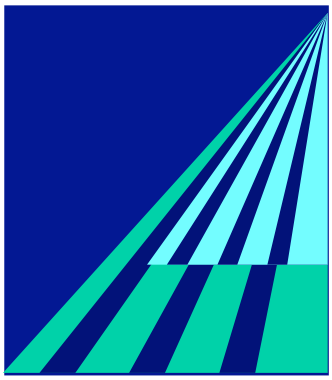
Universität Bielefeld

Short key data and funding overviews for the planned Collaborative Research Centre



Alignment in Communication

Bielefeld University
(Spokesman: Gert Rickheit)

Funding Proposal
2006 – 2007 – 2008 – 2009

Contents

Overview of the project sections

Project section	Title	Research area and direction	Leader, institute, location
A	**Interpersonal aspects of alignment**		
A1	Modelling partners	Artificial intelligence, psycholinguistics	Wachsmuth, Rickheit
A2	Alignment in visual communication	Human-machine communication, interaction modelling, computer vision	Kummert, Ritter
A3	Syntactic-prosodic alignment in language processing and communication	Psycholinguistics, discourse analysis, system theory	Kindt, Weingarten
A4	Syntactic alignment in normal and impaired communication	Psycholinguistics, clinical linguistics, syntax	Hielscher-Fastabend, Sahel, Eikmeyer
A5	Alignment of perceptual and relational components of situation models	Psycholinguistics, computational linguistics	S. Wachsmuth, Sichelschmidt, Rickheit
A6	Alignment of attention in mediated communication	Neuroinformatics, psycholinguistics	Ritter, Sichelschmidt
B	**Intrapersonal aspects of alignment**		
B1	Speech-gesture alignment	Artificial intelligence, psycholinguistics, computational linguistics	Kopp, Wachsmuth, Rieser
B2	Alignment-based accounts of compositionality and indirect interpretation	Cognitive linguistics, psychosemantics, computational linguistics	Eikmeyer, Rieser, Sichelschmidt
B3	Construction of implicit common ground	Psycholinguistics, semantics, artificial intelligence	Eikmeyer, Jäger, Rickheit, Rieser
B4	Neural synchronisation during language processing: Intrapersonal alignment	Neurolinguistics, neuroinformatics	Müller, Ritter
B5	Gestalt approaches to alignment – from visual to verbal processing	Psycholinguistics, cognitive sciences, computational linguistics	Mehler, Rickheit, Sichelschmidt
B6	Alignment of acoustic-phonetic representations in speech perception and productions	Speech recognition, psycholinguistics	Fink, Eikmeyer
B7	Multimodal language processing: Dynamics and integration of parallel processes in verbal communication	Neurolinguistics, neurology	Müller, Wörmann

Project section	Title	Research area and direction	Leader, institute, location
C	**Systemic aspects of alignment**		
C1	Interaction space	Artificial intelligence, interaction modelling	Wachsmuth, Sagerer, Hielscher-Fastabend
C2	Modelling routines, discourse strategies and conversational styles	Text technology, statistical dialogue modelling, pragmatics	Metzing, Sagerer, Wrede
C3	Dialogue games and group dynamics	Semantics, pragmatics, game theory, psycholinguistics	Jäger, Strohner
C4	Emotional alignment. Handicap and optimisation of intra- and interpersonal comprehension of emotional information	Emotional expression, clinical linguistics	Hielscher-Fastabend, Strohner
C5	Repairs and reformulations in dialogue	Pragmatics, psycholinguistics, computational linguistics	Eikmeyer, Kindt, Strohner, Weingarten
C6	Adaptive alignment in human-robot cooperation	Neuroinformatics, learning, robotics	Ritter, Steil, Sagerer
C7	Semantic alignment for keyword-based database front-ends	Human-machine cooperation, dialogue systems	Ritter, Sagerer, Mehler
C8	Visual alignment dynamics for feedback in cooperative search	Multimodal alignment, eye tracking, relevance feedback	Ritter, Sagerer, Jäger
X			
X1	Development of XLM-based multimodal alignment corpora	Linguistics, text technology, computational linguistics	Mehler, Eikmeyer, Rieser, Metzing
X2	System evaluation	Linguistics, informatics	Kopp, Steil, Wrede
Z	**Central services**		Rickheit

Overview of the staff funds and consumables

Financial year	Staff funds	Auxiliary funding Consumables	Investment	Total
2006	2.841,600	254,500	30,000	**3.126,100**
2007	2.841,600	158,800	—	**3.000,400**
2008	2.841,600	158,800	—	**3.000,400**
2009	2.841,600	160,800	—	**3.002,400**

List of instruments costing over € 10,000 (gross)

Project section	Description of the item (with type designation, if applicable)	Requested for financial year			
		Price per item 2006	Price per item 2007	Price per item 2008	Price per item 2009
A6	Eye-Tracker "Eye LinkII"	30,000	—	—	—
	Total:	**30,000**	—	—	—

Alignment in Communication

A Proposal for the Establishment of a Collaborative Research Centre

1. Why Alignment?

1.1 A New Side of Communication

In more than a decade of intensive research on "Situated Artificial Communicators", a team of linguists, psychologists, and computer scientists at Bielefeld University have jointly tackled questions of information processing in task-oriented communication – to the effect that the role of key notions such as integration, context, reference, coherence and robustness has been explored in great depth.

In the course of our research in task-oriented communication, we have frequently observed some rather stunnig phenomena; among them the following:

- There is collaboration: Communicators often complement one another in the production of an utterance, its components being supplied by different individuals (as in [A] "Now you take…" – [B] "…a bolt.").
- There is improvisation: Communicators' contributions occasionally appear "ad hoc" in the sense that they come about as a by-product of activity, without much purposeful planning (as in [A] "No way." – [B] "Must be.").
- There is creativity: Communicators readily agree on the use and interpretation of metaphorical or newly invented terms without any further explanation (as, for instance, the phrase "benzole-shaped bolt" to refer to a hex-bolt)
- There is tolerance: Communicators often do not correct ill-formed or blatantly erroneous utterances – as long as they are useful (as in "…with the three holes the piece of small wood" or in "…fasten the hole in front with an angular bolt").
- There is comprehensiveness: Communicators frequently supplement verbal exchange by nonverbal means, rendering dialogue a multi-modal affair (such as [A, showing an object to B] "You can see it anyway").

Such phenomena clearly demonstrate that communication is, to a large extent, a matter of joint activity. Current theories of cognition by and large focus on the more obvious processes going on in communication; they do not provide sufficient explanations for less overt phenomena like the ones mentioned above. Hence, in our striving for a comprehensive understanding of communication, we have repeatedly encountered points where we felt that the existing paradigms in all of the disciplines concerned lead towards a dead end.

We now understand better than before that communication and human information processing have still another, as yet rather unexplored side. Conceivably, that side is characterised by automatisms rather than by plans, by conventions rather than by intentions, by concord rather than by effort, and by alignment rather than by argument. In fact, interest in that side of communication seems to be growing among the scientific community (as has become apparent in our intensive discussions with pioneers in the field; Rickheit, 2004). In consideration of the demand in research on those aspects, we propose to establish a collaborative research centre on "Alignment in Communication."

Such a research initiative will, in our conviction, substantially broaden our knowledge of the true nature of human information exchange. It will contribute to the theoretical development

in the humanities and, at the same time, bring about practical advancement in technology. Hence we believe that the theme we have to offer is informative and timely, and it is exciting.

1.2 The Notion of Alignment in Communication

The ultimate purpose of the proposed research initiative is to extend our knowledge about the cognitive processes underlying natural language use in human-human communication as well as in language production and comprehension in human-machine interaction. At that, an interdisciplinary approach will be taken, bringing together linguistics, artificial intelligence, neuroinformatics, neurolinguistics, computational linguistics, and psycholinguistics, and employing a conjoint method that integrates description and experimentation with simulation and evaluation.

The common conviction that motivates our commitment is quite straightforward: We believe that *each and every detail in conversation should be taken for granted*. In our conviction, conversation is so easy because people engaged in dialogue have at their disposal an extensive pool of information, plus appropriate ways of transmission (by verbal as well as nonverbal means); moreover, they can take advantage of even the most rudimentary of verbal expressions. In other words, we suggest abandoning the traditional idea that dialogue comprises an abundance of such things as ambiguous, incomplete or incorrect utterances; rather, we think that every single expression is assumed to be a potential contribution to the purpose of making sense. To illustrate, rather than adhering to the traditional idea that a (seemingly erroneous) phrase like *"fasten the hole in front with an angular bolt"* is a deficient manifestation of some underlying well-formed, readily interpretable linguistic structure, we proceed from the assumption that such a phrase, together with the particular situation it is embedded in, provides the listener with cues as to which aspects he or she should specifically attend to in his or her conduct. Thus, even a "bad" expression may be good enough to trigger sensible action. In Hörmann's (1983: 233) words, "the listener does not understand the utterance; he understands the speaker. More precisely, he understands what the speaker, in this situation, wants the listener to think."

In consequence, we suggest focussing on the less obvious aspects of communication that have been addressed above: the role of automaticity, conventions, concord, convergence, and the like. We believe that, crucially, such a theory of communication must take account of the alignment taking place between interlocutors. In broad terms, *alignment can be defined as an ensemble of verbal and non-verbal means that serve to increase the similarity in structure of two interacting dynamic systems without an explicit exchange of information on system states*.

While alignment is a well-established term in several disciplines, notably in the natural and life sciences, in technical signal processing, and in applied sciences such as arts or architecture, the notion of alignment in human information interchange has not received much attention to date – in spite of its fundamental role in mutual understanding and in the striving to make one's environment "intelligible" (Hörmann, 1983). We mean by *alignment in communication* the – seemingly casual – process of common orientation of the participants in a conversation. Precisely here is the crucial point: Conceivably, common orientation, brought about by alignment, is what facilitates conversation. Common orientation exempts agents from constantly discussing their respective points of view.

In an ongoing discussion, Pickering and Garrod (2004) have pushed this argument somewhat further. They claim that dialogue is characterised by a special interactive device which supports the automatic alignment of linguistic representations at diverse levels – which, in turn, facili-

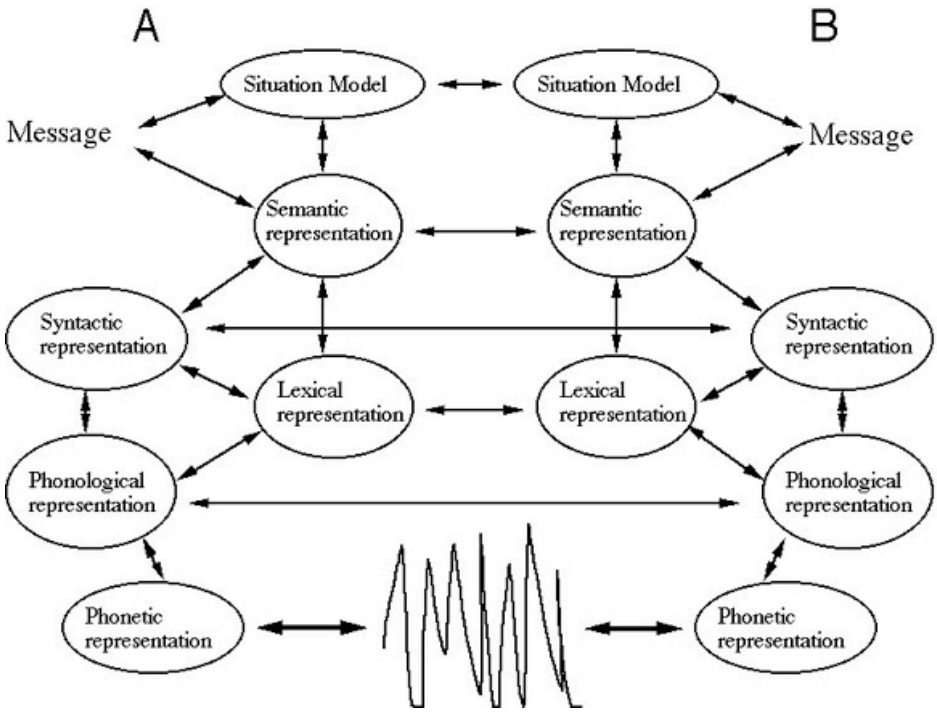


Figure 1: Assumed stages of language comprehension and production for two interlocutors according to the interactive alignment model of Pickering and Garrod (2004).

tates the production and comprehension of verbal utterances. Figure 1 is a schematic visualisation of the assumed stages of language comprehension and production processes for two interlocutors (A and B) engaged in dialogue; arrows indicate the interrelations of diverse levels of representation. According to this view, interactive alignment operates through two kinds of automatic "priming" mechanisms. The first mechanism depends on transient activation of linguistic representations used in both production and comprehension; it produces short-term alignment effects. The second is an implicit memory-based mechanism which leads interlocutors to establish routines that hold throughout that dialogue and that enable the use of semi-fixed expressions. Routinisation benefits production by restricting the decision space in utterance formulation, and it benefits comprehension by disambiguating the verbal input. These mechanisms can be considered automatic in that they operate in a resource-saving fashion while taking into account the goals, means, and actions of the interlocutors; hence the term "mechanistic" as opposed to conscious control.

