

An Alignment-capable Microplanner for Natural Language Generation

Hendrik Buschmeier, Kirsten Bergmann and Stefan Kopp

Sociable Agents Group, CITEC, Bielefeld University

PO-Box 10 01 31, 33501 Bielefeld, Germany

{hbuschme, kbergman, skopp}@TechFak.Uni-Bielefeld.DE

Abstract

Alignment of interlocutors is a well known psycholinguistic phenomenon of great relevance for dialogue systems in general and natural language generation in particular. In this paper, we present the alignment-capable microplanner SPUD *prime*. Using a priming-based model of interactive alignment, it is flexible enough to model the alignment behaviour of human speakers to a high degree. This will allow for further investigation of which parameters are important to model alignment and how the human-computer interaction changes when the computer aligns to its users.

1 Introduction

A well known phenomenon in dialogue situations is *alignment* of the interlocutors. An illustrative example is given by Levelt and Kelter (1982), who telephoned shops and either asked the question “What time does your shop close?” or the question “At what time does your shop close?”. The answers were likely to mirror the form of the question. When asked “At what ...?”, answers tended to begin with the preposition ‘at’ (e.g., “At five o’clock.”). Conversely, when asked “What ...?”, answers tended to begin without the preposition (e.g., “Five o’clock.”). Similar alignment phenomena can be observed in many aspects of speech production *inter alia* in syntactic and lexical choice.

Pickering and Garrod (2004) present the *interactive alignment model* bringing together all alignment phenomena of speech processing in dialogue. According to this model, human language comprehension and production are greatly facilitated by alignment of the interlocutors during conversation. The process of alignment is explained through mutual priming of the interlocutors’ linguistic representations. Thus, it is automatic, efficient, and

non-conscious. A stronger claim of the authors is that alignment — in combination with routines and a dialogue lexicon — is a prerequisite for fluent speech production in humans.

Alignment effects also occur in human-computer interaction. Brennan (1991) and Branigan et al. (in press) present evidence that syntactic structures and lexical items used by a computer are subsequently adopted by users. For this reason, alignment is an important concept for natural language human-computer interaction in general, and for dialogue systems with natural language generation in particular. Integrating ideas from the interactive alignment model into the microplanning component of natural language generation systems should be beneficial for several reasons. First, microplanning may become more efficient since the subsets of rules or lexical items in the dialogue lexicon that have been used before can be preferentially searched. Second, due to *self*-alignment, the output of the system can become more consistent and therefore easier to understand for the user. Finally, mutual alignment of user and dialogue system might make the conversation itself more natural and, presumably, cognitively more lightweight for the user.

In this paper we present a computational model for parts of the interactive alignment model that are particularly important in the context of natural language generation. We describe how this model has been incorporated into the existing SPUD *lite* system (Stone et al., 2003; Stone, 2002) to yield the *alignment-capable* microplanner SPUD *prime*. In Section 2 we describe previous approaches to integrate alignment into natural language generation. In Sections 3 and 4, we present our priming-based model of alignment and its implementation in SPUD *prime*. In Section 5, we describe results of an evaluation on a corpus of task-oriented dialogue, and in Section 6 we conclude our work and describe possible future directions.

2 Related Work

Computational modelling is an important methodology for evaluating and testing psycholinguistic theories. Thus, it is certainly not a new idea to implement the interactive alignment model computationally. Indeed, a call for “explicit computational models” is made as early as in the open peer commentary on Pickering and Garrod’s (2004) paper.

Brockmann et al. (2005) and Isard et al. (2006) present a ‘massive over-generation’ approach to modelling alignment and individuality in natural language generation. Their system generates a huge number of alternative sentences — up to 3000 — and evaluates each of these sentences with a trigram model consisting of two parts: a default language model computed from a large corpus and a cache model which is calculated from the user’s last utterance. The default language model is linearly interpolated with the cache model, whose influence on the resulting combined language model is determined by a weighting factor $\lambda \in [0, 1]$ that controls the amount of alignment the system exhibits.

