive, is a better all-round response because it is informative but neutral in tone.

Example 5

Finally, I shall consider situations where there is no immediate dialogue context for the system to use, i.e. where the user's input or system's output are not part of an interactive sequence. Quite apart from the beginning of ordinary dialogue, these situations may occur naturally with some task systems, even if the task is complex and follow-up dialogue is possible; they need not however imply that the system has no previously acquired information about the user. They also occur with more restricted tasks like conventional database query, where current front ends, apart from question recording to deal with ellipsis or more efficient searching, and possibly ad hoc sub-dialogue to handle question defects like syntax errors, operate on a one-shot exchange basis and cannot treat input sequences as groups and recognise the user goals or plans motivating them. But this is a somehat artificial limitation, as it would be preferable in at least some applications to have a more powerful consultation front end exploiting the database management system as a tuple getter. Operational plant control is a more natural example since there will be occasions where, even with a powerful front end, simple exchanges or unidirectional messages will be appropriate, and where the user is not necessarily seeking advice. For present purposes the assumption is that the system has some available modelling information, say obtained through earlier interaction, so it is not modelling the user solely on the basis of the specific current input.

Now consider:

```
     U1  : The packing machine seems to be producing unusually slowly.
 *   S1  : How much more slowly than normal?
     S1a : How often are the boxes coming off?
```

*S1 could be motivated by a user model that has established the user as a thoroughly experienced observer, i.e. it is not just assuming any operator is experienced (this is easier than in the libel case because the system has an independent task to provide an evaluation context). But while it may be flattering to the agent to be treated as a good judge, from the system's point of view the alternative response S1a, which is not motivated by anything but its domain model, is as good, if not better as capable of getting more precise information. The same would apply to an autonomous system output message, so S1a below is as good in real terms as *S1:

```
 *   S1  : Shut down the feed valve if the flow's abormally lumpy.
     S1a : Shut down the feed valve if any lumps are larger than your hand
           or the gaps between bursts of feed last longer than three seconds.
```

It may be that the former would give a finer plant control, but it is, I maintain, somewhat riskily dependent on the system's capacity to evaluate its human operators. This risk might

be reduced by administering explicit 'examinations', but these could hardly be expected, for a complex plant, to be exhaustive. Of course in this last single output case, the system has no preceding linguistic input to exploit as well as its internal knowledge, so S1a would be based on its application model alone.

4   Example implications

My argument in all these examples has been that while it would in principle be possible for the system to exploit a user model based on more than the data supplied by the user's language inputs plus the application task and domain knowledge it has for its decision making, it can serve its various functions, in this discussion acceptability as well as effectiveness, perfectly well with just the input data and application information. (Whether or not the life of such a user model is extended beyond the session will depend on the application.) This applies even in the advisory case, where the need for as full a user model as possible appears to be strongest, as well as in systems or interactions where consultation is not the primary activity.

But what exactly does this imply for the notion of user model and for the part it should play in generation?

In general, in those cases where there is a human patient, the important point about the application model to which the patient's decision properties are related is that this is a coherent, or at least connected, whole; i.e. it is generally, if not always, possible to make inferences explaining given patient data or to relate given data to other possible properties of the patient. The importance of the connectivity supplied by the application model is clearly seen in the instruction case. If the pupil is providing the wrong answers to the set mathematics problems, and this is not deemed attributable to frivolity, boredom, fatigue, mistyping or 'simple error', the system has to be able to postulate explanatory processes which are couched in terms appropriate to the domain, e.g. that the pupil has carried out an inappropriate operation, but one of a number-manipulating kind: the notion of 'mal-rules' presupposes this kind of domain- relevant systematicity.