Our fascination with the notion of alignment in communication is inspired by the discussions that have evolved around these ideas and by the fact that there is a growing body of evidence on the cognitive processes that govern dialogue (Barr, 2004). The ideas which have been outlined so far may thus be taken as a starting point for a more in-depth exploration of alignment which is meant to heighten and to extend our knowledge of what is going on in everyday human-human communication as well as in human-machine communication.

Proceeding from a framework that is centered around the notion of alignment in communication, we attempt to provide well-founded answers to the following questions:

1. Prerequisites: What are the requirements for a unitary process model of communication, and what is the contribution of the notion of alignment to such a model?

2. Viability: What are the pros and cons of characterising communication in terms of alignment, and what is the benefit of such an endeavour to language users?
3. Inter-personal aspects: What is the advantage of an alignment-based approach in describing and explaining the interactions between agents in communication?
4. Intra-personal aspects: What is the advantage of alignment in describing and explaining individual cognitive processes in language production or comprehension?
5. Versatility: Is it possible to reasonably extend the notion of alignment in communication so as to also cover non-verbal interchange?
6. Evidence: What methods are best suited to empirically substantiate an alignment-based approach to the dynamics of communication?
7. Applicability: How can aspects of alignment be simulated in artificial agents and to what extent can human-machine interaction be improved that way?

In general, we strive to address in detail the question as to what are the prospects of studying alignment in communication with respect to...

- What: How do people agree on salient components of a complex situation?
- Who: How do people address persons and specify referents in an efficient way?
- Where: How do people converge on ways of structuring environmental space?
- When: How do people synchronise verbal and nonverbal events?
- Why: How do people achieve mutual understanding?

The pursuit of research questions like these constitutes a great challenge to cognitive science. Consequently, any answers will mark significant steps on the way towards a comprehensive theory of communication in human-human and human-machine environments.

Such a theory of communication would be of a special quality because it would be grounded on extensive empirical observations of natural language use in lifelike situations. Moreover, such a theory could be subject to evaluation by means of implementation in a natural language processing system or, more comprehensively, in an embodied conversational agent. Above all, such a theory would be truly interdisciplinary because it would not be restricted to verbal expressions of idealised structure and limited complexity (as customary in linguistics) nor would it be restricted to either the production or the comprehension of verbal expressions (as customary in psycholinguistics). It would also not be geared to application in technical systems without taking into consideration the cognitive basis (as customary in artificial intelligence). As a theory of dialogue, it would have to take into account the verbal and non-verbal processes, the participants that engage in conversation, and the particulars of the situation in question. As a theory of multi-modal communication, it will shed more light on the role of channels such as vision, gesture, gaze, and body posture and their cross-modal interdependencies, which facilitate face-to-face communication to such a high degree. And as a theory of human-machine communication, it will provide new approaches for designing visuo-verbal interfaces for more natural human-machine cooperation. Hence, such a theory would go far beyond language. After all, we believe that communication is not just a matter of arranging words but a matter of aligning worlds.

2. Alignment – Emergence of a New Paradigm

With respect to research on dialogue, the notion of alignment has been introduced only recently. For a considerable time there had been a division of labour between theoretical and descriptive linguistics on the one hand and classical discourse oriented disciplines on the other hand. The former was confined to the domain of sentence while the latter dealt with various

aspects of discourse. At least two developments have brought about a dramatic change: the introduction of Dialogue Game Theory (DGT), according to which interlocutors produce dialogue moves according to scripted patterns, and the innovative impetus that Discourse Representation Theory (DRT) with its notion of flexible scope meant for logic and semantics. While, in a sense, both these developments can be viewed as generalisations of established techniques, both approaches, suitably combined, yield interesting variants of current discourse theory (cf. Steels, 2002).

However, theorising in dialogue had not been seriously tied down to natural language data until the seminal ideas proposed by Clark (1996). Perhaps the most important of these innovative ideas is the claim that agents cooperate in conversation – which makes the production of linguistic structure and the making of sense a multi-agent affair; a "joint project" (to take up Clark's suggestive term). According to the joint project approach, coordination at all levels of dialogue, from attention getting to dialogue game structure, is the key to successful communication. The coordination mechanisms brought to the fore in the joint project approach are based on negotiation, tactical control, and explicit verbalisation, i.e. on utterances produced, repaired or extended, or on content accepted or denied by the participants.

In a sense, the joint project approach is licensed by the more obvious processes in dialogue. Close inspection of pertinent conversations shows that the joint project approach is well supported and that it sheds light on a good deal of the empirical data. Clearly, explanations of dialogue cannot do without a notion like coordination-by-negotiation. However, coordination-by-negotiation is only one side of the coin. Beyond negotiation and explicit verbalisation, there is coordination of a quite different sort. This kind of coordination is characterised by automaticity, simplicity, subtlety, and locality; it requires little or no cognitive effort and has an air of smoothness, incidentality and routine about it. Subtle signs of this kind of coordination are phenomena like the ones mentioned in Section 1.1, which have been observed in the multimodal "Baufix" corpora of the Bielefeld CRC "Situated Artificial Communicators", and which can be summarised as follows:

- Help concerning discourse management or domain knowledge may arrive unsolicited.
- Grounding frequently proceeds without explicit acknowledgement or acceptance.
- Completions or continuations by the other participant occur without prompting.
- Introduction, re-use, and adjustment of terms are routinely brought about.
- Participants often tacitly agree on an interpretation of underspecified or novel terms.
- Agents' gestures are closely synchronised with speech and harmonise with dialogue moves.

Such occurrences cannot readily be explained by the joint project approach or any other current theory. The interactive alignment approach, however, maintains that agents coordinate their behaviour and exert control in a direct, efficient, non-deliberate and essentially resource-free way (Barr, 2004; Pickering & Garrod, 2004). This notion of alignment is intended to account for primitive or basic language use resulting from continuous, precise, fine-grained and mostly non-marked coordination. Altogether, coordination-by-alignment is not intended to supplant the notion of coordination-by-negotiation in dialogue but to supplement it.

There are many approaches in the area of natural communication that focus on the interaction with artificial systems, be it robotic or virtual, visualised ones. Particularly relevant here is the work aiming at the issue of the types of interactional features that are necessary in order to render an interaction as natural and non-intrusive as possible. Psycholinguistic studies have demonstrated that linguistic functions such as reference, perspective taking, etc., are achieved collaboratively, and coordination-by-negotiation and coordination-by-alignment have now

been shown to take place on all linguistic levels (Barr, 2004; Clark, 1996; Pickering & Garrod, 2004). These results have been taken up and supported by findings in human-computer interaction (e.g., Hoc, 2001). At present, however, the notion of alignment is only at the very beginning of making its way into the cognitively inspired research on and development of technical systems.

In this regard, it is an important aspect of the research initiative we propose here that we strive to render the results of our theoretical research on alignment in communication into operational models of the involved processes, and we aspire to realise these models in technical systems. From the theoretical point of view, the rationale is that simulating a theoretical model and bringing it to application provides us with a way to evaluate its validity, explanatory power, and tractability, which, in turn, can give rise to iterative revisions. From the practical point of view, the construction of systems based on empirically and theoretically grounded models of alignment will allow for significant advances in human-machine interaction.

3. What is Innovative about Alignment?

3.1 Interpersonal Alignment: A New Approach to Conversation

The interactive alignment approach presents a novel research paradigm on dialogue, covering roughly the same field as do the Dialogue Game Theory and the joint project approach. So what innovative alternatives to these more established formulations does the notion of alignment offer? In order to answer this question, we shall – against the background of what is currently perhaps the most sophisticated version of an interactive alignment approach (Pickering & Garrod, 2004) – discuss issues of language use in dialogue first, then proceed to implications for theory of grammar, touch upon methodological issues afterwards, and finally consider alignment as a general trait of social behaviour.

3.1.1 New Insights on Language Use in Dialogue

In their interactive alignment theory, Pickering and Garrod (2004) claim that successful dialogue depends on aligned representations at all linguistic levels. At that, the alignment of situation models is tied to alignment at the lower levels (although directionality of processes is not an issue). Global alignment arises from local alignment at any level of linguistic representations via priming: Priming at one level triggers priming at another level. Based on incoming information, alignment of situation models is hence automatic.

The authors' stance towards grounding, i.e. the accumulation of mutually accepted information in dialogue, is worked out in a more subtle fashion than is commonplace in linguistics. Interlocutors align on (epistemically and doxastically neutral) implicit common ground, with interactive alignment as the underlying primitive, "non-costly" mechanism (although clarification questions are permitted). In contrast, establishing full common ground is considered a specialised and non-automatic process. As a consequence, it is assumed that speakers and listeners do not normally take into account full common ground. Rather, strategies like elaborate reasoning or inference based on full common ground are viewed as optional strategies that can be employed especially in order to overcome problems that arise from (presumed) misalignment.

Alignment is based on priming: Structured input automatically generates structured representations. In addition, alignment is massively fostered by the interactive nature of dialogue. First of all, there is a sort of copying effect: We have parity between representations used in production and comprehension pointing towards a principle of input-output coordination. Alignment on the level of lexical processing (string, lexical category and interpretation) leads to

alignment on the set of referring expressions used. By the same token we achieve syntactic alignment for constituents. Furthermore, the development of a lexicon of expressions relevant for reference to the actual domain promotes the development of dialogue routines of different complexity: The construction of temporary formulae is common in dialogue. On the more obviously interactive level we have in addition parallelisms chosen, joint constructions and the fixing and monitoring of common ground. Finally, natural language itself provides constraints for information packaging in dialogue to be observed in order to guarantee well-formedness of linked dialogue moves.

So, this is what emerges as an alternative to the joint project approach at all levels: the notion of alignment and the construct of implicit common ground.

3.1.2 New Impetus for a Theory of Language

Theories of dialogue can profit a lot from the interactive alignment approach. However, interactive alignment also has consequences for the established doctrine of linguistic theory: The existence of joint constructions in dialogue (e.g., completions, continuations, or handling of fragments) necessitates the assumption of "sub-sentential turns". As a consequence, linguistics must work with a more flexible notion of constituency than the ones adopted by most paradigms.

Another argument shows the importance of implementing the notion of incrementality into dialogue-bound linguistics: Since interlocutors can only control the alignment process if some sort of feedback exists, we have linked exchanges consisting of verbal contributions and feedback signals within the confines of dialogue moves. Perhaps the *ad-hoc* metaphor of "alignment moves" can shed light on the issues involved here.

Finally, linked representations for syntax, semantics and phonology, as well as the non-directionality of alignment (syntax does not dominate) serve as arguments for the use of constraint-based grammar or tools of comparable strength. Constraint-based grammars can in various ways be hooked onto paradigms using descendants of DRT. Hence, the alignment notion is in good agreement with modern semantics.

Taken together, the interactive alignment approach puts more emphasis on agents' coordination than on questions of well-formedness and coherence in dialogue, contrary to what has been customary practice in linguistics to date. After all, it is agents that coordinate, not structures.

3.1.3 New Ways of Experimenting with Dyads and Groups

Furthermore, a methodological point warrants mentioning. Following a seminal paper on the collaborative nature of reference (Clark & Wilkes-Gibbs, 1986), an increasing number of studies on verbal exchange in dyads or in groups with or without overhearers has employed techniques which open the door to an experimental investigation of communication (e.g., Matessa, 2003). For example, in an early study of alignment in task-oriented dialogue (Garrod & Doherty, 1994; 1995), the composition of dyads was systematically varied (stable vs. swapped pairings) – to the effect that under both conditions there was convergence with respect to communicative strategies; however, convergence was more pronounced with swapped pairings. In a similar vein, speakers in small groups align most with their respective predecessors, whereas in large groups they only align with the dominant person (Fay, Garrod & Carletta, 2000). Above all, such studies prove that it is feasible to do large-scale experimental research on dyads or groups. Beyond that, more in-depth investigations of the time course of individual

or joint processing are within reach since sophisticated chronometric measures on the basis of priming are available. Yet another popular line of research in the "visuo-linguistic behaviour" of interlocutors relies predominantly on spatiotemporal data from eye movement tracking in naturalistic settings (e.g., Henderson & Ferreira, 2004).