Purver et al. (2006) take a more formal approach. They use an implementation of the Dynamic Syntax formalism, which uses the same representations and mechanisms for parsing as well as for generation of natural language, and extend it with a model of context. In their model, context consists of two distinct representations: a record of the semantic trees generated and parsed so far and a record of the transformation actions used for the construction of these semantic trees. *Re-use* of semantic trees and actions is used to model many dialogue phenomena in Dynamic Syntax and can also explain alignment. Thus, the authors declare alignment to be a corollary of context re-use. In particular, re-use of actions is assumed to have a considerable influence on alignment in natural language generation. Instead of looking through the complete lexicon each time a lexical item is chosen, this kind of lexical search is only necessary if no action — which constructed the same meaning in the given context before — exists in the record. If such an action exists, it can simply be re-used, which obviously leads to alignment.

A completely different approach to alignment in natural language generation is presented by de Jong et al. (2008), whose goal is to make a virtual museum guide more believable by aligning to the user’s level of politeness and formality. In

order to achieve this, the virtual guide analyses several features of the user’s utterance and generates a reply with the same level of politeness and formality. According to the authors, lexical and syntactic alignment occur automatically because the lexical items and syntactic constructions to choose from are constrained by the linguistic style adopted.

Finally, Bateman (2006) advocates another proposal according to which alignment in dialogue is predictable for it is an inherently social activity. Following the social-semiotic view of language, Bateman suggests to model alignment as arising from register and microregister. More specifically, in his opinion priming of a linguistic representation is comparable with pre-selecting a microregister that must be considered when generating an utterance in a particular social context.

The approaches presented above primarily focus on the linguistic aspects of alignment in natural language generation. The work of Brockmann et al. (2005) and Isard et al. (2006) concentrates on the surface form of language, Bateman (2006) sees alignment arising from social-semiotic aspects, and Purver et al. (2006) are primarily interested in fitting alignment into a formal linguistic framework. In this paper we adopt a more psycholinguistic and cognitive stance on alignment. Pickering and Garrod (2004) propose that low-level priming is the basic mechanism underlying interactive alignment. Here, we propose that computational modelling of these priming mechanisms also opens up an interesting and new perspective for alignment in natural language generation.

3 A Priming-based Model of Alignment

We are interested here in those parts of the interactive alignment model that are most relevant for microplanning in natural language generation and it is out of our scope to model all the facets and details of direct/repetition priming in the alignment of linguistic representations. For instance, exact timing effects are likely to be not even relevant as, in an actual system, it does not matter how many milliseconds faster the retrieval of a primed lexical item is in contrast to the retrieval of an item that is not primed. For this reason we adopt an idealised view, in which priming of linguistic structures results from two basic activation mechanisms:

Temporary activation This kind of activation should increase abruptly and then decrease slowly over time until it reaches zero again.

Permanent activation This kind of activation should increase by a certain quantity and then maintain the new level.

These two mechanisms of priming are in accordance with empirical findings. Branigan et al. (1999) present evidence for rapid decay of activation of primed syntactic structures, whereas Bock and Griffin (2000) report evidence for their long(er) term activation. In any case, Reitter (2008) found both types of priming in his analysis of several corpora, with temporary activation being the more important one. The assumption that both mechanisms play a role in dialogue is also supported by Brennan and Clark (1996) whose terminology will be followed in this paper: temporary priming will be called ‘recency of use effects’ and permanent priming will be called ‘frequency of use effects’.