But the same point applies in other contexts, and specifically in relation to situations where the user's dynamic mental properties are in question: these are presumed to be motivated by beliefs, goals and so forth that the system has to identify to produce appropriate responses. The user may indeed have ulterior goals and plans which are outside the scope of the application model, for example the putative train traveller wants to visit a loved one, so there is a top-level unexplained user property "wants to know (for ulterior reason)". But otherwise, seeking to identify the user's beliefs, desires, goals, intentions or plans implies identifying those that would be within the scope of the application model, i.e. beliefs, goals etc that are explicable even if false or mistaken (e.g. Pollack 1987). For example in a train inquiry system, that if an inquiry is about trains from X to Y, a possible user goal is to go by train from X to Y. Similarly, with something like UC (Wilensky et al 1988), if the user asks about deleting files, their goal may be one of the things that can be achieved, within Unix, by file deletion, including removing this specific file, making space, deleting old versions etc. (Of course characterising removing a file in terms of these is the hard part of writing the application model.) Moreover, as the application model covers the nature of the task as well as the domain, so that for an inquiry system the generic requirement for the user is to get information and the generic purpose of the system is to supply it, if the user offers information, this is interpreted as a means to this end.

Thus for systems with a human patient, the user's specific given or inferred decision properties constitute the user model required whether the system has a global task (e.g. diagnosis) or is simply responding to whatever local requirement the user's input sets, and which may be exploited to serve acceptability and efficiency as well as effectiveness. The user's inputs are clearly the prime source of information for this decision model. The role of the system's generic application model is in explaining and filling out the user's inputs through the connectivity the domain element supplies and the motivational base the system's task specification offers. But this connectivity is very valuable: the problem in any attempt to use non-decision properties for modelling is that (as illustrated in Sparck Jones 1989) it may be hard to pick up potentially useful properties because there is no recognisable relation between them and others.

A patient model can thus be an individual model since in building a decision model it is legitimate for the system to hypothesise an individual. How far a human patient model should be described as a model of a person is partly a matter of the nature of the decision property descriptions the system allows, and partly a philosophical issue. But the important point is that the system can model on the basis that it is dealing with a single autonomous entity because decision properties are supposed to connect in individuals. (With several patients things are more more complicated, but the basic point holds.) At the same time the system can motivate its specific model, and establish connections between patient properties, through its application (i.e. domain and task) knowledge. This provides the reasons for observed or hypothesised property cooccurrence. The application connectivity can in particular allow, to the extent that this is appropriate for the system context, for the patient's subjective or mental states and attitudes. These points apply whether the patient model is used for its primary purpose, namely effectiveness, or for other purposes like efficiency and acceptability.

They also apply even if the system exploits stereotypes as starting points for decision modelling as considered in Rich 1989, for example. The notion of stereotype has been rather variously interpreted, and stereotypes have been used in different ways. I shall use 'stereotype' here more narrowly than Rich, to refer to fixed user characterisations which are not open to individuation through property value choices or through combination, and so are likely to be of a crude sort covering large classes of individual user. Stereotypes in this sense can still refer to decision or non-decision properties, and so to patient or agent, and can be partial or complete characterisations of the user. Stereotypical properties when combined with user input providing an individual characterisation on other properties can provide an overall individual model, as in Morik 1989 or Chin 1989, but the effect where stereoptype properties impinge may still be rather crude because the stereotypes themselves are crude. However in the present context the important point is that while using stereotypes in the broader sense of Rich is really just one strategy for constructing individual user models, stereotypes in my sense are built into the system specification and individuation is on other bases. (This does not of course imply that these stereotypes cannot be helpful, only that rather different assumptions about user modelling are being made from those in question here.)

My argument is thus that with patients the system can engage in serious modelling even though the specific information it gets about the user is through his linguistic inputs. The user's inputs are primary, and the system can only work from these. But assuming the system has a coherent application model, it can then go further in building the patient's decision-property model. This support is not necessarily provided in a similar way for non-decision properties.

Giving primacy to the user's inputs, because the system is dependent on them, suggests

that it is not going too far to say the user is his inputs, and in the present context his linguistic inputs: what he says and how he says it. But what does this imply for the elements of user inputs that do not bear on the patient's decision properties; and what happens to the agent role, whether the agent is the same as the patient or is independent, or where there is no human patient?