3.1.4 Alignment as a General Principle in Interaction

Considering the effectiveness of alignment as a mechanism for information transfer, one may wonder whether it can be regarded as a general principle in interactive behaviour. Indeed, alignment may not be purely or not even predominantly linguistic. Instances of alignment can be found at every biological level – from the neuronal level with oscillatory processes causing synchronisation of cell assemblies to the ecosystemic level with the complex interplay between predators and prey populations. In a similar vein, alignment appears to be at the very basis of a perception-behaviour link and to be responsible for all sorts of unconscious imitation.

From a biological point of view, alignment is a common principle in the communication of organisms. Even in communication as simple as that among squids or lower mammals, the information transferred is hardly ever complete. Rather, the information functions as a trigger to evoking some specific behaviour or emotional state, or to activating specific knowledge that extends beyond the information given. In addition to redundancy, the constructive nature of information processing is among the reasons why communication is so fast, efficient, and robust. So, communicating organisms sometimes show a complex interplay of aligned behaviours which could be misinterpreted as a planned action.

Alignment can even be tied down to the level of neurocognition. In psychophysiology, for example, investigating the coordination of perception and action in different domains, and based on observations of diverse effects in vision, bimanual coordination, imitation and joint action, a common-coding principle has been proposed which may in part be based on a mirror-neuron system (e.g., Sebanz, Knoblich & Prinz, 2003). By virtue of putting emphasis on shared representations, common coding bears some resemblance to the notion of alignment. From this point of view, learning in infants as well as in neurorehabilitation of motoric and language functions can be redefined in several aspects, ranging from simple motor functions over basic facial and vocal expression (e.g., Studdert-Kennedy, 2002) to complex social and emotional behaviour (e.g., Fonagy, Gergely, Jurist & Target, 2002).

3.2 Intrapersonal Alignment

So far, alignment has been worked out as a fundamental mechanism in interpersonal information exchange. However, any explanation of dialogue would be incomplete without taking into account the cognitive prerequisites for alignment on the side of each of the participants. Therefore, it is necessary to discuss alignment also with respect to intrapersonal information processing in conversation. To date, however, psycholinguistic research has almost exclusively covered monological language, focusing on either the production or the comprehension of utterances in an individual. An alignment-based approach thus constitutes a step towards bridging the gap that exists between the (dialogue-oriented) "language as action"-tradition and the (monologically oriented) "language as product"-tradition (cf. Tanenhaus & Trueswell, 2004).

3.2.1 The Deficiency of Monologue

While the classical monological approach enables researchers to investigate in great detail the various levels of representation as well as the time course and the flow of information in either language production or language comprehension, such an approach has one severe shortcom-

ing: It disregards any interactions that may take place between the interlocutors in a conversation. Therefore, a monological approach can give only a restricted picture of actual language use. Some aspects unavoidably remain unexplored – conceivably, the most vital ones, since speakers produce utterances with respect to (and sometimes with the help of) the recipients, and likewise, listeners take into account the speaker's intentions. In effect, the characteristics of the language used in dialogue are quite different from those of monological language. For instance, interactive alignment promotes the use of underspecified expressions and the development of dialogue routines during a conversation – aspects which escape the notice of an exclusively monological approach. Still, an examination of monologue from either the speaker's or the listener's side might prove fruitful in the endeavour to uncover regularities in language use – provided that monological expressions are viewed as an individual's contributions to some "joint project".

Thus, the notion of alignment in communication provides an opportunity to reconcile present-day dialogue-oriented research with monologue-oriented research in the psycholinguistic tradition. The rationale is that in order to develop an understanding of alignment between interlocutors, one must necessarily develop some understanding of the way in which information from diverse channels, in diverse modes, or of diverse formats gets aligned within an individual. Quite obviously, agents in communication have to be selective with respect to relevance: As speakers, they have to decide on what to express in which way in which channel; as listeners, they have to decide on assigning which weight to which information in which channel. Alignment – through priming or through routines – can be viewed as a means to facilitate language production or comprehension by automatising such decisions so that language users take a minimum of cognitive effort to make as much sense as possible (an aspect that Hörmann, 1976, referred to as "sense constancy").

3.2.2 The Mental Representation Issue

Intrapersonal alignment basically pertains to the alignment of mental representations within an individual either producing or comprehending language. With respect to language production, there are currently three major schools in psycholinguistics that differ in their views on issues of representation:

- According to the autonomous view (e.g., Levelt, 1999), utterances are produced incrementally in a strictly hierarchical fashion, i.e. by going through a fixed sequence of independent stages.
- According to the regulative view (e.g., Herrmann & Grabowski, 1994), utterances are produced incrementally in a heterarchical fashion, i.e. by going through feedback loops within a number of interconnected subsystems.
- According to the interactive view (e.g., Dell, Chang & Griffin, 1999), utterances are produced incrementally in a complex network structure, i.e. by spreading of activation among formal neurons in a constrained way.

With respect to comprehension, there are also competing schools:

- 'Bottom up' models (e.g., Kintsch, 1998) conceive of comprehension as a predominantly stimulus-driven, analytical process which results in a propositional mental representation of the discourse in question.
- 'Top down' models (e.g., Johnson-Laird, 1995) conceive of comprehension as a predominantly knowledge-based, constructive process which results in a comprehensive mental representation of the situation in question.

It should be noted that, regardless of the fundamental differences in design, all current theories in psycholinguistics assume that in the course of language processing, mental representations of diverse kinds play a role. These are also featured in the interactive alignment account (however, without any commitment to a particular theory of the format of mental representations).

At any rate, from a psycholinguistic point of view, the question of alignment brings up the question of how a language user comes to align the diverse mental representations that develop in the course of conversation. An ambiguous phrase like "a heavy smoker", for instance, requires recipients to align the literal interpretation suggested by the syntactic form with the non-literal interpretation which may be contextually appropriate (Frisson, Pickering & McElree, 2004). Similarly, it has been demonstrated that people easily arrive at a relatively clear-cut (and sharable) interpretation of vague quantifiers like "several", balancing various default assumptions about the situation described (Hörmann, 1983).

However, in the endeavour to partly reconceptualise language processing in terms of alignment of mental representations, one must keep two points in mind. First, a distinction is to be drawn between alignment and comprehension in general: While the former pertains to the establishment of implicit common ground, the latter embraces any form of grounding under rationality assumptions. Second, the term "alignment of representations" can be assigned a procedural and a structural reading. Therefore, any attempt to thoroughly investigate intrapersonal alignment will have to pay attention both to the process of aligning mental structures – the dynamics of alignment, so to speak – and to the state of mental structures aligned.

3.2.3 Alignment in Language Production and Comprehension

In the field of language production and comprehension, intrapersonal alignment may refer to such diverse matters as the relationships that hold…

- between representational structures at different levels,
- between representations of a verbal expression and its verbal context,
- between representations of a verbal expression and its nonverbal context.

Accordingly, the range of issues to be discussed with respect to alignment is rather wide.

Perhaps the most fruitful, but also the most difficult, area of study in language processing refers to verbal-representational relationships. It stands to reason that alignment in language processing is to be regarded as a two-way process: In language production, the speaker will typically align the elements of a verbal expression so as to match his or her currently activated knowledge (which includes assumptions about the audience). Whereas in language comprehension, hearers or listeners will typically align their mental representations so as to render them compatible both with the text processed and their general knowledge. This is a dynamic process: In order to attain or maintain comprehension, language users must continuously update their mental representation of the states of affairs. In the course of that alignment, various transformation mechanisms may be employed. One important mechanism is automatic inference, the generation of semantic information beyond what is explicitly given; a mechanism which – according to recent studies – can work in a more or less specific way (Mauner, Tanenhaus & Carlson, 1995). Another transformation mechanism is the transfer of a problem space from one domain to another, which – as the "structural alignment" approach maintains – enables people to deal with analogy and indirect language use as in metaphor or metonymy (Gentner & Markman, 1997). Finally, a third transformation mechanism is to ignore less relevant elements of a message or not to analyse them fully; a mechanism which relates the well-

known "Moses illusion" to alignment (Sanford, Barton, Moxey & Paterson, 1995). – On the other hand, on the side of the speaker, alignment, in the interest of making oneself understood, may result in omitting those elements from the verbal expression which readers or listeners can easily infer (a typical instance would be linguistic ellipsis), or, respectively, putting emphasis on important topics or being particularly thorough on less familiar topics (an example would be repetition). In addition to these automatised processes, speakers typically design their utterances on purpose so as to meet the requirements of the presumptive audience, their level of knowledge, their ways of speaking, and their communicative expectations (Clark, 1996).

As to verbal-verbal-relationships, we feel that the bulk of the psycholinguistic research on coherence lends itself to interpretation in terms of alignment. This holds for local as well as for global coherence. In this sense, local coherence may be viewed as the degree of alignability of neighbouring text elements. Hence, consecutive sentences could be regarded as coherent if they share referential, causal, temporal or local relations or are – at least – similar in structure (Gernsbacher & Givón, 1995). Global coherence, on the other hand, may be viewed as referring to the average degree of alignability of the text as a whole with the structure of a particular knowledge domain. Hence, a text could be regarded as coherent if its elements can be mapped onto a particular knowledge-based schema or scenario (Sanford & Garrod, 1998). Incidentally, researchers have recently proposed a statistical approach ("Latent Semantic Analysis") to measuring coherence of verbal expressions at the local and at the global level by computing what might be called alignability coefficients (Folz, Kintsch & Landauer, 1998).

Finally, as to verbal-nonverbal relationships, we believe that quite a few aspects of contextual dependency can be understood in terms of alignment. One of the best-studied areas is spatial reference. Spatial reference systems (which may be intrinsic, relative, or absolute) are a ready source of communication problems, and thus are subject to alignment or, in the case of major problems, negotiation. Moreover, in describing spatial environments, speakers tend to take an "imaginary tour", and the use of spatial terms such as *left* or *right* indicates that speakers in this case keep in alignment with the "imaginary tourist" (Avraamides, 2003). Another area of research on the interaction of verbal and nonverbal information touches on the processing of illustrated text or multimedia materials in general. Most of the current approaches to multimedia information processing assume that the verbal and the pictorial information is initially represented in separate, modality-specific formats. Later in the course of processing, these representations will have to be aligned in order to be, finally, integrated (Schnotz, Bannert & Seufert, 2002).

3.2.4 Relating Intrapersonal to Interpersonal Alignment

Altogether, a list of some of the phenomena that have a bearing on intrapersonal alignment reveals that there is a large intersection with phenomena that have been discussed with respect to interpersonal alignment and to coordination-by-negotiation in dialogue:

- In production, the use and re-use of referential terms; in comprehension, effects like the "repeated name" penalty in nominal coreference (Gordon, Grosz & Gilliom, 1993).
- In production, frequent use of underspecified expressions; in comprehension, employing a "good enough" criterion as a matter of course (Ferreira, Bailey & Ferraro, 2002).
- In production, the many degrees of freedom in conceptual combination; in comprehension, the ease of resolution of ambiguous compound words (Gagné, 2000).
- In production, frequent shifts of perspective in spatial descriptions; in comprehension, the tendency to do an "imaginary tour" (Avraamides, 2003).
- In production, the occurrence of tropes; in comprehension, their interpretation on the basis of similarity, according to the "structural alignment" theory (Gentner & Markman, 1997).

- In production, spontaneous gestures that accompany, emphasise and complement verbal utterances; in comprehension, their immediate interpretation (McNeill, 1992).

In consideration of this, negotiation and alignment are not to be taken as mutually exclusive strategies for coordination but rather as extremes on a continuum that leaves ample room for transition. In fact, scrutinising the way in which these strategies interact in natural language processing might be of particular relevance.

3.3 Simulating Alignment with Artificial Agents

A growing body of work in artificial intelligence, robotics, and agent research takes up questions relating to communication from a technical perspective. Originally a field of the study of intelligence by computational theories of symbol use, artificial intelligence has, over the past decade, undergone a paradigmatic shift toward the scientific study of artificial agents, which has led to new types of models for the study of communication. Cognitive robotics has begun to construct systems that illustrate how behaviours reproducing important abilities of natural intelligent agents can be implemented. The main goal of the CRC is to get a detailed understanding of cognitive abilities in general and of alignment in particular. It can be pursued by constructing both pure software agents and embodied agents acting in real or in simulated environments.

Embodied conversational agents (Cassell et al., 2000) or virtual humans (Gratch et al., 2002) are software entities that look and act like people. Such systems have expressive faces, limbs and hands and they can engage in conversation and collaborative tasks with human interlocutors. In order to align verbal and nonverbal utterance behaviour, it is, e.g., necessary to determine the specific spatial and temporal relations among modalities: speech-related gestures must closely follow the voice cadence. First attempts have been made to integrate these intrapersonal multimodal behaviours in computer-animated human models with sufficient articulation to effect both gross and subtle movements with visual acceptability and real-time responsiveness. The realisation of such synthetic agents engaging in interactive communication has also drawn attention to questions of interpersonal alignment in mixed-initiative dialogue, feedback signals, and turn-taking.