Reitter (2008) assumes the repetition probability of primed syntactic structures to depend logarithmically on the distance between priming and usage. Here, recency of use effects are modelled by a more general *exponential decay* function, modified to meet the needs for modelling activation decay of primed structures:

$$ta(\Delta r) = \exp\left(-\frac{\Delta r - 1}{\alpha}\right), \quad (1)$$

$$\Delta r \in \mathbb{N}^+; \alpha > 0; \quad ta \in [0, 1]$$

$ta(\Delta r)$ is the temporary activation value of a linguistic structure depending on the distance Δr between the current time T and the time r at which the structure was primed. The slope of the function is determined by the parameter α . Additionally, the function is shifted right in order to yield an activation value of 1 for $\Delta r = 1$. This shift is due to the assumption of discrete time steps with a minimal distance of 1. A plot of $ta(\Delta r)$ with different values for α is given in Figure 1a.

Using exponential decay to model temporary activation appears to be a sensible choice that is often used to model natural processes. The advantage of this model of temporary activation lies in its flexibility. By changing the slope parameter α , different empirical findings as well as variation among humans can be modelled easily.

Next, a mathematical model for frequency of use effects is needed. To prevent that frequency effects lead to an ever increasing activation value, a maximum activation level exists. This is also found in Reitter’s (2008) corpus studies, which indicate that

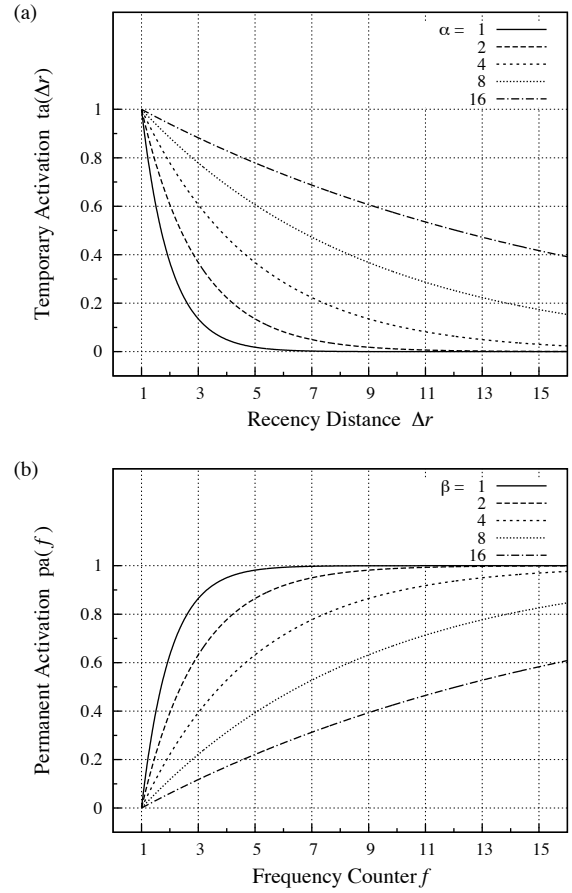


Figure 1: Plots of the mathematical functions that model recency and frequency effects. Plot (a) displays temporary activation depending on the recency of priming. Plot (b) shows permanent activation depending on the frequency count. Both are shown for different values of the slope parameter α respectively β .

the frequency effect is inversely connected to the recency effect. Here, we model recency effects with a general *exponential saturation* function, modified to meet the requirements for modelling permanent activation of linguistic structures:

$$pa(f) = 1 - \exp\left(-\frac{f - 1}{\beta}\right), \quad (2)$$

$$f \in \mathbb{N}^+; \beta > 0; \quad pa \in [0, 1]$$

The most important point to note here is that the permanent activation value $pa(f)$ is not a function of time but a function of the frequency-counter f attached to each linguistic structure. Whenever a structure is primed, its counter is increased by the value of 1. Again, the slope of the function is determined by the parameter β and the function is

shifted right in order to get an activation value of 0 for $f = 1$. A plot of equation (2) with different slope parameters is given in Figure 1b. Similar to the advantages of the model of temporary activation, this model for frequency effects is very flexible so that different empirical findings and human individuality can be expressed easily.