The most straightforward view of linguistic inputs is that they are just linguistic inputs; so where they do not say anything (directly or indirectly) about decision properties they are not saying anything material about the user as a person. The non-decision properties of linguistic inputs are thus just properties of a participant in the interaction as such, and have no more personal individuality, as opposed to discourse individuality, than this. [Though this does not imply the discourse model is the user model; see also Sparck Jones 1988 and other papers in that issue of *Computational Linguistics*.]

Thus as far as non-decision information is concerned, the user has no more personality than that given by the discourse itself. But this in turn has implications for agents. Where there is only an agent, and a non-human patient, the agent is simply the discourse participant, with no properties beyond his linguistic inputs apart from the decision information they convey. Where there is a human patient and the agent is the patient, the distinction between patient and agent roles beyond the patient's decision properties is submerged in the characteristics of the discourse participant. Where there is a distinct agent, the only viable position is that the properties of the discourse, unless they are explicitly signalled as the patient's, are all the agent's, so the patient has nothing but decision properties.

These conclusions do not necessarily imply that the linguistic inputs cannot convey factual or substantive information about the non-decision, and even objective as well as subjective, properties of the user. They may do this. The point is rather that where this information is not manifestly referable to the patient, it must be assumed to be about the agent; and that in either case while it may be exploited by the system, it need not be, and as it is non-decision information it can should only be exploited with caution because it is not ordinarily connected in a well-founded way with other non-decision, or indeed decision, information.

In practice, whatever the patient/agent relationship (since this could apply even where patient and agent are the same), it may be useful to build in an agent type model or stereotype or, if the application makes it appropriate, to allow for several types, perhaps with explicit checking for the relevant type at the beginning of the session. Agent typing of this sort would be used in a uniform way, for example to justify an assumed level of domain knowledge, with individuation only as called for by the linguistic input. (As noted in previous papers, an agent model of this kind may be important for system efficiency as well as effectiveness.) Thus though there is an independent agent model, it is strictly limited in relation to the modelling aims and assumptions with which I started, and so does not materially undermine the proposition that user modelling can be based on decision properties plus the user's inputs, and indeed should be, to avoid error. Indeed managing an agent type model of this kind could be regarded simply as part of the system's task specification. Certainly it is hard to imagine a system without some default assumptions about the agent built in, for example (in systems of the kind illustrated) that he knows English, is an adult, is not playing games. But the effect on interaction is uniform and is not particular to the individual user.

More specifically, as realism suggests one cannot generally look for any underlying non-decision characterisation of the agent, the agent can be no more than what his linguistic inputs - in their syntactic, semantic and pragmatic form and content - add up to; and he is an individual only to the extent that specific discourse is individual. The agent model is

what the linguistic discourse is above any information, and especially decision information, attributable to the patient, or what can be inferred from this discourse on general grounds, e.g. about any human, or via a task stereotype. These general bases for characterising discourse participants indeed allow for all the broader pragmatic information associated with speech acts and communicative intentions and so, for example, for the use of metaplanning knowledge in the style of Litman 1985. So when these generic properties of language use are exploited in conjunction with the specific information supplied by the content and form of the discourse, it may be possible to build a relatively rich, discourse-motivated model of the agent (e.g. deriving information about the lawyer from "The stupid fellow says ..."). But we can only consider, even in principle, information about the agent beyond this if it is given explicitly, since the system cannot be expected to have more general knowledge to support inferences on additional information; and even then it may be impossible to make any use of this explicit non-decision information about the user (e.g. the lawyer saying "As a bachelor I think ..."), because it cannot be related to any of the system's functions. The fact that non-decision information about the agent may rarely figure in inputs in practice is a separate matter.