The design of human-machine interactions with robotic agents and virtual humans is of great heuristic value in the study of communication because it allows researchers to isolate, implement, and test essential properties of inter-agent communications. Creating and systematically modifying artificial systems that reproduce certain aspects of a natural system can help us to understand the internal mechanisms that have led to the particular results. From a basic research perspective, such operational models can advance our understanding of the key aspects of interactive alignment. From an application perspective, they are positioned to provide well-grounded support to enable "anthropomorphic" interfaces for assistance systems that are better tailored to human needs and expectations.

3.4 Alignment as a General Principle for Information Processing

Language is only one – but a very pronounced – example of a more general situation: the need to coordinate two or more very complex structures on the basis of a very narrow communication channel, or, in more neutral terms, a weak coupling. In technology, we encounter this situation in many forms: In consumer electronics, it is a general rule that the inner complexity of most products has to be hidden from the user. This requires crafting a suitable interface with operations that have to be simple and natural, yet sophisticated enough to give the human user sufficiently fine-grained control over the functions offered by the device. Examples

range from the (in-)famous video recorder to mobile phones, modern cars, PCs and – not yet really here, but on the horizon – user-friendly robots for domestic environments.

We view alignment as a far more general and powerful principle of organising the efficient cooperation of complex systems under the constraint of an only very restricted and "informationally narrow" coupling. (The apparent informational richness of language comes from its *effects* in our brains; per se, language has a very low bit rate compared to our other modalities, such as vision or touch, not to speak of the tremendous bandwidth of inner-brain couplings.) In pure informatics, we witness the development of toolkit libraries of ever-increasing size, facing us with the challenge to "hide" their control function in lean interfaces; in society, we see the formation of very complex organisations, and it is a major research issue which interaction structures are best suited for their effective management. And, looking for good solutions to these questions, we may hope to obtain inspiration from modern biology, which, from a particular perspective, may be viewed as a "science of ingenious interfaces", ranging in complexity from the surface of complex proteins over cell membranes to the amazing "informational interfaces" set up even by the brains of simple species.

We believe that, although these domains differ strongly in their specific details, alignment is a perspective that has significant potential to understand an important class of underlying dynamic structure for establishing functional coherence in complex, composite systems. It is the mission of this CRC to unravel these principles from the side of language research and its neighbouring computer science fields, but we will keep an open eye to cross-cutting developments in other fields to gain insights also into the more general and overarching *systemic* aspects of alignment as a general principle for information processing.

3.5 Summing Up

We feel that an alignment-based approach to the study of conversation is a viable alternative to the current paradigms in cognitive science. It is obvious that both coordination-by-negotiation and coordination-by-alignment are indispensable to successful communication, so that research will have to take into account both strategies, with special attention directed to their interface in natural dialogue. At the same time, the cognitive processes underlying alignment need to be treated on both an interpersonal as well as on an intrapersonal level, paying equal attention to the overarching aspects of interacting systems. Altogether, the notion of alignment is pioneering and innovative in character; it has much explanatory power, and thus opens a new, and potentially fruitful, avenue of research into cognition and into human-human and human-machine interaction.

4. The Research Agenda

Innovative as it is, the interactive alignment approach does not only open new perspectives on conversation; it also gives rise to a number of questions about the cognitive mechanisms at the core of alignment, mainly because the account is not yet very well developed (this is particularly true for its morphological, semantic, and pragmatic aspects). Above all, it remains to be shown that the interactive alignment approach is sound. It goes without saying that this cannot be accomplished with a single experiment or two; rather, such an endeavour requires extensive, systematic, and careful multi-facetted research. We believe that the proposed research initiative is worth the effort. As a prospect, such a research initiative will bring about valuable new insights into the functioning of human cognition and broaden our knowledge of the regularities that govern communication.

With respect to the notion of alignment, the research initiative we are proposing is aimed at two overall goals:

- First, to *substantiate the notion of alignment* as a pioneering explanation of natural language use in conversation by corroborating assumptions that require elaboration or empirical support, thus providing answers to questions yet unanswered.
- Second, to *extend the notion of alignment* as a general principle in information interchange by putting to the test the interactive alignment approach in situations that go beyond verbal conversation between humans.

Before introducing the research agenda in more detail, we shall present a brief overview of the topics to be addressed, and discuss the most important research topics against the background of the interactive alignment framework.

4.1 Substantiating the Notion of Alignment

With respect to substantiating the notion of alignment, we shall focus on two fields of research. One: the concept of *self-alignment* shall be investigated in the context of multi-modal data. Two: we propose that *attempts at a foundation* of the alignment approach be made in two ways: On the one hand, by scrutinising the central issues of compositionality and indirect interpretation, implicit common ground, and the role of situation models in interactive alignment; on the other hand, by "anchoring" the notion of alignment to a physiological substrate, to classical Gestalt ideas, and to theories concerning the interplay of focusing and attention.

4.1.1 Self-alignment

Figure 1 depicts two kinds of links; on the one hand, links between participants at various grammatical strata, and on the other hand, links within each participant that connect the respective levels. The links transport information via priming. Hence, every bit of information intentionally generated in dialogue can become operative on either the intrapersonal or the interpersonal level. Alignment between participants is further mediated by implicit common ground construction and so-called "dialogue inference", i.e. a sort of interactive check on which information can be considered as given and mutually accepted.

In Pickering and Garrod's (2004) view, intrapersonal alignment also provides the basis for an interesting model of self-monitoring, since speakers can control the information assembled on every level. The idea concerning self-monitoring applies to language production as well as to language comprehension: One can check the use of one's phonetic representation, one's concepts, their morphological and syntactic encoding, etc., or one can check one's understanding of meaning extracted from the morphological and syntactic material at hand. This is well confirmed empirically by the possibility of having repairs on all of these levels. Furthermore, it is at the core of the joint projects approach with respect to generating and maintaining information on these strata. – Starting with clarification questions concerning phonetic realisation, a parallel argument can be constructed for other-monitoring. Self-monitoring and other-monitoring operate on the same representational level due to the interactive construction of implicit common ground.

Reliable experimental data concerning intrapersonal alignment susceptible to statistical treatment are still sparse. A suitable field of investigation is provided by co-ocurrences of speech and concomitant spontaneous gesture. We suggest extending the notion of intrapersonal alignment by taking into account the relationships between verbal expressions and co-verbal deictic and iconic gesturing **(project section B1)**. Research on temporal and semantical relations between speech and gesture has shown that speakers coordinate gesture stroke with

phonetic stress and syntactic, semantic or pragmatic information; a case in point being e.g. the coordination of perspective or focus (i.e. genuine pragmatic information) and stroke. These findings are essential for building up an integrative, yet still missing multi-modal theory of speech and gesture. The same is true for the corresponding simulation with embodied conversational agents (ECA) in virtual reality (VR). Congenial to attribute-value-matrix accounts in theories of grammar, VR simulation can work with an inventory of image description features. The semantics of description features can then be unified with linguistic feature information on arbitrary levels of composition. As a consequence, we can model fine-grained alignment between gesture and speech. Gesture-speech alignment can finally feed into a self-monitoring model for multi-modal expressions, a model that is well motivated by corpus data showing that gestures can be corrected by gestural or verbal means.

4.1.2 Broadening the Foundation

Interactive alignment presents a novel alternative to psycholinguistic and linguistic explanations of production and comprehension processes in dialogue, with emphasis on thoretical parsimony. Not surprisingly, the empirical backing for its key concepts varies considerably. In the following, we shall address a few concepts which, in our opinion, require substantiation.

One of the questions that awaits an answer is that of the physiological substrate of alignment. If alignment is indeed to be conceived of as an ensemble of automatic processes, coarse measures like processing time or choice reactions are not likely to yield any quantitative evidence of service. Instead, a promising way to study the time course of alignment at various levels is to employ neurophysiological methods such as electroencephalography **(project section B4)**. Though limited to the study of monologue for technical reasons (simultaneous measurement of two interlocutors would probably result in inextricably confounded data), measurement of event-related potentials may enable researchers to empirically distinguish between alignment processes at different levels. In doing so, spectral analytic techniques provide a means to disentangle parallel processes. The promise of brain imaging techniques is even greater **(project section B7)**. Thus, the investigation of neural correlates of alignment processes may ultimately lead to an understanding of alignment that goes substantially beyond what can be figured out by the observation of overt behaviour.

Another point of interest concerns the cognitive mechanisms at the bottom of alignment. With priming listed as the short-term mechanism and routinisation as the long-term mechanism of prime importance in alignment, these two mechanisms are left largely unexplained. In face of the many forms that priming can take (from phonological or syntactic to semantic or cross-modal priming), and because of the close relationship in kind between routinisation on the one hand and repetition, imitation and learning on the other, it is advisable to study more closely the cognitive prerequisites of those mechanisms. In doing so, classical Gestaltist notions like proximity, similarity, and Prägnanz (which easily transfer from vision to language) provide a promising conceptual background **(project section B5)**. The advantage of such notions is that they can accommodate both generalisation and expansion – two fundamental principles that govern language processing and that come to bear in alignment-related phenomena like underspecification and automatic inference.

One of the functions fundamental to communication – one that is addressed in detail by Clark (1996) – is attention-getting and, in due course, attention maintenance. In order to attain common ground, interlocutors must establish a joint focus of attention. In one respect, the term "attention" can be given an intra-individual interpretation, referring to giving weight to or selecting among the available channels of information, or to the relative relevance assigned

to mental representations at various levels. From such an intrapersonal point of view, misalignment of attention should become most obvious in the case of conflicting information. In another respect, however, the term "attention" can be viewed as an inter-individual concept. From such a point of view, the first interpretation that comes to mind is a topological one, referring to those referents that interlocutors currently attend to. In a broader sense, joint attention relates to the degree of similarity of the interlocutors' mental representations (e.g., in terms of the current distribution of activation in an "attentional landscape"). An empirical investigation of attentional alignment is possible by systematically varying the available information. Such an investigation will provide data on specific conditions of alignment, i.e. on necessary and sufficient information **(project section A6)**.

At present, the bulk of the data cited in favor of the interactive alignment approach comes from research on syntax-in-dialogue. There is much less empirical foundation concerning the lexicon, morphology, semantics or pragmatics. In face of this, we suggest corroborating the alignment approach in two respects **(project section B2)**: Starting with the investigation of nominal compounds and their underspecified relational semantics, we propose to test alignment with respect to compositional morphology. In other words, we plead for an extension of the alignment approach to a level below the word. Further corroboration is bound up with indirect interpretation and its Gricean resolution. Since compounds frequently encapsulate literal as well as non-literal meaning, we need tools for indirect interpretation below the word level as well. As an alternative to Gricean metonymy resolution via maxims we will put to use the notion of routinisation. Routines are expected to provide the conventionalised meaning of certain classes of metonymies. Routines will also be used to treat matters of indirect interpretation at the level of dialogue moves or within dialogue games (speech acts, rhetorical relations, or enthymemes).

Information exchange in dialogue is, according to the interactive alignment approach, regulated via the construction of implicit common ground (and, if so required, full common ground). Implicit common ground—a structure containing purely factual information—is easily and quickly searched and updated. As a first attempt, it might therefore be encoded in a non-modal DRT-format or as an information state lacking private information. In contrast, full common ground contains propositional attitudes of various sorts and is therefore paired with inference across fully developed mental states. Employing empirical studies, VR simulation and formal modelling, we shall work on a theory of implicit common ground and will thus contribute to the foundation of the central notion of the interactive alignment approach on which much else hinges. One special matter of interest in this context will be the development of information states using hybrids of game-theoretical and logical structures **(project section B3)**.

Fundamental work will also be done regarding situation structure and situational alignment. Although situations are considered the motor of alignment, it is not entirely clear how situation models should be conceived of. Here, the contribution of perceptional information to situation models, their constitutive parameters, the relations and constraints (concerning e.g. spatiotemporal and causal relationships) that exist among these are the main focus of interest. Situation models are at the very basis of the attunement process, and if the information on this matter is aligned, it must enter implicit common ground, thus enabling appropriate verbalisation **(project section A5)**.