Now, both priming models need to be combined for a model of alignment. We opted for a weighted linear combination of temporary and permanent activation:

$$ca(\Delta r, f) = v \cdot ta(\Delta r) + (1 - v) \cdot pa(f), \quad (3)$$

$$0 \leq v \leq 1; \quad ca \in [0, 1]$$

Different values of v allow different forms of alignment. With a value of $v = 0.5$ recency and frequency effects are equally important, with a value of $v = 1$ alignment depends on recency only, and with a value of $v = 0$ alignment is governed solely by frequency. Being able to adjust the influence of the different sorts of priming on alignment is crucial as it has not yet been empirically determined to what extent recency and frequency of use affect alignment (in Section 5.2 we will exploit this flexibility for matching empirical data).

In contrast to the models of alignment presented in Section 2, the computational alignment model presented here will not only consider alignment between the interlocutors (interpersonal- or *other-alignment*), but also alignment to oneself (intra-personal- or *self-alignment*). Pickering et al. (2003) present results from three experiments which suggest self-alignment to be even more important than other-alignment. In our model, self-alignment is accounted for with the same priming-based mechanisms. To this end, four counters are attached to each linguistic structure:

- Δr_s : recency of use by the system itself
- Δr_o : recency of use by the interlocutor
- f_s : frequency of use by the system itself
- f_o : frequency of use by the interlocutor

The overall activation value of the structure is a linear combination of the combined activation value $ca(\Delta r_s, f_s)$ and the combined activation value $ca(\Delta r_o, f_o)$ from equation (3):

$$act(\Delta r_s, f_s, \Delta r_o, f_o) = \lambda \cdot (\mu \cdot ca(\Delta r_s, f_s) + (1 - \mu) \cdot ca(\Delta r_o, f_o)), \quad (4)$$

$$0 \leq \lambda, \mu \leq 1; \quad act \in [0, 1]$$

Again, by changing the factor μ , smooth interpolation between pure self-alignment ($\mu = 1$) and pure other-alignment ($\mu = 0$) is possible, which can account for different empirical findings or human individual differences. Furthermore, the strength of alignment is modelled with a scaling factor λ , which determines whether alignment is considered during generation ($\lambda > 0$) or not ($\lambda = 0$).

4 The Alignment-capable Microplanner SPUD *prime*

The previously described priming-based model of alignment has been implemented by extending the integrated microplanning system SPUD *lite* (Stone, 2002). SPUD *lite* is a lightweight Prolog re-implementation of the SPUD microplanning system (Stone et al., 2003) based on the context-free tree rewriting grammar formalism TAGLET. Not only the microplanner itself, but also the linguistic structures (the initial TAGLET trees) are represented as Prolog clauses.

SPUD *lite* carries out the different microplanning tasks (lexical choice, syntactic choice, referring expression generation and aggregation) at once by treating microplanning as a search problem. During generation it tries to find an utterance which is in accordance with the constraints set by its input (a grammar, a knowledge base and a query). This is done by searching the search space spanned by the linguistic grammar rules and the knowledge base until a goal state is found. Non-goal search states are preliminary utterances that are extended by one linguistic structure in each step until a syntactically complete utterance is found which conveys all the specified communicative goals. Since this search space is large even for relatively small grammars, a heuristic greedy search strategy is utilised.

Our alignment-capable microplanner SPUD *prime* extends SPUD *lite* in several ways. First, we altered the predicate for the initial TAGLET trees by adding a unique identifier ID as well as counters for self/other-recency/frequency values (r_s , f_s , r_o and f_o ; see Section 3). The activation value of an initial tree is then calculated with equation (4).

Furthermore, we have created a mechanism that enables SPUD *lite* to change the recency and frequency information attached to the initial trees on-line during generation. This is done in three steps with the help of Prolog's meta-programming capabilities: Firstly, the clause of a tree is retrieved

from the knowledge base. Secondly, it is retracted from the knowledge base. Finally, the clause is (re-)asserted in the knowledge base with updated recency and frequency information. As a welcome side effect of this procedure, primed initial trees are moved to the top of the knowledge base and — since Prolog evaluates clauses and facts in the order of their appearance in the knowledge base — they can be accessed earlier than unprimed initial trees or initial trees that were primed longer ago. Thus, in SPUD *prime* recency of priming directly influences the access of linguistic structures.