The conclusion, therefore, is that because it is difficult to get a hook on any user properties other than decision ones and those that are expressed as the linguistic discourse properties of user inputs, a user model subsumes, where this is appropriate, a patient model defined by decision properties, and an agent model defined by the discourse properties of the system's linguistic inputs. Overall, modelling the user means modelling the user as, and only as, a user of the system, and considering only those of his properties that are relevant to this. Restricting modelling in this way still allows for the user's subjective properties, and especially his dynamic mental ones, i.e. for treating the user as a thinking and purposive agent; but the scope of the user's activities as an agent in this general sense of "reasoning agent" is defined by the system. This means not only that his specific beliefs, goals etc are limited to ones relevant to the system, but that the notion of belief, goal and so forth are being given a rather shallow, behaviouristically- oriented interpretation in terms of the way users use systems, and specifically a linguistically-oriented interpretation in terms of their expression in the user's inputs. The system's inferences about what lies behind the user's input, where they are not exploiting his previous inputs, are based on the system's application knowledge and are constrained by the limits of this knowledge and by the practical limits on its ability to test and extend its model through interaction with the user. In an important sense the user is his inputs.

But what does all this imply for generation?

5   Generation

As the user model is what the user says and how he says it, insofar as generation should exploit a user model, this is what it has to exploit. Where there is no human patient, the content and form of a generated output may not be determined solely by the need to respond to the user, in this case the agent; and even where there is a human patient, a spontaneous output, though motivated by a user model, need not be a response to a prompting user input. But where the system is generating output in response to a user input, in a single shot or dialogue, all it really has for the specific user is the linguistic input representing the user as agent, plus any patient model for the application. [I am assuming that if the system obtains

any non-decision information about the patient only, this will be incidental in every sense.]

Of course if the patient and agent are identical, the system has a richer model to support its interaction with the user in his agent role. A system may have independent agent stereotypes which can influence output as in Cohen and Jones 1989, where explanations about the patient's educational performance are tailored to types of agent. (In Paris 1989 the choice of explanation is rather different as it can be viewed as a system decision for the patient user who is also the agent.) A system like Cohen and Jones' can treat agents as very important - they perhaps verge on being patients - but they are still treated primarily through the dialogue linguistic evidence, and the system has little independent basis for modelling them as individuals. Agent stereotypes are part of the built-in system task specification, and however valuable, do not affect my general arguments for realism in constructing and using individual user models; more generally, while there may be very good reasons for providing for more sophisticated, independent agent models through a connectible property set, without this explicit provision allowing an agent analogue of decision- property operations on the patient, the default for agents has to be that the system has no information about them beyond what comes directly from their behaviour as discourse participants.

But it is clear that many of the constraints on an output generated as a response to an input are motivated by principles governing discourse, whether more narrowly as in adherence to structural rules, for example constraining topicalisation, or in adherence to Gricean principles which in this case would be primarily based on the system's view of the user, for example of the state of his knowledge of the problem domain, as this has been derived from his previous inputs. Thus it would be the session data, which includes the system's own previous output, and the view of the user's state extracted from this, which would determine what constituted unnecessary repetition, for example. Aspects of the generated output not motivated by application needs are therefore determined by the linguistic characteristics of inputs, but these bear on discourse coherence at the textual level, and on cooperative behaviour of the Gricean sort between dialogue participants, i.e. rest on generic considerations applied to the specific discourse and not on anything to do with the originator of the linguistic inputs as persons in their own right.

Generation as a whole is determined by the system's application need to know decision information, and the system's discourse need to maintain communicative interaction. What is an appropriate system output to generate at any point thus depends on, i.e. is the result of the interaction between, the system's application state, i.e. its decision-making state in relation to its global or local task model, and its linguistic discourse state, the significant point being that while the former, with a human patient, may in a real sense represent a person, the latter should not be seen as representing any more than what the linguistic properties of the discourse themselves express, i.e. as representing a discourse participant and not an individual person. Generation, in relation to users, is driven by a mix of person model for patients and a (session) participant model for agents.

All of this leads to generation as illustrated in the following example.