4.2 Extending the Notion of Alignment

With respect to extending the notion of alignment beyond the ideas developed so far, we shall address three aspects, namely, sub-dialogue levels of linguistic analysis, alignment in dialogue,

and, finally, constraints on alignment. Levels of grammar serve as tracks for alignment. As to these, we will study the phonetic preconditions of alignment, investigate matters of compositionality and indirect interpretation at the word level and below, and discuss syntax-prosody mapping under alignment conditions. *Alignment in dialogue* focuses on routinisation and on the routines established, bringing to the foreground aspects like patterns of dialogue structure, the spread of dialogue games in dyads and groups, and reformulations and other-repairs as routines central for dialogue. *Constraints on alignment* may arise due to characteristics of the individual interlocutors or to properties of particular types of channels. As to constraints of an individual nature, the relevance and limits of memory in modelling partners and the influence of language impairment on syntactic or emotional alignment will be investigated. As to constraints of a more technical nature, the topology and multi-dimensionality of the interaction space emerging in face-to-face dialogue, the feasibility of studying alignment in nonverbal exchange, and the difficulties imposed on alignment by the different channels in multimedia communication are on the research agenda.

4.2.1 Focusing on Specific Linguistic Levels

According to the by now familiar picture, priming processes percolate up and down intrapersonal and interpersonal linguistic grids. As modelled in constraint-based grammar, information introduced at any point can then distribute freely in the graph-like structure. One of our intentions in extending the notion of alignment is to put those levels of linguistic analysis on the agenda that warrant closer inspection.

This holds true for phonetic and phonological information. There is ample evidence in favour of the assumption that speakers align their phonetic representations: In dialogue, articulation becomes reduced among speakers; accents and speech rates become attuned. These clues concerning phonetic alignment can be used in the construction of an automatic system for the perception and production of speech. Its main source of inspiration will be Pickering and Garrod's (2004) idea about input-output coupling at the phonetic level (which so far has not been implemented in automatic systems). Another research line to be pursued in this context will be to substantiate the idea of self-monitoring at the phonetic level **(project section B6)**.

Syntactic and prosodic maps have been a topic of long standing in descriptive linguistics; to date, however, these have not been treated in terms of alignment. When mapped onto the interactive alignment scheme, it is evident that any attempt to do so will have a production and a comprehension side, subject to the principle of input-output parity. The pursuit of the alignment notion in prosody will be of prime interest since this is a domain still in demand of empirical data and appropriate models **(project section A3)**.

Attempts to ground the alignment approach with respect to morphology and pragmatics, which come down to extending alignment to the sub-word level, have been discussed elsewhere **(project section B2)**.

4.2.2 Alignment in Dialogue

Routines, script-like structures, non-compositional semantic entities and the like have been frequently looked upon with suspicion. However, it is clear today that one cannot seriously doubt their existence and that one cannot do without them. As argued above, we will put emphasis on research on routinisation and routines in various respects. In terms of alignment, routines exist with respect to morphology, the lexicon, collocations, referential expressions, indirect speech acts and rhetorical relations connecting dialogue moves or even dialogue games. In consideration of the frequency of their occurrence in natural conversation, routines should

hence be investigated statistically on a larger scale with a larger corpus of annotated dialogue. Dialogue modelling techniques should be applied on top of descriptive techniques in order to identify patterns of alignment for preferred dialogue moves and dialogue games. Regarding both routines and dialogue games, the field of investigation of alignment will be extended far beyond its present state **(project section C2)**.

There is evidence that the dialects of dyads engaged in task-oriented conversation converge with respect to the use of referential expressions and syntax. Similar observations have been made with respect to morphology, the lexicon, semantics, and pragmatics: Particular styles of compounding do emerge in the course of conversation, neologisms and lexicalised expressions are introduced with special readings barring polysemy, reduced forms gradually substitute long definite descriptions or complex predications, metonymy underspecified is automatically fixed, fragments and ellipses do the job of indirect speech acts, the meaning of whole dialogue games is condensed in referential or predicative expressions, there is frequent reference to the domain's ontology, to situations, actions or results. It goes without saying that we are dealing with routinised structures here as well, sometimes atomic, sometimes of considerable complexity. With larger groups, essentially the same effects, namely, convergence of dialects, have been observed. Given the state of the art in evolutionary linguistics, the systematic investigation of language use in larger groups is of interest for various reasons: On the one hand it shows how conventions of language use are established; on the other hand it shows how preferences for particular verbal structures such as compounds or metonymies emerge, spread and stabilise **(project section C3)**. Adhering to the distinction between cooperation-by-negotiation and cooperation-by-alignment, we might ultimately be able to tell which conventions are determined by rational factors like Gricean maxims and which are due to purely mechanistic "spread". This will also foster the development of hybrid modelling techniques that combine both rational and mechanistic elements.

Among the routines to be investigated, repairs and reformulations in dialogue will have a prominent place. Both can be viewed as interactive techniques that serve to establish and maintain common ground. The investigation of these routines **(project section C5)** will hopefully also shed light on the difference between implicit and full common ground. Intuitively, other-repairs appear to be on the negotiative side (and hence closely intertwined with the attribution of plans, mental states, intentions and the like), but they also might display features of a more automatic nature (such as a repair indication, respecting the boundaries of structures, re-cycling non-repaired material, indicating the scope of repairing, and acknowledging or denying repair attempts). If this is true, it indicates that at the bottom of complex conventions involving mental states we may have a fairly mechanistic stratum. The picture which then might well emerge is that higher order structure is grafted onto more basic proceedings. Such an outcome would be a substantial contribution to our understanding of the preconditions of negotiative structure.

4.2.3 Constraints on Alignment

Proponents of an alignment-based approach argue that dialogue, unlike any other form of verbal communication, is characterised by interactive alignment. Unfortunately, however, the distinction between dialogical and monological conversation is not very well specified – in particular, when it comes to accommodating multimodal communication, mediated communication, and communication within larger groups. In consideration of this, it may be appropriate to conceive of the dialogue-monologue-distinction as a gradient rather than as a dichotomy. As a consequence, it is obvious to pay attention to intermediate forms of verbal communication. In other words: Imposing specific constraints on communication may yield deeper insight into the

nature of alignment, in particular, into necessary and sufficient conditions, and hence, make headway towards clarifying what exactly distinguishes dialogue from monologue.

One obvious direction that such an extension of the original interactive alignment approach could sensibly take would be to study constraints that lie in the nature of the agents who engage in conversation. Individuals may differ in their information processing capacity, they may differ in their ability to develop or employ particular routines, in their sensibility to different types of information, in their efficiency in aligning representations, and in many other respects. Such constraints will not only influence dyadic conversation; they will likely take an effect also on the communicative behaviour within larger communities (and thus, on the utilisation of routines). The extent to which interlocutors rely, e.g. in modelling partners, on episodic or semantic memory (by virtue of shared experience or social attitudes) is a point in question. Pertinent evidence suggests that memory is an important factor which merits investigation **(project section A1)** – both under quantitative aspects (e.g., as to the duration of memory effects) and under qualitative aspects (e.g., alignment of contradictory information). Another point in question concerns any constraints on alignment that are due to language impairment on the side of (one of) the interlocutors. This pertains both to the communication of handicapped people as well as to the communication with handicapped people **(project section A4)**: On the one hand, one might want to study the alignment capabilities of, say, aphasic patients; on the other hand, one might want to investigate the consequences of impairment for the people in the environment of aphasics. At that, denotative and connotative (or, more generally, emotional) aspects might likewise play a role. For this reason, the emotional dimension lends itself to study within the interactive alignment framework **(project section C4)**. If, as a result of pertinent studies, specific deficits in alignment could indeed be traced back to specific impairments, this would provide far-ranging clues about which level of representation exerts which function in conversation.

Another, equally plausible direction that an extension of the original interactive alignment approach could take would be to study constraints that are to be attributed to the various communication channels. These are constraints of a more technical kind. One such restriction is of a topological nature. It concerns the interaction space of the interlocutors **(project section C1)**. As to that, the question arises if it is possible to identify sub-spaces that license different modes of alignment with respect to verbal and nonverbal behaviour (e.g., the use of deictic or iconic gestures that complement underspecified expressions). For a naturalistic, cognitively adequate model of face-to-face conversation, it would be essential to know the factors that define such a space and its sub-spaces. Another, related, constraint ensues from extending the original interactive alignment approach to nonverbal behaviour **(project section A2)**. One may want to ask which forms of alignment are common to the verbal and the non verbal domain and which, in contrast, are specific to the one or the other. More precisely, one might ask if the regularities that hold in the nonverbal domain are roughly the same as those that hold for dialogue. Likely, there are instances of alignment of posture, gesture, facial expression, or gaze by means of priming and routinisation to be observed in nonverbal exchange – but then, what exactly do they contribute to successful communication? Last but not least, the extension of the notion of alignment from the domain of dialogue to the nonverbal domain gives rise to an in-depth investigation of the interactions between domains. A reasonable way to proceed is to systematically vary the constraints that hold in a given situation – from largely unconstrained face-to-face dialogue via the introduction of increasingly severe constraints on the visual channel (simulating video-mediated communication as in, say, video conferencing) to disabling any visual feedback (thus simulating telephone conversations or internet chat) – and to study accordingly the effects of this manipulation on alignment as manifested in the verbal expres-

sions **(project section A6)**. In such a way, it will be possible to empirically assess both the interface between cooperation-by-alignment and cooperation-by-negotiation as well as the gradual transition from dialogue to monologue.

4.2.4 Alignment and Artificial Systems

While language has been "crafted by evolution" as an ingenious complexity-hiding interface into our brains (superbly successful to the extent that even for accessing our own brain we rely to an amazing extent on "inner speech" – impressively, by the way, illustrating the two inter- and intra-facets of alignment), artificial systems challenge us with the task to newly devise similar facilities, so that they can be "connected" to human users (or among themselves) as seamlessly as possible. With the background of the previous sections it is easy to recognise this challenge as nothing less than calling for the understanding, design and, finally, replication of mechanisms that can support an *extended* and *generalised form of alignment* between two or more agents.

Focusing on the case when one of the interacting parties is a human, we may from this generalisation of alignment immediately draw several benefits for our research perspective:

- One party being an artificial system, we are free to scale its complexity within wide bounds. This offers us a means to study simplified and restricted forms of alignment in a controlled way.
- In particular, we face issues of alignment also in non-verbal, e.g., visual or motor domains. Insights into these generalised forms of alignment can be translated back for the human-human case.
- A further advantage results from the full accessibility of the artificial communicator (as compared to a human one). This opens interesting avenues to study alignment dynamics, at least in its "visible half".
- Last but not least, generalisations of alignment suggested from human-machine interaction situations will be of significant application value.

These perspectives have motivated the inclusion of quite a number of projects located at the human-technology interface: The **project sections C7** (Semantic Alignment for search engines) and **C8** (Alignment in Cooperative Search) address the situation when the machine-part is a large database system, either in the form of a web-engine or in the form of a large database. While the first project focuses on the verbal domain, framing the Turing Test as a special (and very important!) instance of an alignment task, the second project addresses the complementary modality of vision, considering alignment as an organising principle to efficiently support interactive search for pictorial material in a database. With regard to attentional and emotional processes in mixed, human-machine settings, one may argue that embodiment matters. The **project sections A1, B1, B3**, and **C1** employ a virtual humanoid agent using advanced VR-rendering techniques to approach alignment at such levels, as well as to study and replicate alignment processes in multimodal communication. In contrast, **project sections A2** and **C6** use a set-up involving a human partner and a humanoid robot to investigate alignment processes in real-world settings. **Project section C1** will take an integrated view and will be concerned with both a virtual human as well as a robotic agent. While these projects are largely focused on communication in the verbal and visual domains, e.g., in **project A2** in conjunction with eye-tracking techniques, **project section C6** generalises alignment into the domain of action (including the important aspect of adaptivity): Here, the aim is to study mechanisms that can make cooperative physical actions between a human and a robot partner efficient on the basis of again only a weak coupling given by the situational context together with dialogue and sensori-motor interaction.

References

Avraamides, M.N (2003). Spatial updating of environments described in texts. *Cognitive Psychology, 47*, 402-431.

Barr, D.J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science* (in press).

Bühler, K. (1934). *Sprachtheorie. Die Darstelungsfunktion der Sprache*. Jena: Fischer.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (eds.) (2000). *Embodied Conversational Agents*. Cambridge, MA: MIT Press.

Clark, H.H. (1996). *Using language.* Cambridge, UK: Cambridge University Press.

Clark, H.H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39.

Coventry, K.R. & Olivier, P. (Eds, 2002). *Spatial Language: Cognitive and Computational Perspectives.* Dordrecht: Kluwer.

Dell, G.S., Chang, F. & Griffin, Z.M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science, 23,* 517-542.

Fay, N., Garrod, S. & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science, 11*, 487-492.

Ferreira, F., Bailey, K.G.D. & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11-13.

Folz, P.W., Kintsch, W. & Landauer, T.K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes, 25*, 355-361.

Fonagy, P., Gergely, G., Jurist, E.L. & Target, M. (2002). *Affect regulation, mentalization, and the development of the self.* New York: US Other Press.