Most importantly, the activation values of the initial trees are considered during generation. Thus, in addition to the evaluation measures used by SPUD *lite*'s heuristic state evaluation function, the mean activation value

$$\overline{act}(S) = \frac{\sum_{i=1}^N act_{t_i}(\Delta r_{s_i}, f_{s_i}, \Delta r_{o_i}, f_{o_i})}{N}$$

of the N initial trees $\{t_1, \dots, t_N\}$ of a given search state S is taken into account as a further evaluation measure. Hence, when SPUD *prime* evaluates (otherwise equal) successor search states, the one with the highest mean activation value is chosen as the next current state.

5 Evaluation

In order to find out whether our priming-based alignment model and its implementation work as intended, we evaluated SPUD *prime* on a corpus that was collected in an experiment designed to investigate the alignment behaviour of humans in a controlled fashion (Weiß et al., 2008). The part of the corpus that we used consists of eight recorded and transcribed dialogues between two interlocutors that play the ‘Jigsaw Map Game’, a task in which different objects have to be placed correctly on a table. Speakers take turns in explaining each other where to place the next object in relation to the objects that are already on the table. Each speaker has to learn a set of name–object relations before the game, such that both use the same names for all but three objects. Due to this precondition, both speakers use the same lexical referring expressions for most objects and the speaker’s lexical alignment behaviour for the differently named objects can be observed easily.

In our evaluation, we concentrate on the generation of nouns by simulating the uses of the three differently learned nouns in the eight dialogues

from the perspective of all sixteen interlocutors. In each test, SPUD *prime* plays the role of one of the speakers talking to a simulated interlocutor who behaves exactly as in the real experiment. With this test setup we examined, first, how well SPUD *prime* can model the alignment behaviour of a real speaker in a real dialogue context and, second, whether our model is flexible enough to consistently emulate different speakers with different alignment behaviour.

In order to find the best model (i.e., the best parameter set $\{\alpha, \beta, \mu, \nu\}$) for each speaker, we simulated all tests with all parameter combinations and counted the number of mismatches between our model’s choice and the real speaker’s choice. To make this exhaustive search possible, we limit the set of values for the parameters α and β to $\{1, 2, 4, 6, 8, 10, 14, 18, 24, 30\}$ and the set of values for the parameters μ and ν to $\{0, 0.1, 0.2, \dots, 1\}$, resulting in a total of $11^2 \times 10^2 = 12100$ different parameter sets. Since we want to investigate alignment, λ is constantly set to 1.

5.1 An Illustrative Example

To illustrate our evaluation method, we first present and discuss the simulation of one particular dialogue (from the Jigsaw Map Game corpus) from the perspective of participant (A). Before the experiment started, both interlocutors learned the name–object relations ‘Raute’ (rhombus), ‘Ring’ (ring), ‘Schraube’ (bolt) and ‘Würfel’ (cube), additionally participant (A) learned ‘Spielfigur’ (token), ‘Ball’ (sphere) and ‘Block’ (cuboid) and participant (B) learned ‘Männchen’ (token), ‘Kugel’ (sphere) and ‘Klotz’ (cuboid). In our simulation, we focus on the use of the differently learned names (the targets) and not on the other names (non-targets). Table 1 shows the sequence of target nouns as they occurred in the real dialogue (non-targets omitted).

