Adaptation to affective factors: architectural impacts for natural language generation and dialogue

John A. Bateman
Faculty of Linguistics and Literary Science &
Sonderforschungsbereich/Transregio SFB/TR8
'Spatial Cognition'
University of Bremen, Germany
bateman@uni-bremen.de

Cécile L. Paris
CSIRO - ICT Centre
Locked Bag 17, North Ryde
NSW 1670, Australia
cecile.paris@csiro.au

ABSTRACT

We are currently developing automatic natural language dialogue capabilities for assistence systems where the perceived 'naturalness' and ease of use for the systems' users is a central consideration. We are, in addition, aiming at generic solutions that will enable a broad range of users to be supported. Our applications focus on issues that arise in the spatial domain—that is, our assistence systems are concerned with shared user-system tasks that are to be carried out in both shared and non-shared spatial configurations. The potential for miscommunication in these kinds of situations is extremely high since the perception and representation of space by our systems and by their human users is generally quite different; nevertheless, the interaction between user and system must be made to work even in the face of mutual non-comprehension. This potentially frustrating situation must be explicitly recognised and catered for in order to keep the system and user in an effective working relationship, thus raising issues of producing language with appropriate affect and of recognising the user's affective state. In this paper, we present how this capability is being added into our dialogue architecture as an integral part of the mechanisms for dialogue support in general. We relate this to existing mechanisms for affective language generation/interpretation, user-adaptability and the importance of maintaining a tight inter-relationship between language and situation.

1. INTRODUCTION: CONTEXT OF OUR WORK

We have recently started some cooperative work which is investigating how generated language can be made maximally appropriate for the users of computational dialogue systems by employing generic mechanisms of adaptation. Our aim is to construct dialogue systems capable of actively supporting a broad range of users with varying levels of linguistic, perceptual and behavioural skills by means of natural and inobtrusive natural language interaction. As one of our application domains, we are building a dialogue system that allows a user and an autonomous wheelchair (the Bremen 'Rolland' developed by the Informatics department of the University of Bremen [26]) to interact via natural language concerning spatial tasks, such as finding particular rooms, jointly constructing maps of the (indoor) environment, etc. [39].

Robotic agents of this kind provide a particularly challenging set of requirements. First, the perceptual systems of the robotic agent and its human users are radically different. This means that there is a crucial need for flexible mediation between the system and its user in order to negotiate workable common terms and constructs. Second, there is the expectation that system and user will be interacting over a prolonged period: therefore the agent must function as what has been described a *Relational Agent* [7], i.e., one which maintains a 'long-term socio-emotional relationship' with its user(s). Since it has now been established that users respond very favourably to systems which actively deal with sources of frustration or difficulty in their interactions (see, for example, the results cited by Liu and Picard [28]), it becomes crucial that the high potential for miscommunication present in scenarios such as our autonomous wheelchair application be appropriately channelled so as not to compromise the effectiveness of the user-system relationship. In interactions of the kind explored here, our empirical investigations to date have shown that affective issues are constantly arising; mismatches between the expectations of the user and the abilities of the computational device lead to dialogic situations where it is important that the interaction be maintained even in the face of communication breakdowns (cf., [15, 16, 5]). We now see highly adaptive natural language dialogue as one means of securing this safety channel.

This broad appreciation of the possible role of language has now been taken in several systems (e.g., [14, 36, 28]). In our own work, we are focusing particularly on the architectural design implications that this degree of adaptivity brings for natural language components. We suggest that it is necessary to go substantially further than classifying an computational agent's dialogic turns as being in need of demonstrating particular emotional dispositions rather than others (e.g., showing 'sympathy'/'empathy' when giving bad news). We need in addition to be constantly monitoring both users' utterances and system responses concerning their affective 'commitments' in order to recognise trouble as soon as possible (cf., [16, 5]). That is, we seek ultimately to recognise the linguistic indications of particular task-relevant and interaction-relevant affective states and to respond to these appropriately, all within a generic dialogue architecture.

As our work is only in its very early stages, we will concentrate in this paper on exactly how we are attempting to con-

struct the fine-level monitoring of affect and its consequences for language within a generic dialogue architecture as an intrinsic component. We currently see affective adaptation and its utilisation as a necessary consequence of our overall architectural decisions rather than as some non-essential 'add-on' and this raises a number of issues and design goals generally relevant for discussion in this area. Our starting point is therefore a reappraisal of 'stylistic' natural language generation. While generating under 'stylistic' constraints has been the focus of a number of generation projects in the recent past (e.g., [33], the main use of style was to characterise the language of a particular domain or text types (e.g., labels of pharmaceutical products). We believe that it is equally important to produce text appropriate to the *affective state* of a user, thus adding a further dimension to the kinds of 'style' at issue.