Frisson, S., Pickering, M.J. & McElree, B. (2004). *Global and local influences on adjective-noun combination: A limit on incrementality.* Manuscript submitted for publication; available from URL: http://ila.psych.nyu.edu/users/bdm/Mans_PDF/adjnoun.pdf

Gagné, C.L. (2000). Relation-based combinations versus property-based combinations: A test of the CARIN theory and dual-process theory of conceptual combination. *Journal of Memory and Language, 42*, 365-389.

Garrod, S. & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition, 53,* 181-215.

Garrod, S. & Doherty, G. (1995). Special deteminants of coherence in spoken dialogue. In G. Rickheit & C. Habel (Eds.), *Focus and coherence in discourse processing* (pp. 97-114). Berlin: de Gruyter.

Garrod, S. & Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8,* 8-11.

Gentner, D. & Markman, A.B. (1997). Structural alignment in analogy and similarity. *American Psychologist, 52,* 45-56.

Gernsbacher, M.A. & Givón, T. (Eds., 1995). *Coherence in spontaneous text.* Amsterdam: Benjamins.

Gordon, P.C., Grosz, B.J. & Gilliom, L.A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17,* 311-347.

Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., & Petajan, E. (2002). Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems,* July/August 2002 (2 1094-7167/02).

Henderson, J.M. & Ferreira, F. (Eds., 2004). *The interface of language, vision, and action: Eye movements and the visual world.* New York: Psychology Press (in press).

Herrmann, T. & Grabowski, J. (1994). *Sprechen. Psychologie der Sprachproduktion.* Heidelberg: Spektrum.

Hoc, J.M. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human-Computer Studies, 54*, 509-540.

Hörmann, H. (1976). The concept of sense constancy. *Lingua, 39,* 269-280.

Hörmann, H. (1983). The calculating listener – or: How many are *einige*, *mehrere*, and *ein paar* (*some*, *several*, and *a few*)? In R. Bäuerle, C. Schwarze & A. v. Stechow (Eds.), *Meaning, Use, and Interpretation of Language* (pp. 221-234). Berlin: de Gruyter.

Johnson-Laird, P.N. (1995). Mental models, deductive reasoning, and the brain. In M.S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 999-1008). Cambridge, MA: MIT Press.

Kintsch, W. (1998). *Comprehension – a Paradigm for Cognition.* New York: Cambridge University Press.

Levelt, W.J.M. (1999). Producing spoken language: A blueprint of the speaker. In C.M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 83-122). Oxford: Oxford University Press.

Lewis, D. (1969). *Convention: A philosophical study.* Cambridge, MA: Harvard University Press.

Matessa, M. (2003). Measures of adaptive communication. In S. Weibelzahl & A. Paramythis (Eds.), Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems. Available from URL: http://human-factors.arc.nasa.gov/ihi/cognition/

Mauner, G., Tanenhaus, M.K., & Carlson, G.N. (1995). Implicit arguments in sentence processing. *Journal of Memory and Language, 34*, 357-382.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought.* Chicago: University of Chicago Press.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue [with discussion: The interactive alignment model – developments and refinements]. *Behavioural and Brain Sciences* (in press).

Rickheit, G. (2004). Alignment und Aushandlung im Dialog. *Zeitschrift für Psychologie* (in press).

Sanford, A.J., Barton, S.B., Moxey, L. & Paterson, K. (1995). Cohesion processes, coherence, and anomaly detection. In G. Rickheit & C. Habel (Eds.), *Focus and coherence in discourse processing* (pp. 203-213). Berlin: de Gruyter.

Sanford, A.J. & Garrod, S.C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes, 26,* 159-190.

Schnotz, W., Bannert, M. & Seufert, T. (2002). Toward an integrative view of text and picture comprehension: Visualization effects on the construction of mental models. In J. Otero, J.A. Leon & A.C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 385-416). Mahwah: Erlbaum.

Sebanz, N., Knoblich, G. & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition, 88*, B11-B21.

Steels, L. (2002). Grounding symbols through evolutionary language games. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 211-226). London: Springer.

Studdert-Kennedy, M. (2002). Mirror neurons, vocal imitation, and the evolution of articulate speech. In V. Gallese & M.I. Stamenov (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 207-227). Amsterdam: Benjamins.

Tanenhaus, M.K. & Trueswell, J.C. (Eds., 2004). *World-situated language processing: Bridging the language-as-product and language-as-action traditions.* Cambridge, MA: MIT Press (in press).

Cooperation Matrix between the project sections

	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6	B7	C1	C2	C3	C4	C5	C6	C7	C8	X1	X2
A1		•			•	•	•		•			•	•	•		•	•	•	•				•
A2	•				•	•	•						•	•					•				
A3				•			•			•		•	•		•			•				•	
A4			•				•			•		•	•		•			•				•	
A5	•	•				•		•	•		•			•		•		•			•	•	•
A6	•	•			•				•					•		•			•			•	
B1	•	•	•	•					•			•	•	•			•		•			•	•
B2					•				•		•							•		•		•	
B3	•				•	•	•	•			•				•	•	•		•	•		•	•
B4			•	•									•		•		•	•				•	
B5					•			•	•								•	•			•	•	
B6	•		•	•			•						•		•				•			•	•
B7	•	•	•	•			•			•		•		•	•		•		•			•	
C1	•	•			•	•	•						•		•		•			•	•		•
C2			•	•					•	•		•	•	•		•						•	
C3	•				•	•			•						•			•					
C4	•						•		•	•	•		•	•					•			•	•
C5	•		•	•	•			•		•	•					•						•	
C6	•	•				•	•		•			•	•				•			•	•		•
C7									•					•					•		•		•
C8					•						•			•					•	•			•
X1			•	•	•	•	•	•	•	•	•	•	•		•		•	•					•
X2	•				•		•		•			•		•			•		•	•	•	•	

Project area A: Interpersonal aspects of alignment

Interpersonal communication happens in a multidimensional, multi-modal dynamic system, which complexity seems to overtax each interlocutor. At least linguists and psychologists have been wondering for decades, how communication can function successfully, when so many different aspects from different modalities have to be combined in a subtle way in a very short time. A very differentiated view of language use is formulated by Clark (1996), who stresses the social purposes and the multidimensional character of language use. He treats language use as a species of joint action. In this way, one has to be careful about treating phenomena of joint activities only as characteristics of language use. As important phenomena he stresses for instance coordination, cooperation, conventions and the accumulation of common ground.

It remains a matter of debate, how precise information is given and how well it is really understood by the interlocutors in a social dialogue situation. But several findings from SFB 360 showed that the case seems to be not as bad, as it was thought of by some theoretical positions. While research focused on explicit processes of establishing meaning and constructing a shared mental model for the given task, several findings showed up, that can be taken as hints for a very fast and spontaneous coordination of the communication partners.

Communication is functional – and it seems to be even easier, if both interlocutors are present in the communicative situation in a shared space. Project section A asks, why this can be the case. What makes communication easier, faster or more successful, if the communication partners have the chance to be present in the same place at the same time when communicating about the situation or even about topics not represented in the communicative setting. Some basic ideas for this facilitation are formulated by Pickering and Garrod (2004) and can be summed up by the concept of interpersonal processes of alignment, or as the authors have put it recently, as *interactive alignment*. In addition to these automatic and routinised processes contributing to the development of communicative common ground, in many cases human communication involves the construction of certain types of partner models. These models may apply to specific individuals, certain social groups, generic expectations, or even to egocentric generalisations (Keysar, 1997). Modeling partners therefore will be of central relevance for the study of interpersonal alignment as the main topic of Project area A.

In more detail, Project area A is concerned with processes of alignment on different levels of language comprehension and production. It discusses factors that may influence the "amount of alignment" taking place for verbal and nonverbal components of communication. Several project sections analyse in more detail personal, situational and thematic variations and limitations of alignment aiming at a better understanding a) of the cognitive processes underlying alignment, b) of the structures being aligned and c) of factors enhancing or limiting alignment. In this way, each of the following projects contributes interesting aspects for different levels of language processing, for limitations or enhancing factors of the situation or of the interlocutors.

Project section A1 is generally concerned with modeling partners on nonverbal and verbal levels of communication. A2 is concerned with processes of aligning gestures and facial expression in situations, demanding a coordination of actions or a discussion of some spatial constructional task. Projects A3 to A6 concentrate on different levels and modi of verbal coordination and processes of aligning verbal structures and conceptual representations. All projects are based on experiments and/or empirical data from studies on "normal" or "impaired" communication including systematic variation of relevant factors. Some projects especially use simulation and

computer modelling as an additional tool for theoretical abstraction. In more detail the projects are concerned with processes of interpersonal alignment in the following sense:

In Project section A1 (Modelling partners; Wachsmuth & Rickheit), some fundamental questions of interpersonal alignment will be addressed: How are adequate partner models developed, which take into account implicit effects of autobiographical memory? Processes will surely rely on implicit memory representations, including information about space, time, causality, and intentionality. Alignment in interactions with human and artificial agents will be analysed.

Project section A2 (Alignment in visual communication; Kummert, Ritter & Wachsmuth) concentrates on mimic and facial expressions accompanying dialogue about e.g. controversial film sequences or episodes. Alignment of intra-personal information processing and interpersonal routines will be analysed.

Project section A3 (Syntactic-prosodic alignment in language processing and communication; Kindt & Weingarten) is concerned with alignment of prosodic and syntactic structures with special emphasis on the coordination of prosodic and syntactic forms and functions.

Project section A4 (Syntactic alignment in normal and impaired communication; Hielscher-Fastabend, Sahel & Eikmeyer) takes well-documented syntactic priming effects as a basis and compares these effects in the context of natural dialogue and experimental communicative settings. Effects in especially handicapped groups (aphasic adults, second language learners, children with specific language impairments) will be analysed to learn more about the underlying functions of the effects and to formulate a model of grammatical encoding and decoding in communication.

Project section A5 (Alignment of perceptual and relational components of situation models; Wachsmuth, Sichelschmidt & Rickheit) tackles the question of how alignment processes at "low levels" (word retrieval, syntax) lead to alignment at the level of the situation model. Empirical studies and computational modelling focus on the interplay between the visual perception of a situation and low level language processes leading to the construction of shared situation models through conversation.

Project section A6 (Alignment of attention in mediated communication; Ritter & Sichelschmidt) aims at investigating functional aspects of attentional alignment in multi-modal environments. This project analyses visually mediated communication in comparison to face-to-face communication by using the eye movement paradigm, which will help to distinguish diverse modes of alignment and their facilitative potential.

References

Clark, H.H. (1996). *Using language.* Cambridge, UK: Cambridge University Press.

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes, 24,* 253-270.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* (in press).

Project section A1: Modelling partners

Ipke Wachsmuth, Gert Rickheit

Motivation

When persons communicate, they constantly evolve models of their communication partners, covering aspects on different levels and being partial in that only relevant aspects are stored. The project A1 is concerned with establishing and maintaining individual partner models that make use of implicit memory recall in order to allow for alignment processes. Partner models comprise mechanisms and representations which are crucial for adaptation to the situational context and they play an important role for effective conversation. Traditionally, explicit partner models have been investigated, with logical inference used to reason about the interlocutor's intentions and beliefs. But as these processes tend to be effortful, we want to investigate how better and faster alignment between familiar communication partners can take place, e.g., by implicit, automatic activation processes of an autobiographical memory.

Objectives

In this project, we address high-level, cognitive dimensions of alignment phenomena arising in the context of a conversation and especially with familiar persons. We want to investigate face-to-face situations with up to three interlocutors. In this context two aspects are relevant: On the one hand, partner models on *general aspects* that can be applied to any communication partner and in which conventions such as turn-taking and focus of attention are important; on the other hand, partner models on *individual features* which help to align with interlocutors previously known. A typical example for the latter is the activation of the interlocutor's anterior intentions from a similar situation. This cognitive process aids in anticipating what the interlocutor is now referring to, thinking about, or trying to achieve. To this end, the project aims at designing and implementing an individualised partner-model which takes into account implicit effects of an autobiographical memory.

Methods

The project will be carried out in close cooperation between informatics and psycholinguistics using the *empirical-simulative* method. Psycholinguistic experiments will be conducted in a variety of settings to examine relevant factors leading to fast alignment among familiar persons. In face-to-face conversations with up to three interlocutors, we want to evaluate the degree of alignment between the individual conversational partners subject to their familiarity. Findings will be applied to a partner model implemented in the virtual embodied conversational agent "Max" in iterated design steps and extensions of the model. Simulations will build on a *belief-desire-intention* architecture developed for Max in previous work. It provides a good premise to model intentional behavior resulting out of goal attainment or failure, involving emotional reactions. As autobiographical memories are often seen primarily as records of success and failure in goal attainment, this appears to be a good starting point for memorising events. An emotion module, an episodic short-term memory as well as a turn-taking module (Leßmann et. al., 2004) developed in the SFB 360 are resources to build on.