For each parameter set $\{\alpha, \beta, \mu, \nu\}$ the dialogue is simulated in the following way:

- When participant (A) used a referring *non-target* noun in the dialogue, self-priming of the corresponding rule(s) in SPUD *prime*'s knowledge base is simulated (i.e., the recency and frequency counters are increased).
- When participant (A) used a referring *target* noun in the dialogue, SPUD *prime* is queried to generate a noun for the target object. Then it is noted whether the noun actually generated

	B: der Klotz	14	A: der Klotz
1	A: die Spielfigur	15	A: die Kugel
2	A: der Klotz	16	A: der Klotz
	B: das Männchen		B: der Klotz
	B: der Klotz		B: die Kugel
3	A: die Spielfigur		B: der Klotz
	B: das Männchen	17	A: der Klotz
4	A: das Männchen		B: das Männchen
5	A: das Männchen		B: der Klotz
6	A: das Männchen	18	A: das Männchen
7	A: das Männchen	19	A: der Klotz
8	A: das Männchen		B: das Männchen
	B: das Männchen	20	A: der Ball
9	A: das Männchen	21	A: das Männchen
10	A: der Ball		B: der Ball
	B: der Ball		B: das Männchen
11	A: der Ball	22	A: die Kugel
12	A: der Ball	23	A: der Ball
	B: die Kugel		B: der Klotz
	B: das Männchen	24	A: der Ball
13	A: der Ball		B: der Klotz
	B: die Kugel	25	A: der Klotz

Table 1: Sequence of referring target nouns used by participants (A) and (B) in our example dialogue.

is the noun used in the actual dialogue (match) or not (mismatch).

- When participant (B) used a referring noun (target or non-target), priming of the corresponding rule(s) in SPUD *prime*'s knowledge base is simulated.

The evaluation measure for a specific parameter set is the number of mismatches it produces when simulating a dialogue. Thus the parameter set (or rather sets) which produce the least number of mismatches are the ones that best model the particular speaker under consideration. For participant (A) of our example dialogue the distribution of parameter sets p producing m mismatches is shown in Table 2. Four parameter sets produce only two mismatches (in phrase 15 and 22; cf. Table 1) and thus our priming-based alignment model can account for 92% of the target nouns produced by speaker (A). However, it must be noted that these two mismatches occur at points in the dialogue where the alignment behaviour of (A) is not straightforward. At target noun 15, both interlocutors have already used the name 'Ball' and then both switch to 'Kugel'. The mismatch at target 22 is a special case: (A) used 'Kugel' and immediately corrected himself to 'Ball', the name he learned prior to the experiment. It seems as if the task instruction, to use the learned nouns, suddenly became prevalent.

m	0	1	2	3	4	5
$\# p$	0	0	4	833	3777	2248
m	6	7	8	9	10	...
$\# p$	3204	1105	478	148	294	0

Table 2: Number of parameter sets p leading to m mismatches for participant (A) in dialogue 7.

5.2 Simulation Results

To evaluate our alignment-capable microplanner, we simulated the noun production for each of the interlocutors from the experiment. One dialogue has been excluded from the data analysis as the dialogue partners used nouns that none of them had learned in the priming phase. For each of the remaining 14 interlocutors we varied the parameters α , β , μ and ν as described above to identify those parameter set(s) which result in the least number of mismatches.

Each interlocutor produced between 18 and 32 target nouns ($N=14$, $M=23.071$, $SD=3.936$). Our simulation runs contain between 0 and 19 mismatches overall ($N=169400$, $M=6.35$, $SD=3.398$). The minimal number of mismatches for each speaker simulation ranges between 0 and 6 ($N=14$, $M=2.286$, $SD=1.684$). That is, our model can simulate a mean of 89.9% of all target nouns ($N=14$, $M=.899$, $Min=.667$, $Max=1.000$, $SD=.082$), which is an improvement of 24.6% on the baseline condition (alignment switched off), where 65.3% of the target nouns are generated correctly ($N=14$, $M=.653$, $Min=.360$, $Max=1.000$, $SD=.071$). As already illustrated in Section 5.1, mismatches typically occur at points in the dialogue where the alignment behaviour of the interlocutor is not straightforward.