In order to achieve this, we have been led to consider issues of affective language control from three main perspectives. First, although language does not provide direct access to emotional states, it does provide a rich variety of clues that indicate an 'abstract emotional disposition' as a connotation of language use. These clues are now receiving increasing study; work in this area blends into work on style, affective style and subjective language in general [13, 42]. Second, we can see these clues as another example of the highly systematic and coded relationship between language and situational context that has received considerable study under the guise of **register**; this has been explored functionally by, for example, Gregory, Ure, Matthiessen, Martin and many others [21, 40, 31, 30]. By maintaining a close link between the mechanisms that we propose and existing linguistic work that addresses the relationship between language and its contexts of use, we believe that the computational mechanisms developed can be made to have the widest possible scope of application as well as being able to draw on a considerable body of previous linguistic results. Third and finally, we require that our automatic dialogic partner be highly adaptive—that is, the wheelchair is expected to conform as far as is possible to the language requirements of the user. Affective adaptation is then just one aspect of a more general treatment of adaptation, and we are seeking to ensure that it finds a natural place within the generic mechanisms for adaptation that we are developing.

2. SOME DEFINITIONS

There are a broad range of conditioning influences that shape the appearance of real texts in context over and above the bare bones of what is to be communicated. These variations have long been dealt with as aspects of 'style' [24]. However, since 'style' remains a very broad term, it has not as yet led to particular mechanisms for its coordinated control. This is partly due to the very heterogeneous nature of the kinds of phenomena that it has been made to include. To produce appropriate text, we prefer to exploit the theoretical construct of **register** to the less formal notion of style. By adopting the functional interpretation of register rather than the more quantitative definitions suggested in work such as that of Biber [6], we are able to define more rigorously both the dimensions of variation relevant for identifying specific situations that must be taken into consideration when generating language and the specific consequences of these dimensions for the language that occurs.[1] Style, in contrast, refers to a broad range of linguistic phenomena, and it is often unclear what is covered by it.

Register has been studied at length in linguistics, and we can thus draw on a long history of previous work. Essentially register is a theoretical construct that characterises a situation together with its consequence on language in terms of 3 dimensions: ideational (the field of discourse), textual (the mode of discourse) and interpersonal (the tenor of discourse, related to the participants and their relationship). While a register can affect all levels of the linguistic system, we restrict our focus here to the consequences of a register choice at the lexico-grammatical level (i.e., the phrasing level). The functional view of register allows us to position quite precisely aspects of language use that are to be included under particularly 'affective' factors of text production. Affective factors fall under the **tenor** of discourse, as it is this area that is particularly concerned with interpersonal relationships, evaluation and appraisal—all areas that are particularly vulnerable to affective variation.

According to register theory, language is taken to co-vary with situational factors. Affective language use can therefore be said to co-vary with tenor. We start, then, by characterising tenor along the following three standard dimensions proposed for this dimension of situations:

- Status (equal, unequal)
- Contact (involved, distant)
- Affect (marked: positive/negative, or unmarked)

With each value of these dimensions one can identify consequences of the language that realises it and so can both constrain appropriate language production during generation and recognise a speakers construal of the situation during understanding—this therefore contributes to understanding a users affective state during an interaction and guide the production of language with appropriate affect. While a range of linguistic consequences in this area are already known and have been classified according to the registerial 'values' of tenor (cf. [29]), there is much still to investigate empirically in order to improve our understanding further. In this latter case, the setting out of the kinds of variation possible serves as a valuable framework for integrating empirical results as they become available and for guiding the formulation of new hypotheses for precisely targetted empirical investigation.

The next definitional issue with which we are concerned is that of 'adaptation' itself. Here we choose to organise approaches to adaptation along a continuum of responsiveness. At one extreme, and as implemented in computational interactive systems increasingly frequently, adaptation means that the style of interaction between a system and its user can vary according to a broad classification of the type of user involved. Traditionally in this area we find distinctions such as those between 'experts' and 'novices' in various areas [35]—such as, for example, between doctors and patients

[1] An excellent review of the alternative positions and treatments of the complex area of 'register' is given by Lee [27].

in healthcare situations [10]—or according to 'interest profiles' [34, 12]; broad settings of such profiles for website information presentation are probably the most frequently seen realisations of this style of adaptivity at present. Most sophisticated systems involving dialogic interaction and user modelling already move away from the *a priori* setting of user levels, however, in order to accomodate the *changing* state of a user during an interaction. Most commonly this involves the marking of particular pieces of information as 'known' (or at least 'already presented') so that in subsequent information offerings a modified form of presentation can be adopted. This shows adaptation over the course of a single interaction between system and user.