Work programme

The conception of the autobiographical memory will be based on work by Conway and Pleydell-Pearce (2000) who describe how pre-stored event-specific knowledge influences the cognition process, either directly, or indirectly in the form of priming processes. For storing situations and events, *situation models* provide a good starting point. Pickering and Garrod (in press) assume that in a successful dialogue, interlocutors develop aligned situation models. As

Zwaan and Radvansky (1998) point out, situation models may serve as a framework for autobiographical memory. In order to enable the activation of events shared by the interlocutors, identification of the conversation partner is necessary and thus included in the work programme:

- development of a *general* partner model comprising turn-taking and focus of attention,
- investigations on partner recognition on the basis of visual and aural stimuli,
- development of structured representations of situation models in the course of dialogue to enable memorisation of *individual aspects* in an autobiographical memory,
- investigation of situated recall of autobiographical memory content.

Insights from the empirical studies that lead to the revision of existing models will be incorporated inline in the technical Max model. Advancements in the alignment of Max with familiar interlocutors will be a good indicator of progress of the project. Furthermore, the project will also contribute to the development of motor-level alignment processes in collaboration with the project C1 "Interaction space". Further collaboration is planned with project B3 "Construction of implicit common ground" to investigate how the partner model can contribute to solutions to problems concerning the interface of implicit common ground and full common ground.

References

Conway, M. A. & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*(2), 261-288.

Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.

Leßmann, N., Kranstedt, A., & Wachsmuth, I. (2004). Towards a cognitively motivated processing of turn-taking signals for the embodied conversational agent Max. In *Proceedings of W12: Embodied Conversational Agents: Balanced Perception and Action, AAMAS 2004*, pp. 57-65.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Student assistant with 19 hours per week

Project section A2: Alignment in visual communication

Franz Kummert, Helge Ritter

Motivation

Non-verbal modalities of human communication that are visually perceptible such as eye gaze, (body) posture, face mimics and gesture are indispensable for the understanding of human social behavior. The use of non-verbal modalities allows humans to express emotions, sensations and moods in interpersonal communication and interaction. Communication partners can, for example, exchange information about the nature of their relationship and express consent or dissent with each other. Furthermore, non-verbal communication modalities may support verbal utterances with respect to the fulfillment of a common goal, for example, during joint problem solving. Thus both emotional and pragmatic factors characterise non-verbal communication.

Alignment in non-verbal communication occurs at an intra-personal as well as an interpersonal level. At an intra-personal level, the different modalities are internally adjusted so as to achieve coherency in the communicative behavior. At an inter-personal level, the visually perceptible communication behaviour of one communication partner determines one's own visually perceptible communication behaviour and vice versa. Additional factors such as social position, gender or the communication environment may also affect inter-personal aspects of non-verbal communication.

Objectives

During the first phase of the project, we will focus on the coordination of eye gaze and deictic gestures of two communication partners during joint route planning. This task requires communication partners to explicitly align different interests or preferences regarding, for example, journey time or sights to be visited en route. We will initially address the following research questions in human-human studies experiments: Do communication partners align gaze and gestures and, if yes, how is this achieved? How can we model the empirical observations regarding eye gaze and gestures and the interactions between these two modalities in each communication partner?

After empirical data acquisition and modelling, the models will be integrated into an artificial communication system. This system, available to the project at the end of 2004, consists of an artificial head and torso with two arms and a simple hand capable of pointing. We will address the following research questions in human-machine experiments: Do gaze and gestures differ in human-machine communication when compared with human-human communication? Do humans and machines align gaze and gesture similarly and as well as humans do? How realistic is the model and does the artificial system adequately reproduce human gaze patterns and gestures? We expect that the integration of these essential non-verbal communication modalities in human communication and interaction should significantly enhance the proposed models and thus further help to better understand alignment both in verbal and non-verbal communication.

Methods

Methodologically, an extended version of a smart board will serve as a common interaction medium for the communication partners. This large interactive rear-projection screen allows for the presentation of stimuli and for their manipulation, either in touch-screen mode or, contact-free, using pointing gestures. Two high-precision eye trackers will simultaneously monitor

and record eye movements. The project already avails of one such eye-tracking system (EyeLinkI). The second system (EyeLinkII) will jointly be acquired with the project A6: "Alignment of attention in mediated communication".

The comparison between data from human-human and human-machine experiments based on statistical evaluation of eye movements will help to identify possible shortcomings and difficulties of the artificial system during communication and interaction. Body posture and mimics will be investigated in addition to eye gaze and gestures in the subsequent project phases. The artificial system can further be optimised by on-line adaptation during the course of communication using appropriate learning algorithms to allow for adaptation of the system to a specific human communication partner such that a more personalised emotional and pragmatic alignment in non-verbal communication can be achieved.

Work programme

Starting from empirical observation of human-human communication in front of the smart board, the implementation and modelling of the artifical agent's behaviour will progress in an iteration of testing and modifying the to optimise the behaviour of the artificial communication system. In particular the following steps have to be performed:

- setup of smart board and dual eyetracking system,
- empirical study of human-human communication by means of eyetracking,
- modelling of visual behaviour and implementation on artificial head,
- addition of gesture, body- and head posture,
- modelling and simulation on artifical head/torso system,
- comparison of human-human and human-machine communication,
- inclusion of online adaptation,
- evaluation and comparison experiments to verify adequacy of the model to establish natural communication.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	5,000	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**10,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Hardware for the realisation of the scenario: Smartboards and 2 cameras
- Student assistant with 19 hours per week

Project section A3: Syntactic-prosodic alignment in language processing and communication

Walther Kindt, Rüdiger Weingarten

Motivation

A relevant issue of the planned CRC is the question in which way dialogue partners align syntactic and prosodic structures in order to make communication more effective (i.e. robust, economic, fast, consuming less ressources). Alignment of syntactic and prosodic structures is a well-established topic in linguistics. In earlier approaches this type of alignment is conceived of as a rule based process, later on alignment was formulated in an OT-framework in terms of constraints (Truckenbrodt, 1999). In recent studies syntactic-prosodic alignment is considered from a processing perspective (e.g. Clifton et al. 2002). In this project section we want to extend these approaches of intra-personal alignment, following the concept of alignment proposed by Garrod and Pickering, to dialogue conditions (inter-personal alignment). From a theoretical point of view we expect that concepts of rules and constraints have to be contrasted with a pattern matching approach.

Objectives

In this project we want to investigate forms and functions of prosodic-syntactic alignment in language processing and communication. After developing a mathematical tool for prosodic similarity ("distance") and categorisation, we will review linguistic assumptions on alignment of prosodic and syntactic structures and test these hypotheses in dialogue experiments on prosodic-syntactic processing. In a further step we want to find out in a corpus study in which dimensions communication partners align their prosodic-syntactic patterns in the course of communication. Finally in a series of psycholinguistic experiments we will test the hypothesis that prosodic-syntactic alignment can facilitate language processing and communication.

Methods

The research in the proposed project section combines methods of mathematical system theory, corpus analysis and psycholinguistic experiments.

Work programme

The work programme is divided into five sections:

Section	Objective	Method
1.	*Development of mathematical tools for the description of prosodic and syntactic units ("gestalts")* Generally, linguistic models of prosody assume *discrete* units. In the empirical studies planned in this project we need a tool to describe prosodic units as *continuous* physical data. Furthermore, we have to calculate the *similarity* of continuous prosodic patterns. Henceforth, the construction of prosodic and syntactic units can be related to generalised principles of gestalt theory (cf. Kindt, 2001).	mathematical

2.	*Testing of linguistic hypotheses of prosodic-syntactic alignment in language processing* Whereas linguistic studies conceive alignment as a phenomenon of structures, the concept of alignment proposed by Pickering and Garrod aims at processes in communication. For example, the "constraint on edge alignment" (Truckenbrodt, 1999) can be tested experimentally by violating it under various dialogue conditions. On the basis of this type of experiment the relationship between structural and dialogue alignment can be clarified.	experimental
3.	*Analysis of prosodic-syntactic alignment in the course of communication* In a number of studies it was shown that communication partners align their utterances at the level of lexical items. Until now there have been no investigations on prosodic-syntactic alignments in the course of communication. As a consequence we want to analyse in a corpus study processes of prosodic-syntactic alignment between two dialogue partners. The mathematical tools for the description of prosodic and syntactic units shall serve as a basis.	corpus analysis
4.	*Investigation of facilitory effects of prosodic-syntactic alignment* A central assumption of alignment proposed by Pickering and Garrod aims at facilitory effects: participants align their linguistic representations to save processing resources and to speed up processing. These effects were not yet demonstrated experimentally for prosodic-syntactic alignment.	experimental
5.	*Development of a formal model of prosodic-syntactic alignment in communication* The results of the empirical studies shall be utilised to construct a formal model of prosodic-syntactic alignment in communication. Central aspects of this model will be: discrete vs. continuous units, incremental processing, status of rules and patterns, use of cognitive resources, robustness.	mathematical

References

Clifton, Ch., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech* 45 (2): 87-114.

Kindt, W. (2001), Syntax und Pragmatik: Eine zu entdeckende Verwandtschaft. In F. Hundsnurscher & F. Liedtke (Eds.), *Pragmatische Syntax* (pp. 5-29) Tübingen: Niemeyer.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological patterns. *Linguistic Inquiry* 30 (2), 219-255.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 SHK	1 SHK	1 SHK	1 SHK
Total	**70,800**	**70,800**	**70,800**	**70,800**
Small equipment	5,000	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**10,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Equipment: Audio equipment, software for phonetic analysis and mathematical simulation
- Student assistant with 19 hours per week

Project section A4: Syntactic alignment in normal and impaired communication

Martina Hielscher-Fastabend, Said Sahel, Hans-Jürgen Eikmeyer

Motivation

According to Pickering & Garrod (in press) communication involves automatic processes of alignment on different conceptual and linguistic levels. Partners may align their linguistic representations at many levels: at phonological, lexical, syntactic and semantic levels. The investigation of alignment at one of these linguistic levels, e.g. syntactic level, may provide insight into the issues of how alignment of linguistic representations can be accomplished and which mechanisms underlie these alignment processes. Partners may align their syntactic representation by adapting their syntactic structures, e.g. in terms of complexity and politeness, to each other and by coordinating their utterances in several ways. One basic mechanism which is assumed to cause or support the process of *syntactic alignment* is syntactic priming.

Experimental studies dealing with syntactic priming have found within-speaker local syntactic consistency. They showed that when people are given a message to produce that can be realised using more than one syntactic structure they are more likely to use the syntactic structure recently employed. Priming has been shown for structures at the phrase and sentence level, in spoken and written language. Most available studies on syntactic priming focused on normal adults, while children and adults with language disorders are nearly neglected.

The local syntactic consistency reported in the literature is based on priming experiments conducted with single speakers producing isolated phrases or sentences outside a discourse context. However, only few studies concerned with local syntactic consistency in dialogue are available (e.g. Cleland & Pickering, 2003). These studies showed that syntactic priming effects also occur between speakers in dialogue, indicating that speakers coordinate the syntactic structures of their contributions and therefore align with their interlocutors. The reported experiments demonstrated first tendencies for participants to use the same syntactic structure recently used by their interlocutors.

Objectives

A first aim of our project is to elucidate the conditions, in which syntactic alignment occurs in verbal interactions between adults and adult-child communication with normal language capacity. Structures of interest that will be focused on in the first stage of the project are a) structures on the phrasal level (e.g. a1: Genitiv vs. PP *("der Hund des Mannes"* vs. *"der Hund von dem Mann"*); a2: NS Apposition vs. PP ("*der Mann, der das Auto fährt" vs. "der Mann im Auto"*), b) structures on the sentence level (e.g. b1: active vs. passive; b2: word order variation) and maybe c) complex structures in forms of politeness, varying modus and directness of wishes.

The second, central aim of our project asks for syntactic priming and syntactic alignment in communication with language impaired persons. In subsequent stages, syntactic priming effects in dialogue should be investigated in aphasics, especially in Broca's aphasics with agrammatic language production, and in children with specific language impairment (SLI). A third group of interest may be communication with second language learners of German. The results obtained from experiments with normal adults will serve as control data.