Example 6

The (notional) system here is designed to help someone wanting scientific literature by formulating a proper characterisation of what they want, which will then be submitted to a search service. The patient and agent are therefore one. The example is also for more

than a straightforward advisory case or simple action- achieving situation in Morik's sense, since the system has to engage in complex decision-making in its own right, exploiting technical knowledge the user cannot be expected to possess, to follow up the user's needs with literature-searching actions; and I am assuming the system makes use of a range of domain and task knowledge sources relevant to different aspects of this type of literature search situation (Belkin, Seeger and Wersig 1983; Brooks 1986; Daniels 1987). [Note that these authors interpret "user model" more narrowly than I am doing. Daniels' user model characterises static, essentially objective, decision properties of the patient providing a context for his actual literature need.] The system outputs in the dialogue which follows are annotated to indicate, summarily, the respective contributions made to them by the application model and by linguistic discourse mechanisms assumed to cover broader pragmatic aspects of a Gricean sort as well as narrower matters of expression, both within and between sentences. The application contributions to the system output can be either seeking or supplying application information. In the annotations A = indicates application contributions, L = linguistic contributions.

```
U1  : I'm looking for literature on prefabricated housing.
S1  : What are you looking for it for?
  A : intended use of literature
  L : question form including anaphors and word "looking"
U2  : I'm doing a project for my geography course.
S2  : So you're a student?
  A : appropriate literature level
  L : question form
U3  : Yes. I've looked at a couple of books Dr Smith recommended.
S3  : What were they? Did you find them helpful?
  A : literature coverage; topic specification
  L : question forms; anaphors
U4  : I found Robin's 'Housing for all' very interesting.
S4  : Why?
  A : topic specification
  L : question form
U5  : It showed all kinds of houses I'd never thought of as
      prefabricated are really, like Zambian huts made with regular
      poles and ordinary market grass mats which come in standard sizes.
S5  : So you don't just want stuff on Western factory made units,
      but on any kinds of houses made anywhere out of existing elements
      bigger than bricks?
  A : topic specification
  L : sentence structure, lexical choice etc
```

Imagining now a later stage in this dialogue, the system outputs which follow show how the patient model built up over the session may be applied, in a manner which is also natural from a linguistic point of view in relation to the dialogue history, and how the system's application knowledge can be exploited for cooperative response to avoid possible misunderstanding by the

user. [As Steve Draper has pointed out to me, I have not concerned myself in this example, or the previous ones, with the manner of the user's adaptation to the system, though adaptation is clearly if informally assumed in the putatively natural dialogue. The extent to which a user's conscious adaptation can itself assist the system's modelling is a matter for further investigation.]

```
S21 : Do you want weekly trade magazines?
U21 : Why should I?
S22 : Their technical specification sections ads give a lot
      of detail. But perhaps that would be more than you want.
  A : refer to purpose (mentioned far back)
U22 : Are there a lot of them?
S23 : I don't think so, but the system's coverage could be poor.
  A : give domain knowledge
  L : cooperativity
```

The illustrations in Sparck Jones 1984 and 1989 showed how the system's outputs could be motivated by modelling serving the different functions mentioned, for example to acquire more information in order to promote effectiveness, and to express the request in a particular way, to promote acceptability. During a dialogue the system could produce outputs exploiting modelling designed both the serve functions directly themselves and to obtain new information supporting the development and future application of its models. The illustrative dialogues also showed how a single output could serve several functions at once. The more restricted person and participant models for patient and agent respectively considered in this paper may also be exploited to serve the different system functions, effectiveness, efficiency, and acceptability, in the varied ways illustrated in Sparck Jones 1984 and 1989. So, for example, even the restricted agent model may be exploited for efficiency, or acceptability to patient or agent, the restricted patient model for acceptability to the patient as well as effectiveness.