Research on natural linguistic interaction strongly suggests that this latter kind of adaptability is a pervasive aspect of successful communication. Speakers respond to previous utterances in an interaction to demonstrate how they have understood what has been said and automatically take on features exhibited previously in the discourse. We consider this facet of natural language use to be turn-by-turn micro adaptation on the part of speakers. In essence, each turn in a dialogue is 'adapting' to previous turns.

Work in the tradition of Garrod, Sanford, Anderson and others [18, 19] has long argued for such fine-scale co-ordination of speaker turns during interaction. For example, and particularly relevant for our application scenarios, particular conceptualisations of spatial situations are negotiated, ratified or rejected during dialogue and form an essential basis for the interactive achievement of shared meaning. This work has recently been taken considerably further in its implications for dialogic interaction by the 'alignment model' proposed by Garrod and Pickering [17, 37]. Pickering and Garrod [37] argue that successful dialogic interaction is a result of an *automatic process* of converging situational representations of speaker and hearer: this process, termed **alignment**, is brought about by an automatic convergence, or alignment, occuring at other levels of representation (e.g., lexical, syntactic, semantic) during interaction. This convergence, in turn, causes the linked situational representations to converge. The main psychological mechanism posited for driving alignment at all levels is priming, whereby the production of one form by one speaker makes it more likely that such a form be produced by another speaker.

In our work we are currently exploring a closer relationship between the alignment proposal and register negotiation between speaker and hearer. This direction is also suggested by Pickering and Garrod's remarks concerning the relationship between previous psychological research and generative syntax—and, in particular, approaches that prioritise decontextualised sentences. Pickering and Garrod argue that prioritising decontextualised sentences has made it more difficult for theoretical accounts to see the natural processes of alignment by which dialogue functions. From our perspective, however, since the functional view of register adopted is drawn from a linguistic orientation which insists on the centrality of relating use of language to context [22], it becomes more natural to consider possible interconnections between its linguistic models and Pickering and Garrod's proposed architecture.

Much of the theoretical foundation for such a consideration has already been considered in Bateman [1], where the relationship between fine-scale adaptation and the functional view of register was formalised in terms of **micro-registers**. The essential idea here is that just as register describes the relationship between broad situations and the language that occurs in those situations, so does micro-register describe the additional relations that occur *within* a situation, particularly a dialogic situation, as that dialogue/situation unfolds in time. Register is seen pre-formally as a co-variation of situational features (drawn from the areas of field, tenor and mode) with lexico-grammatical features. Micro-register sees further the relationships imposed by register as underspecified sets of constraints that are then filled in and stabilised in particular ways during any given interaction.

Pickering and Garrod argue that alignment makes both production and comprehension easier by providing 'routines' that, once negotiated between speaker and hearer, can be employed without the need for further decision-making. This ability is achieved by aligning linguistic levels of representation so that both speaker and hearer can rely on common communicative strategies having a particular effect over a particular stretch of interaction. Registers, on the other hand, have traditionally been associated with particular patterns of linguistic features or constructions co-occuring more frequently in some situations than others. Language produced within one register will show these co-occurence patterns more than language produced within other registers. Now, when we add micro-registers to this picture, we can see that the features that are to co-occur become sufficiently specific as to include particular (partial) constructions, lexical selections and semantic strategies. This is precisely the same effect of making available shared 'routines' as proposed for alignment. The difference between register and micro-register is then largely one of specificity.[2]

As a brief example of these concepts, we can adapt one of the examples given by Pickering and Garrod. If one speaker in a dialogue uses the phrase "the sheep that's red" rather than "the red sheep" to assign a colour to some sheep under discussion, then alignment predicts that, via priming, the other speaker will subsequently be more likely to use the first strategy rather than the second, too. Within the semantic formalism that we employ, the intended meaning for these alternatives has a common representation:[3]

```
(s / sheep
   :property-ascription (r / (color red)))
```

Then, within our linguistic model and the description of lexico-grammar employed (essentially systemic-functional grammar as set out in Halliday and Matthiessen [23] and

[2]Given that within our linguistic model the levels of representation are already in a strong inter-relationship of *realization*, we also predict that alignment at any level will bring about alignment at other levels, just as Pickering and Garrod propose. Further discussion of this is, however, beyond the scope of the current paper.

[3]This semantic representation is based on the sentence planning language (SPL) originally defined by Kasper [25], and subsequently used in several natural language generation systems.

semantics	*lexicogrammar*

`(s / animal`
`:property-ascription`
`(r / colour ))`
	`post-modification`

Table 1: A simple microregisterial setting that pairs an underspecfied semantic expression with a grammatical constraint

described computationally for natural language generation in Matthiessen and Bateman [32]), we can characterise the production of an associated utterance as follows.