As displayed in Table 3 the parameter assignments resulting in least mismatches differ considerably from speaker to speaker. However, there are some remarkable trends to be observed in the data. As concerns the parameter μ , which determines the combination of self- and other-alignment, the majority of values are in the upper range of the interval $[0,1]$. For 8 of 14 speakers the mean is above 0.7 with relatively low standard deviations. Only for one speaker (P13) the mean μ is below 0.3. Thus, the parameter values indicate a considerable tendency toward self-alignment in contrast to other-alignment.

For the parameter ν that interpolates between recency and frequency effects of priming, the res-

	m	# p	α		β		μ		ν	
			M	SD	M	SD	M	SD	M	SD
P13	2	4	3.0	1.155	19.5	9.14	.1	.0	.3	.0
P14	1	72	5.53	1.52	14.32	9.61	.819	.040	.901	.108
P17	1	200	1.66	.823	12.94	9.529	.353	.169	.955	.069
P18	3	2445	15.37	8.758	10.98	9.76	.597	.211	.706	.236
P19	0	4321	11.81	9.492	11.01	8.929	.824	.148	.387	.291
P20	2	8	1.0	.0	15.75	9.285	.737	.052	.388	.146
P23	6	987	6.85	6.681	12.08	9.354	.331	.374	.4	.33
P24	3	256	12.95	9.703	13.63	8.937	.537	.201	.468	.298
P39	5	1	1.0	.0	2.0	.0	.9	.0	.8	.0
P40	0	3504	12.08	9.33	10.30	8.753	.843	.147	.343	.282
P41	2	609	11.37	8.475	15.34	8.921	.770	.106	.655	.213
P42	3	30	6.0	1.486	17.53	9.016	.783	.059	.760	.122
P47	2	326	13.75	7.794	13.53	9.508	.772	.095	.816	.166
P48	2	2478	12.87	9.545	10.74	8.538	.764	.175	.166	.148

Table 3: Mean parameter values for those simulation runs which result in a minimal number of mismatches for each speaker.

ults are less revealing. For two speaker simulations (P13 and P48) the mean ν is 0.3 or lower, for another four speaker simulations the mean ν is above 0.7. That is, our model produces good matching behaviour in adopting different alignment strategies, depending either primarily on frequency or recency, respectively. All other simulations, however, are characterised by a mean ν in the medium range along with a relatively high standard deviation.

6 Conclusion

In this paper, we introduced a priming-based model of alignment which focusses more on the psycholinguistic aspects of interactive alignment and models recency and frequency of use effects — as proposed by Reitter (2008) and Brennan and Clark (1996) — as well as the difference between intrapersonal and interpersonal alignment (Pickering et al., 2003; Pickering and Garrod, 2004). The presented model is fully parameterisable and can account for different empirical findings and ‘personalities’. It has been implemented in the SPUD *prime* microplanner which activates linguistic rules by changing its knowledge base on-line and considers the activation values of those rules used in constructing the current utterance by using their mean activation value as an additional feature in its state evaluation function.

We evaluated our alignment model and its implementation in SPUD *prime* on a corpus of task-oriented dialogue collected in an experimental setup especially designed for alignment research. The results of this evaluation show that our priming-based model of alignment is flexible enough to simulate the alignment behaviour of different human

speakers (generating target nouns) in the experimental setting. It should be noted, however, that our model tries to give a purely mechanistic explanation of lexical and syntactic choice and that it, therefore, cannot explain alignment phenomena that are due to social factors (e.g., politeness, relationship, etc.), audience design or cases, in which a speaker consciously decides whether to align or not (e.g., whether to use a word or its synonym). While the evaluation has shown that our model can reproduce human alignment behaviour to a high degree, it remains to be investigated which influence each parameter exerts and how exactly the parameters vary across individual speakers.

Nevertheless, the development of the alignment-capable microplanner is only one step in the direction of an intuitive natural language human-computer interaction system. In order to reach this goal, the next step is to combine SPUD *prime* with a natural language understanding system, which should ideally work with the same linguistic representations so that the linguistic structures used by the interlocutor could be primed automatically. This work is underway.