These functional distinctions are moreover important even in generation in advisory situations (especially where patient and agent are one) they appear difficult to maintain or apply: thus if we subsume communicating the system's decision under effectiveness for the consultation case, what constitutes proper communication? The advice to be given is the system's decision, but it may be hard to separate the advice given from the way it is given, though presentation appears to be an acceptability issue and not a decision matter. For example, deciding what information to select for output from a relevant pool can be viewed as a decision choice or a presentation choice. But it depends on the criterion used to select, which is determined by what the system task specification is in relation to what constitutes advice, which suggests there is still a real distinction between taking a decision and expressing it for the user for any particular system, or in any particular case.

On the other hand, if communicating a decision implies communicating it properly, i.e. effectively, it is more difficult to draw the line clearly, and while preempting false inferences (Joshi et al 1984), or overanswering (Morik 1989), may appear to have to do with effectiveness, where choices of speech act (Morik 1989) or of lexical item may seem to be matters of acceptability, these cannot be general rules, so each instance would have to be categorised independently. It could on the other hand be claimed that if preempting or overanswering

17

are motivated by Gricean principle, as in Joshi et al and Morik, this follows from the generic system task specification and not from the specific decision, and so has more to do with effectiveness in a global sense than in the sense relating to particular decision making in which I have defined it. All of this is a tricky issue which needs further analysis; but I shall assume here that it is reasonable and possible to make a working distinction between effectiveness and acceptability in generating output, even in communications about or of system decisions. However in the advisory context, unlike other expert system ones, decisions may be final or intermediate, and definite or provisional; and effectiveness is of course also a motive for system outputs designed to gather information as well as to provide it. This all makes the functional status of system outputs very complex.

Considering now the use just of application and linguistic resources as illustrated in the Example 6 literature search dialogue, it is clear that the system outputs can serve various functions, and also for example that application knowledge can serve different functions. If communicating a decision at least in some neutral, information-giving sense is subsumed under effectiveness, then linguistic knowledge can contribute to more than just making the system's behaviour acceptable. This would fit the idea of explanation generation, for instance as illustrated in Paris 1989, as a seamless web with subject content and linguistic form all stemming from the system's decision about what explanation to give. But in other cases linguistic knowledge is deployed in a tactical manner which may be more reasonably described as serving acceptable dialogue behaviour. It could also serve efficiency through economical and acurate information gathering.

Thus in Example 6 (but as before concentrating only on effectiveness and acceptablility and disregarding efficiency), the system output S5 - about the user not wanting only Western prefabrication - while it is essentially motivated by a concern with effectiveness both in communicating a local decision and gathering information for an ultimate one, can also be seen to be using application knowledge to promote acceptability by showing the user the kinds of distinctions and subject specifications the system is capable of considering in relation to the particular consultation topic. [It might also be taken as seeking to promote confidence in the system as a whole.] S22 - about the technical advertisement detail - shows a similar combined concern using application knowledge for acceptability as well as effectiveness through the explanatory information associated with the reference to the search purpose. S23, on the other hand, on the system's coverage, is simply an acceptability-oriented explanation.

The linguistic contributions on the other hand are equally dual purpose in many cases since they are partly serving effectiveness, as communications about decision information, and partly acceptability: maintaining discourse coherence works both ways, for example. The linguistic form of S3 - asking the user about the books read - is designed both to extend the existing model information, i.e. is effectiveness motivated, and also to promote acceptability, through its careful form. Similarly S5, on the building types, by opening with "So", is exhibiting the system's consequential response both as a matter of content and of conversational linking.