If we do not provide any further constraints, then both of the possible utterances above (and several others) can be generated with our English grammar. However a selection between these can be forced (in this case) by the choice between contrasting grammatical features: for example, somewhat simplified for the purposes of discussion, 'pre-modification' *vs.* 'post-modification'. By default the grammar tries to make a sensible choice between these on the basis of how much semantic material is to fit in the property ascription (e.g., 'the red sheep' *vs.* 'the sheep that used to be red every other day'), but we can also choose to *pre-select* the relevant feature in advance. Such pre-selection has precisely the effect of priming for one construction rather than another that Pickering and Garrod associate with alignment. This, then, is a minimal micro-register: we can state that the production of this grammatical form primes for the actually selected lexico-grammatical features rather than those that would in principal be possible but which were not selected.

This can be made arbitrarily more complex. The actual example given by Pickering and Garrod draws on experimental results from Cleland and Pickering [11] which showed that the priming effect was much stronger when ascribing a colour to a *semantically similar entity.* That is, "the sheep that's red" was produced far more often after hearing "the goat that's red" than it was after hearing "the book that's red". This shows that the micro-register must consist not only of preselected lexico-grammatical features but, instead, of (at least) *pairs* of semantic:lexico-grammatical expressions that are contingently associated during an interaction. The micro-register established in the current case might then be summarised by the pair shown in Table 1.

The exact degree of specificity for the semantic types (i.e., any 'animal' or just 'mammals', any 'colour' or some particular range, etc.) must be ascertained empirically; the basic mechanism for the formation of such locally active 'routines' or micro-registers is, however, relatively clear. We therefore can import accounts of 'partially idiomatic' expressions and fixed phrases (all interpreted as more or less underspecified fragments of syntactic structure) and combine these with our notion of dynamically grown micro-register pairings for tracking spontaneously created routines during dialogue.

The final link to our main theme in the current paper, that of affective adaptability, is then to focus on just one area of linguistic variation: that is, the area of linguistic variation picked out by slicing through the *tenor* component of register and micro-register. Tenor values are taken up and refined, changed, negotiated during an interaction just as has been indicated. Whereas the micro-registerial pairing shown above involves meanings taken from the *field* of'register concerned with colour ascription, we can equally consider such pairings for semantics drawn from the affective areas. This is then to be captured by adaptive/aligning micro-registerial variation in precisely the same way so as to produce text that matches the affective state of the user, or text that reflects a different affect in an attempt to change the affective state of the user.

3. SOME MECHANISMS

Although we are now beginning to see more systems that explicitly take on the problem of controlling 'style' in generation, we suggest that this is still mostly seen as some extra functionality that is grafted on to an existing account. In contrast, we argue that since language must always be produced with respect to a register and that micro-registers will necessarily be established during any interaction, phenomena such as alignment will also necessarily follow.

Several kinds of mechanisms have now been proposed in the literature on affective generation that include approaches to 'style'. We can distinguish, for example, following loosely on the categorisation provided by Paiva and Evans [33], between those approaches to style that rely on the identification of co-occuring linguistic features in texts taken to exhibit stylistic varation and those approaches that see style effects as strategic choices in order to realise non-propositional, including 'affective', communicative goals. The former, Paiva and Evans describe as adopting a 'corpus-based' definition of style. Although this is attractive for several reasons—particularly, for example, in its ability to cover a more or less continuous *space* of stylistic variation—we also see potential drawbacks with the degree of control that can be brought about. Paiva and Evans follow the factor analysis approach of Biber [6] and uncover, on the basis of their corpus analysis, two dimensions for controlling stylistic variation. This can only provide variation with very broad brush strokes, and there is as yet little evidence in favour of reducing style in general to two parameters. That this is not simply a problem of the size of the corpus adopted is shown by Biber's larger study, which still only managed to produce six or seven dimensions of variation with reducing returns as the number of dimensions were increased.

However, the generation mechanism pursued by Paiva and Evans is claimed to be generally applicable. All that is relied upon is that the adopted generation system makes available choice points that control variation with respect to particular inputs; such choice points are then typically places where non-propositionally related grammatical alternatives are covered. Certainly within our own adopted grammars there are very many such choice points, and so we can envisage employing a hybrid approach where strategic control of non-propositional (non-ideational in our terminology) choice points is augmented with selections from a broad stylistic space. This may be a more appropriate model than more discrete control of linguistic features by stylistic parameter values [24, 20].

Among the generation systems that rely on the strategic approach to controlling style, several have recently attempted

to include politeness within their capabilities in a more sophisticated fashion. These systems commonly adopt results from Brown and Levinson [8] and their discussion of the influence of **face** on the communicative strategies employed.[4] In Porayska-Pomsta and Mellish [38], for example, two dimensions for describing face are isolated, autonomy and approval, and these are adopted as the contextual variables which are set in order to guide generation. The actual matching of contextual variables with linguistic choices is managed by a Bayesian Network. Again, we see this as addressing one possible aspect within a more generic approach to relating language to contextual variables: approval and autonomy could be envisaged as particular finer features within the tenor component of register. We thus need to investigate whether the kind of mechanisms employed by Porayska-Pomsta and Mellish might also usefully apply to the other features of tenor.