Furthermore, user studies should be carried out in order to evaluate SPUD *prime* in a more sophisticated way. Branigan et al. (in press) found that human-computer alignment was even stronger than human-human alignment. But how would the alignment behaviour of human interlocutors change if the computer they are speaking to also aligns to them? Further, would integration of an alignment-capable dialogue system into a computer interface make the interaction more natural? And would an embodied conversational agent appear

more resonant and more sociable (Kopp, 2008) if it aligned to users during conversation? The work presented here provides a starting point for the investigation of these questions.

Acknowledgements – This research is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence in ‘Cognitive Interaction Technology’ (CITEC) as well as in the Collaborative Research Center 673 ‘Alignment in Communication’. We also thank Petra Weiß for making the ‘Jigsaw Map Game’ corpus available and three anonymous reviewers for their helpful comments.

References

- John A. Bateman. 2006. A social-semiotic view of interactive alignment and its computational instantiation: A brief position statement and proposal. In Kerstin Fischer, editor, *How People Talk to Computers, Robots and Other Artificial Communication Partners*, SFB/TR 8 Report No. 010-09/2006, pages 157–170, Bremen, Germany.
- J. Kathryn Bock and Zenzi M. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129(2):177–192.
- Holly P. Branigan, Martin J. Pickering, and Alexandra A. Cleland. 1999. Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, 6(4):635–640.
- Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. in press. Linguistic alignment between people and computers. *Journal of Pragmatics*.
- Susan E. Brennan and Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482–1493.
- Susan E. Brennan. 1991. Conversation with and through computers. *User Modeling and User-Adapted Interaction*, 1(1):67–86.
- Carsten Brockmann, Amy Isard, Jon Oberlander, and Michael White. 2005. Modelling alignment for affective dialogue. In *Proc. of the Workshop on Adapting the Interaction Style to Affective Factors at the 10th Int. Conf. on User Modeling*.
- Markus A. de Jong, Mariët Theune, and Dennis Hofs. 2008. Politeness and alignment in dialogues with a virtual guide. In *Proc. of the 7th Int. Conf. on Autonomous Agents and Multiagent Systems*, pages 207–214.
- Amy Isard, Carsten Brockmann, and Jon Oberlander. 2006. Individuality and alignment in generated dialogues. In *Proc. of the 4th Int. Natural Language Generation Conf.*, pages 25–32.
- Stefan Kopp. 2008. From communicators to resonators – Making embodied conversational agents sociable. In *Proc. of the Speech and Face to Face Communication Workshop in Memory of Christian Benoit*, pages 34–36.
- Willem J. M. Levelt and Stephanie Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology*, 14(1):78–106.
- Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–226.
- Martin J. Pickering, Holly P. Branigan, and Janet F. McLean. 2003. Dialogue structure and the activation of syntactic information. In *Proc. of the 9th Annual Conf. on Architectures and Mechanisms for Language Processing*, page 126.
- Matthew Purver, Ronnie Cann, and Ruth Kempson. 2006. Grammars as parsers: Meeting the dialogue challenge. *Research on Language and Computation*, 4(2–3):289–326.
- David Reitter. 2008. *Context Effects in Language Production: Models of Syntactic Priming in Dialogue Corpora*. Ph.D. thesis, University of Edinburgh.
- Matthew Stone, Christine Doran, Bonnie Webber, Tonia Bleam, and Martha Palmer. 2003. Microplanning with communicative intentions: The SPUD system. *Computational Intelligence*, 19(4):311–381.
- Matthew Stone. 2002. Lexicalized grammar 101. In *Proc. of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, pages 77–84.
- Petra Weiß, Thies Pfeiffer, Gesche Schaffranietz, and Gert Rickheit. 2008. Coordination in dialog: Alignment of object naming in the Jigsaw Map Game. In *Proc. of the 8th Annual Conf. of the Cognitive Science Society of Germany*, pages 4–20.