6 Conclusion

This discussion has been a reaction to what I believe is the assumed, if not stated, long-term goal in natural language interface design, which has been encouraged by a concern with advisory systems, namely that the aim is personalisation. My conclusion is that the elements of

modelling are those I identified before: it is necessary to distinguish user roles, user property types, and modelling functions; but also, reiterating my conclusion in Sparck Jones 1989, that modelling itself should be approached simply, partly because information about the user is not available but also, more importantly, because fancy modelling chasing the real person is unnecessary. My view is that personalisation in a strong sense is a misconceived aim. Seeking to tailor the system's generated output ever more tightly to an ever more refined characterisation of the individual, and especially to one going beyond what is required for decision effectiveness, is a mistake, for two reasons. The first is that in human interaction many socially relevant properties of those involved are recognised as important and are exploited, notably supposed class, status, authority and nous; but where these are not decision properties, systems are unlikely to be able to perceive them. At the same time, many ordinary human interactions, as in shops, do not bother about properties of the user as an individual behind what he says (and does) and what it is necessary to know for the specific task in hand, and at most take note in addition of stereotypical properties like sex.

Thus, for example, if sex is a decision property it has to be and can normally be obtained; and it may then be exploited for non-decision functions like acceptability; if it happens, though not a decision-property, to be known, it may be usefully exploited; but if it is not known, general politeness in generated output will be an adequate substitute.

In other words, there are too many constraints on man-machine interaction which are not just those imposed by our current system-building limitations, for it to be rational to seek model- based generation beyond what in fact suffices for many ordinary human situations. Thus what we are actually doing now, which is modelling confined to decision properties and discourse information, which is sometimes seen as an inferior substitute for the real thing, should instead be seen as an appropriate strategy in its own right, and not as a way station to something better. Though this of course is not implying that model-based generation, even if limited to decision information and discourse pragmatics, is easy.

References

Belkin, N.J., Seeger, T. and Wersig, G. 'Distributed expert problem treatment as a model for information system analysis and design', *Journal of Information Science* 5, 1983, 153-167.

Brooks, H.M. *An intelligent interface for document retrieval systems: developing the problem description and retrieval strategy components*, PhD Thesis, City University, London, 1986.

Carbonell, J. et al 'The XCALIBUR project: a natural language interface to expert systems', *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, 1983, 653-656.

Chin, D.M. 'KNOME: modelling what the user knows in UC', in Kobsa and Wahlster (1989).

Cohen, R. and Jones, M. 'Incorporating user models into expert systems for educational diagnosis', in Kobsa and Wahlster (1989).

Daniels, P.J. *Developing the user modelling function of an intelligent interface to document retrieval systems*, PhD Thesis, City University, London, 1987.

Grosz, B.J. and Sidner, C.L. 'Assertions, intentions, and the structure of discourse', *Computational Linguistics* 12, 1986, 175-204.

Joshi, A., Webber, B. and Weischedel, R. 'Preventing false inferences', *Proceedings of*

COLING84, 1984, 134-138.

Kobsa, A. and Wahlster, W. (eds) *User models in dialogue systems*, Berlin: Springer, 1989.

Litman, D.J. *Plan recognition and discourse analysis: an integrated approach to understanding dialogues*, PhD Thesis, TR 170, Department of Computer Science, University of Rochester, 1985.

Morik, K. 'User models and conversational settings: modelling the user's wants', in Kobsa and Wahlster (1989).

Paris, C.L. 'Tailoring object descriptions to a user's level of expertise', in Kobsa and Wahlster (1989).

Pollack, M.E. *Inferring domain plans in question answering*, Technical Note 403, SRI International, Menlo Park, 1986.

Rich, E. 'Stereotypes and user modelling', in Kobsa and Wahlster (1989).

Sparck Jones, K. *User models and expert systems*, Technical Report 61, Computer Laboratory, University of Cambridge, 1984.

Sparck Jones, K. 'User models, discourse models, and some others', *Computational Linguistics* 14, 1988, 98-100.

Sparck Jones, K. 'Realism about user modelling', in Kobsa and Wahlster (1989); also Technical Report 111, Computer Laboratory, University of Cambridge, 1987

Webber, B.L. 'Questions, answers and responses: interacting with knowledge base systems', in *On knowledge base systems* (ed Brodie and Mylopoulos), Berlin: Springer, 1986.

Wilensky, R. et al 'The Berkeley UNIX Consultant project', *Computational Linguistics* 14, 1988, 35-84.