The approach of Walker *et al.* [41] is broadly similar, although it includes a more interesting view of the function of style as such. Walker *et al.* propose that the particular stylistic choices made are a key aspect of an agent's character. Making such selections appropriately and consistently is therefore an important part of presenting a coherent picture of self. This approach draws on both standard speech act theory and various approaches to social interaction, including again that of Brown and Levinson. They adopt a traditional distinction between illocutionary force and alternative surface realisations (corresponding in our account to tenor and mode features and the diverse grammatical forms that these can be mapped into), augmented by an 'emotional' stance that additionally controls finer variation. As would be expected, some of the particular dimensions of contextual variation employed for selecting between speech act realisations are as in our account: it is well known that the dimensions of social distance and power influence linguistic variation [9, 29]; the breakdown of face into approval and autonomy is also found here. The variation particularly concerned with *affect* is treated orthogonally to the contextual variables. Agents are simply allocated an emotional 'disposition' (e.g., 'angry', 'annoyed', 'sad' and so on) that provides additional constraint on the language produced, particularly involving the acoustic properties of the speech synthesised.

Walker *et al.* then present a *threat function* which evaluates the values of contextual variables for a given situation combined with a ranking of the 'imposition on face' that particular speech acts invoke. This function carries the load of selecting both among the possible grammatical realisations of particular speech acts and among the speech acts themselves. The actual choice of speech act is mediated by the classification of politeness strategies formulated by Brown and Levinson. The lowest values of the threat function lead to direct strategies; the highest values lead to off-record strategies. This provides a sophisticated way of selecting among speech acts during conversational planning and supports contextually appropriate linguistic selection. However, the kinds of realisations associated with particular speech acts appear to be relatively fixed; each speech act type has an associated default syntactic form. Moreover, the kind of ongoing adaptability necessary for genuinely natural dialogue is not yet present.

[4]Note that 'face' can be seen as a crucial abstraction over tenor relationships, one that is primarily concerned with maintaining certain values of interpersonal relationships and evaluation of self.

In contrast to these approaches, we attempt to preserve a broadly strategic functional control of the registerially-conditioned linguistic alternatives while, at the same time, doing justice to the flexibility in expression observed for this phenomenon. We have already investigated several of the mechanisms required to produce language under register-control in our previous work [2, 3]. Building on this, we propose that we need 3 distinct mechanisms in a generation system in order to allow register to effectively control phrasing:

1. the selection of which 'size' (more technically, *rank*) of grammatical unit is to be used for given semantic classes;

2. the construction of a subgrammar, which controls the grammatical options available; and

3. a controlled mapping of instances in the world (i.e, concepts in a domain model) to a linguistic ontology which will guide the grammar during generation.

These mechanisms are quite general and are sufficient for providing a very rich and varied range of linguistic phrasing variation that nevertheless remains under functional control.

We have explored these mechanisms separately in several prototype natural language generation systems. The first two were originally presented in Bateman and Paris [2], while the latter played a central role in the architecture described in Bateman and Teich [4]. We are now building on this work in two ways. First, we combine them all within an integrated architecture; second, we utilise their full potential *during interaction*. It is the second extension that now makes contact with, on the one hand, the interactive alignment model of Pickering and Garrod and, on the other hand, with adaptive language production.

The first example of variable rank semantic-grammar associations we explored involved such cases as logical quantifiers in a semantic representation being mapped to clauses or determiners depending on the tenor variable of expertise. That is, the single formula component:

$$\exists x.\mathrm{signal}'(x)$$

could appear in a surface text as either *there exists a signal* (clause association) or *a signal* (determiner association). We can now relate this directly to the example of adaptive alignment and micro-registers given above. The semantic expression involving the attribution of a color to an entity can be associated either with a clause (giving *that is red*) or with an adjectival group (giving *red* as in *a red sheep*). Formerly we hypothesised that such variation is an essential part of register variation across full registers; now we add to this the possibility that the association be established locally during a single interaction.

This may naturally combine with the selection of a subgrammar. When we have a greater degree of constraint on the grammatical side of the association, we can naturally refer to this as a 'subgrammar'. In the trivial case of the 'red' *vs.* 'that's red' variation we can talk of the former as involving a subgrammar where the postmodification option is not present, and the latter as involving a subgrammar where the premodifying adjectival group is not present. Crucially, this constraint can apply either for entire registers or for particular semantics-grammar associations as developed during interactive alignment. Whereas in the present case the result is almost over-determined, it is not always the case that the grammatical constraints can be expressed so simply. This may lead to more interesting subgrammar selections that involve a variety of feature (pre-)selections and restrictions.

This has the following consequences when applied to affective generation. Alignment refers to the phenomenon whereby a hearer changes his or her speech to 'match' that of the speaker, so that, eventually, both participants in a dialogue will converge. Using our theoretical perspective, alignment can be construed as a person adopting the register of their interlocutor, or the participants co-creating a new micro-register, starting from the registers they individually had at the beginning of the interaction. Computationally, then, this means a system needs to (1) be able to recognise the register of its user and (2) employ the mechanisms introduced above to move smoothly into the new register.

In a dialogue, a number of strategies may be adopted by the system, depending on the situation: for example, the system may choose to align its register to that of the user; alternatively, the system may choose a register different from that of the user in an attempt to make the user change his or her own register. This may be important in affective computing: for example, a system may choose to be very calm and polite when the user is showing signs of anger, in an attempt to smooth the user. It is important that a system be able to make this choice explicitly, reasoning about it and about its potential consequences on the user and the interaction. In our current work, we are planning series of experiments to see in what ways, and under which precise conditions, a user is indeed influenced by a computational systems utterances to the extent of aligning with it.

4. SOME ISSUES

In this section, we briefly suggest some of the issues that arise within our approach and which are in need of further discussion.

First, the approach to register and micro-register we employ clearly places register at a central position in the generation architecture. There has been discussion of whether descriptions in terms of register function causally for generation or whether they are epiphenomenal. We can see from our discussion above that, unless there is some shared representation between interlocutors, then there is little basis to expect alignment to occur. Therefore registers in this sense play a causal role in the generation process.

Second, we see that for interaction all the mechanisms employed need to be active during run-time. The consequences of this need to be explored. Whereas they add to the complexity of the process, it remains to be seen if the corresponding simplification caused by the availability of constrained 'routines' and preselections of mappings and subgrammars provides a counter-balancing reduction in complexity overall.

Third, although we have focused more on language production, we believe that the constructs appealed to are also essential during interpretation. This also follows from the Pickering and Garrod model since they argue strongly that production and interpretation must be closely linked. We see in our model that it is certainly the case that an interpretation process must deliver potential micro-register associations: this is a task over and above that normally expected from analysis components, and the extent to which this is feasible needs to be investigated more closely. We also see clear advantages of a micro-registerial specification for analysis, for example in the provision of very finely-grained subgrammars that might be used to constrain speech recognition on an almost turn-by-turn basis. This also places certain requirements on speech recognition technology: components that are not able to readily change their grammar models will not be able to benefit from these potential sources of constraint.

Fourth and finally, Pickering and Garrod have not addressed particularly the notion of affective alignment. We can see, therefore, how a grounding in the more finely articulated notion of register can lead to clear extensions in the areas of concern to be investigated and, hopefully, subsequently controlled. An area that we will investigate particularly is to what extent the notion of alignment proceeding automatically needs to be modified when we are considering affective language production. A simple alignment model could well lead to an explosive spiral of emotion-laden language: for example, if a speaker's utterance contains many lexico-grammatical clues for a highly agitated and frustrated emotional disposition and the system aligns to this behaviour, its utterances will *also* be highly agitated and frustrated! It is clear here, therefore, that a simple carry over of the alignment model will not be fully adequate for sensible affective adaptation. It must equally be possible to encourage strategies of anti-alignment, or counter-alignment, that can aim at positively influencing the user. The exploration of the deliberate construction of counter-aligning moves in a dialogue is one of our current main areas of concern. This falls under the broad area of **shaping** in the sense of Zoltan-Ford [43].

5. NEXT STEPS

In this paper, we have set out an approach to affective and adaptive language production that draws together results in a variety of areas. In particular, we have argued for some general constructs and mechanisms for controlling the production (and potentially the recognition) of affective language. We have further argued that the interactive alignment model of Pickering and Garrod provides a convincing basis for taking seriously notions of automatic convergence of language behaviour. Moreover, we suggest that Pickering and Garrod were forced to try and reinvent much of an appropriate linguistic account because of their assumption of a "traditional" individualistic, generative account of the linguistic system. In contrast to this position, we have been able to draw on and instantiate linguistically a model

of adaptive behaviour that already commits to much of the inter-individual properties that Pickering and Garrod are at pains to demonstrate. Register and micro-register as essentially inter-individual phenomena provide a natural grounding or linguistic home for the psychological model Pickering and Garrod set out—one that already includes much of the apparatus required for the more complex notions of (partially fixed) routines and shared levels of representation foreseen.

Within this context, we have provided an overview of some of our current work started within a collaborative project. Although quite preliminary, we nevertheless believe that sufficient issues are raised of both theoretical and practical for valuable workshop discussion. Within this framework, we must regard the dialogic contributions of system users as necessarily being indicative of affective states in an abstract sense by virtue of their positioning along the dimensions of tenor. The response of the system must also then determine whether to align or counter-align, thus adapting its interaction strategy according to what will be considered most effective for the continuation of the working relationship between system and user. The system must then be able to control its lexico-grammatical resources to produce the language reflecting the chosen affect. For these reasons, we now believe that affective language adaptation and dialogic alignment will need to play an increasingly important role as affective computing develops in the future.

6. ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR 8 Spatial Cognition — Subproject I1-OntoSpace (http://www.sfbtr8.uni-bremen.de/). We also thank particularly Kerstin Fischer, Thora Tenbrink, and Shi Hui (SFB/TR8, subproject I3) for dialogue-related discussion and advice.

7. REFERENCES

[1] J. A. Bateman. *Utterances in context: towards a systemic theory of the intersubjective achievement of discourse*. PhD thesis, University of Edinburgh, School of Epistemics, Edinburgh, Scotland, 1986. Available as Edinburgh University, Centre for Cognitive Science In-House Publication EUCCS/PhD-7.

[2] J. A. Bateman and C. L. Paris. Phrasing a text in terms the user can understand. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1511–1517, Detroit, Michigan, 1989. IJCAI'89.

[3] J. A. Bateman and C. L. Paris. Constraining the development of lexicogrammatical resources during text generation: towards a computational instantiation of register theory. In E. Ventola, editor, *Recent Systemic and Other Views on Language*, pages 81–106. Mouton, Amsterdam, 1991.

[4] J. A. Bateman and E. Teich. Selective information presentation in an integrated publication system: an application of genre-driven text generation. *Information Processing and Management: an international journal*, 31(5):753–767, Sept. 1995.

[5] A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. How to find trouble in communication. *Speech Communication*, 40:117–143, 2003.

[6] D. Biber. A typology of English texts. *Linguistics*, 27:3–43, 1989.

[7] T. Bickmore and R. Picard. Establishing and maintaining long-term human-computer relationships. *ACM transactions on Computer Human Interaction (ToCHI)*, 2004.

[8] P. Brown and S. C. Levinson. *Politeness: some universals in language usage*. Cambridge University Press, Cambridge, 1987. Revised version of 'Universals in language usage: politeness phenomena', originally published in E. Goody (ed.)(1978) *Questions and Politeness: Strategies in Social Interaction*, Cambridge University Press.

[9] R. Brown and A. Gilman. The pronouns of power and solidarity. In T. Sebeok, editor, *Style in Language*. The M.I.T. Press, Cambridge, MA, 1960.

[10] A. Cawsey, B. L. Webber, and R. B. Jones. Natural language generation in healthcare. *Journal of the American Medical Informatics Association*, 1998.

[11] A. Cleland and M. J. Pickering. The use of lexical and syntactic information in language production: evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49:214–230, 2003.

[12] N. Colineau, C. Paris, and M. Wu. Actionable information delivery. *Revue d'Intelligence Artificielle (RSTI - RIA)*, 18(4):549–576, 2004. Special Issue on Tailored Information Delivery.

[13] F. de Rosis and F. Grasso. Affective natural language generation. In A. Paiva, editor, *Affective interactions: towards a new generation of computer interfaces*, pages 204–218. Springer, Berlin / London, 2000.

[14] F. de Rosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. De Carolis. From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59(1-2):81–118, 2003.

[15] K. Fischer. Repeats, reformulations, and emotional speech: Evidence for the design of human-computer speech interfaces. In H.-J. Bullinger and J. Ziegler, editors, *Human-Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction*, pages 560–565. Lawrence Erlbaum, London, 1999.

[16] K. Fischer and A. Batliner. What makes speakers angry in human-computer conversation. In *Proceedings of the Third Workshop on Human-Computer Conversation*, Bellagio, Italy, 2000.

[17] S. Garrod. The challenge of dialogue for theories of language processing. In S. Garrod and M. Pickering, editors, *Language Processing*. Psychology Press, Hove, 1999.

[18] S. C. Garrod and A. Anderson. Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition*, 27:181–218, 1987.

[19] S. C. Garrod and A. J. Sanford. Discourse models as interfaces between language and the spatial world. *Journal of Semantics*, 6:147–160, 1988.

[20] S. J. Green and C. DiMarco. Stylistic decision-making in natural language generation. In G. Adorni and M. Zock, editors, *Trends in Natural Language Generation: an artificial intelligence perspective*, number 1036 in Lecture Notes in Artificial Intelligence, pages 125–143. Springer, Berlin, New York, 1996. (Selected Papers from the 4th. European Workshop on Natural Language Generation, Pisa, Italy, 28-30 April 1993).

[21] M. Gregory. Aspects of varieties differentiation. *Journal of Linguistics*, 3:177–198, 1967.

[22] M. A. K. Halliday. *Language as social semiotic*. Edward Arnold, London, 1978.

[23] M. A. K. Halliday and C. M. Matthiessen. *An Introduction to Functional Grammar*. Edward Arnold, London, 3rd edition, 2004.

[24] E. H. Hovy. *Generating natural language under pragmatic constraints*. Lawrence Erlbaum, Hillsdale, New Jersey, 1988.

[25] R. T. Kasper. A flexible interface for linking applications to PENMAN's sentence generator. In *Proceedings of the DARPA Workshop on Speech and Natural Language*, 1989.

[26] A. Lankenau, O. Meyer, and B. Krieg-Brückner. Safety in robotics: The Bremen autonomous wheelchair. In *Proceedings of AMC98, 5th Int. Workshop on Advanced Motion Control*, pages 524–529, 1998.

[27] D. Y. Lee. Genres, registers, text types, domains, and styles: clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology*, 5(3):37–72, Sep 2001.

[28] K. Liu and R. W. Picard. Embedded empathy in continuous, interactive health assessment. In *Proceedings of the CHI Workshop on HCI Challenges in Health Assessment*, Portland, Oregon, April 2005.

[29] J. R. Martin. *English text: systems and structure*. Benjamins, Amsterdam, 1992.

[30] J. R. Martin. Language, register and genre. In A. Burns and C. Coffin, editors, *Analysing English in a Global Context: a reader*, Teaching English Language Worldwide, pages 149–166. Routledge, Clevedon, 2001. .

[31] C. M. I. M. Matthiessen. Register in the round, or diversity in a unified theory of register. In M. Ghadessy, editor, *Register Analysis. Theory and Practice*, pages 221–292. Pinter, London, 1993.

[32] C. M. I. M. Matthiessen and J. A. Bateman. *Text generation and systemic-functional linguistics: experiences from English and Japanese*. Frances Pinter Publishers and St. Martin's Press, London and New York, 1991.

[33] D. S. Paiva and R. Evans. A framework for stylistically controlled generation. In A. Belz, R. Evans, and P. Piwek, editors, *Natural Language Generation: Third international Conference (INLG 2004)*, number 3123 in Lecture Notes in Artificial Intelligence, pages 120–129. Springer, Berlin, New York, July 14-16 2004.

[34] C. Paris, S. Wan, R. Wilkinson, and M. Wu. Generating personal travel guides - and who wants them? In M. Bauer, P. Gmytrasiewicz, and J. Vassileva, editors, *Proceedings of the International Conference on User Modelling (UM2001)*, pages 251–253, Sonthofen, Germany, July 13-18 2001.

[35] C. L. Paris. *User modelling in text generation*. Pinter Publishers, London, 1993.

[36] C. Pelachaud, V. Carofiglio, and I. Poggi. Embodied contextual agent in information delivering application. In *Proceedings of AAMAS'02*, Bologna, Italy, July 2002. ACM.

[37] M. J. Pickering and S. Garrod. Towards a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, 27(2):169–190, 2004.

[38] K. Porayska-Pomsta and C. Mellish. Modelling politeness in natural language generation. In A. Belz, R. Evans, and P. Piwek, editors, *Natural Language Generation: Third international Conference (INLG 2004)*, number 3123 in Lecture Notes in Artificial Intelligence, pages 141–149. Springer, Berlin, New York, July 14-16 2004.

[39] R. Ross, H. Shi, T. Vierhuff, B. Krieg-Brückner, and J. Bateman. Towards dialogue based shared control of navigating robots. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, and T. Barkowsky, editors, *Spatial Cognition IV: Reasoning, Action, Interaction. International Conference Spatial Cognition 2004, Frauenchiemsee, Germany, October 2004, Proceedings*, Berlin, Heidelberg, 2005. Springer.

[40] J. Ure and J. Ellis. Language varieties – register. In *Encyclopedia of Linguistics (Information and Control)*, volume 12, pages 251–259. Pergammon Press, Oxford, 1969.

[41] M. A. Walker, J. E. Cahn, and S. J. Whittaker. Improvising linguistic style: social and affective bases for agent personality. In *Processings of the Agents '97 Conference*. ACM, 1997.

[42] J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin. Learning subjective language. *Computational Linguistics*, 30(3):277–308, Sep 2004.

[43] E. Zoltan-Ford. How to get people to say and type what computers can understand. *International Journal of Man-Machine Studies*, 34:527–547, 1991.