Why should priming and alignment processes be different in these populations? A psycholinguistic model of language production and perception explaining syntactic priming effects has to

consider the availability of different syntactic structures for phrases and sentences and their differences in planning capacity and the amount of computational resources. Assuming that agrammatic language of Broca's aphasics reflects a capacity problem (e.g. Kolk & Heeschen, 1992; Schade & Hielscher, 1992), syntactic priming should be effective in Broca's aphasics, too. Otherwise, if it was a representational deficit, as it is hypothesised e.g. by Friedman (2001), it should be impossible to elicit specific syntactic structures. First empirical evidence for even enhanced priming effects of passive and dative structures in Dutch has been reported by Hartsuiker & Kolk (1998), thus supporting a capacity and accessibility deficit in agrammatic patients. At least, there is no empirical evidence yet for priming effects in normal and therapeutic communication as well. Two questions will be of main interest in this context: 1. Which syntactic structures can be primed in experimental conditions? 2. Can (these) syntactic structures be transferred into active communication by processes of alignment?

In a similar way it can be assumed, that second language learners of German profit from syntactic priming in communicative settings, as their passive knowledge of many syntactic structures is much higher than their active usage.

A third group of special interest in this context are children with specific language impairment (SLI). For these children it has been reported that mothers adapt their communicative style to the child's reduced grammatical style. On the other hand, language therapy for children frequently makes use of a principle, called elaboration of utterances, which is assumed to help the children to formulate more complete and complex structures. Here, empirical evidence is needed to clarify the conditions, in which children may profit from a more elaborated or from a restricted language code. Experimental data to confirm the effects of syntactic priming and alignment in these children are as yet completely missing.

Methods

As basic methodological tools we will use:

- Psycholinguistic experiments on syntactic priming,
- Development of a German screening for special grammatical competences in aphasia (perception and production), which can be adapted to children,
- Assessment of spontaneous communication,
- Psycholinguistic studies on syntactic alignment in dialogue situations.

Work programme

In the first stage of our project we make use of a classical syntactic priming paradigm with adults (normal controls, aphasics, second language learners), which has to be adapted in a second step to the capabilities of children (normal children age 3-6; children with SLI). The grammatical competence for the specific structures, that will be primed, has to be assessed for control persons and patient groups. If priming effects show up with sufficient consistency for these patients a neurophysiological basis can be gained in cooperation with project A4.

In the second stage a confidante has to be trained to use special grammatical structures in a communicative setting. Again, control subjects, aphasic patients, SLI children and second language learners will be compared in settings, that will allow sufficient syntactic complexity. In the dialogue situations with adults this can be a role play situation in which furniture has to be placed in a defined room (Rehabilitation centre, holiday flat, space ship, Barbie house, Playmobil hospital). With children we have to define game situations, where a common goal has to be reached by alternated instructions. Probably, several different situations have be used to elicit sufficient critical utterances. – At least three conditions have to be compared in each of these settings: monologue instruction, dialogue with confidante, natural dialogue.

References

Cleland, A.A. & M.J. Pickering (2003). 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.

Friedman, N. (2001). Agrammatism and the psychological reality of the syntactic tree. *Journal of Psycholinguistic Research, 30,* 71-90.

Hartsuiker, R.J. & H.H.J. Kolk (1998). Syntactic facilitation in agrammatic sentence production. *Brain and Language*, 62, 221-254.

Kolk, H.H.J. & Heeschen, C. (1992). Agrammatism, paragrammatism and the management of language. *Language and Cognitive Processes, 7,* 89-129.

Pickering, M.J & S. Garrod (in press). Toward a mechanistic psychology of dialogue.

Schade, U. & Hielscher, M. (1998). Die Modellierung des Agrammatismus. In M. Hielscher et al. (Eds.), *Störungen des Mediums Sprache.* Tübingen: Stauffenburg.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	3 SHK	3 SHK	3 SHK	3 SHK
Total	**122,400**	**122,400**	**122,400**	**122,400**
Small equipment	3,000	—	—	—
Consumables	1,000	1,000	1,000	500
Travel	5,000	5,000	5,000	3,500
Office supplies	2,000	800	800	800
Other	—	—	—	—
Investment	—	—	—	—
Total	**11,000**	**6,800**	**6,800**	**4,800**

Remarks:

- 1 IIa BAT: planning and conducting experiments with patient groups. The potential assistant has to have a double qualification for clinical diagnostic and empirical experience.
- 1 IIa/2 BAT: planning and conducting control experiments, statistical analyses.
- 3 SHK: assisting experiments, corpus transcriptions etc., 19 hours per week
- small equipment : equipment for outdoor recordings in clinical contexts:
- 2 high quality mini Net-MD recorders
- 2 high sensitive active microphones
- 1 camcorder

Project section A5: Alignment of perceptual and relational components of situation models

Sven Wachsmuth, Lorenz Sichelschmidt, Gert Rickheit

Motivation

This project section is concerned with the question of how shared situation models arise in communication. Language processing involves the formation of a representation of the matters talked about. A coherent and integrative representation has to be constructed of the elements of a described or perceived situation, which is composed of *entities* (objects, protagonists, etc.) within a *spatial-temporal framework* and their *properties* as well as *relations* between them (Zwaan & Radvansky, 1998). For a dialogue to be successful, it is essential that both interlocutors develop partially aligned models of the situation under discussion as proposed by Pickering and Garrod (in press); indeed, there is some evidence, reviewed by them, for alignment of different components of situation models in speech production or comprehension.

The main focus of the project designed is on the process of building a shared mental model of the situation under discussion through conversation. People normally do not achieve an alignment of situation models through explicit negotiation (e.g., Garrod & Anderson, 1987). Instead, Pickering and Garrod argue in their paper that aligning situation models comes about partly through an interactive process of aligning linguistic representations at different levels. But how does alignment at "low levels" (word choice, syntax) lead to alignment at the level of the situation model? Several aspects of this question will be addressed in the project by combining psycholinguistic experiments with theoretical and conceptual work in cognitive science, and computational modelling.

Objectives

Of particular concern is the interrelation between verbal and visual information in real-world (non-imaginary) situations, and especially the alleviation of the information-processing burden brought about by essentially resource-free and automatic alignment processes. Current computational approaches to an integrated processing of verbal and visual information are based on a structural mapping of both types of information (e.g., Siskind, 1996). For reference resolution, this process can be very resource-expensive in the case of a great number of possible referents and structural relations. Contrary to this, reference resolution usually does not seem to impose a high processing load in successful natural dialogue. The main idea is that linguistic processes in dialogue can lead to a pre-structuring of the representational state, which facilitates the processing of the current utterance. On this account, an efficient mapping is achieved by way of a large amount of pre-processing and pre-selection. For example, instead of initiating a visual analysis process from a referential expression or considering a complete scene reconstruction on any level of detail in referential resolution, a situation model is adapted that reflects expectations about relevant scene aspects. Such processes that enable a massively resource-reduced reference resolution are one of the main focuses of this project.

These processes are intertwined with both the organisation of different pieces of information into a coherent representation, providing robustness of information processing, and the focusing of attention, which allows part of the scene or possible referents to be ignored. On the other hand, the tendency for people to consider only what is represented in their models of a situation can also affect their decision-making and the drawing of inferences. Therefore, one aspect investigated will be the selectivity of processing and the focusing of attention.

Of equal importance is the question of how pan-situational world knowledge and the situation model interact. One aspect concerns the effects of different category levels activated on the situation model built. One particular object can be categorised at different levels of abstraction as indicated by the denotation chosen. As language depends heavily on categorisation, the question of how this affects the granularity and selectivity of dynamically constructed models. arises. Another aspect concerns typicality effects of perceptual properties. In a dialogue scenario where a situation model has to be constructed on the basis of a verbal description of an event, the question is whether deviations from prototypical values are taken into account, and how a temporal misalignment is resolved by the interlocutors.

Methods

The question how shared situation models arise in communication will be examined in an interleaved way combining psycholinguistic experiments and the development of computational moduls for system demonstrators. For experiments, psycholinguistic techniques like different priming paradigms (semantic, perceptual, spatial priming) and eye-tracking methods will be applied. Probabilistic graphical models are a promising candidate for the computational modelling of various influences on the formation of situation models. Based on the sound mathematical basis of probability theory, perceptual as well as relational components can be represented including their time-dependent behavior.

Work programme

To explore the mechanisms involved in the interactive construction of situational models, different communication scenarios will be compared. These include monologue vs. dialogue, dialogue with simultaneous perception of an event vs. retrieval from episodic memory, and talking about a scene imagined or previously perceived in dependence on concurrent perception. Such a stepwise variation of factors possibly influencing the formation and updating of situation models allows the interplay of the contributing processes to be investigated and, on this basis, the development of a model of the construction of aligned situation models with emphasis on perceptual and relational components. Relevant cognitive processes to be taken into account include semantic and syntactic as well as visual and spatial priming, and correction processes. One problem that has to be dealt with is the proper representational format for situation models, in which perceptual and linguistic sources of information are merged, and episodic memory representations are accessed and possibly changed. Besides the empirical studies that drive the development of the computational framework, the project aims at two different application scenarios that provide a testbed for algorithmic solutions, i.e. image/video-retrieval and human-robot-communication.

References

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181-218.

Pickering, M. J. and Garrod, S. (in press). Toward a mechanistic psychology of dialogue.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings, *Cognition*, 61, 39-91.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	2,500	—	—	—
Consumables	1,200	1,200	1,200	1,200
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	1,500	1,500	1,500	1,500
Investment	—	—	—	—
Total	**8,700**	**6,200**	**6,200**	**6,200**

Remarks:

- Staff includes one PhD student in Computer Science and one PostDoc in Psycholinguistics
- Student assistant with 19 hours per week
- Equipment: PC for psycholinguistic experiments
- Other: costs for subjects in experiments

Project section A6: Alignment of attention in mediated communication

Helge Ritter, Lorenz Sichelschmidt

Motivation

This project aims at the empirically founded distinction of diverse modes of alignment and the assessment of the facilitatory role of alignment by systematically contrasting mediated communication (in which there are technical constraints to alignment) to face-to-face communication (in which there are no such constraints). For one, this is a scientifically interesting issue in itself because it is an extension of the interactive alignment account to multimedia communication. For another, it presents a challenge for speech processing systems and, in addition, it is an issue of some practical significance.

Objectives

- Investigation of functional aspects of attentional alignment in multimodal environments.
- Examination of mediated communication (e.g., video conferencing) in comparison to direct (face-to-face) dialogue.
- Exploration of possibilities to empirically distinguish between diverse modes of alignment and to assess their facilitative potential.
- Practical utilisation of the empirical results to the effect of optimising the flow of information in mediated communication.
- Modelling of human eye movements with respect to cognitive adequacy and implementation in artificial systems.

Methods

The proposed empirical studies are based on the registration of eye movements.
Eye movements can be viewed as indicating the importance of the various modalities in communication. Eye movements provide valid spatio-temporal clues to the interlocutors' current focus of attention as well as to the time course of processing (anticipations, synchronicity, "lagging", etc.). Ultimately they yield insight into the role of alignment in information processing. Eye movements may even show overt misalignment and attempts to realign. In principle, the methodological background of the proposed research is provided by the experimental–simulative approach.

Experimentation: Simultaneous tracking of interlocutors' eye movements – including gaze position transfer and gaze-contingent display – is suggested as the essential research method (one eyetracker system, EyeLinkI, is available; a second system, EyeLinkII, is yet to be obtained). The eye-movement data will be matched to linguistic events in the ongoing dialogue, and will be supplemented by global measures such as reaction times or error probabilities.

Simulation: Models of eye movement behaviour (in terms of parameters like fixation probability or gaze duration) are to be developed that encompass a range of relevant situational and verbal predictors. These models are to be evaluated with respect to cognitive adequacy and to be implemented into artificial interlocutors, which allows us to test and optimise the models in "real-world" settings.

Work programme

Against the background sketched by Pickering and Garrod (2004), alignment can be viewed as serving various functions, most notably the efficient control of implicit common ground, focus of (joint) attention, and progress of dialogue. The benefits of alignment should be most pro-

nounced in direct or face-to-face communication. However, mediated communication – as in telephone conversations, video conferences or tele-guidance/tele-robotics – constrains typical communication modalities, and may therefore complicate the development of implicit common ground. Telephone conversations, for example, are strictly verbal with communication partners forming individual scene representations which are not necessarily identical.

The rationale behind the proposed project section is to contrast technically constrained to face-to-face communication situations. Mediated communication scenarios form a convenient environment for the investigation of functional aspects of attentional alignment in multimodal communication for various reasons. For one, such scenarios enable researchers to study selected communication modalities in a non-artificial, ecologically plausible way. For another, eye movement measurement technology permits a fine-grained investigation of the transition between direct and mediated communication (for instance by stepwise reduction of the information available on the visual channel from body posture over facial expression to gaze position). Thirdly, the research programme could lead to empirically substantiated recommendations as to the relative contribution of the diverse modalities to alignment. These recommendations could, in effect, lead to the optimisation of mediated communication, both when implementing "remote" human-machine interaction systems and with respect to the provision of an optimal set of communication modalities.

In extension of current research by Kraut et al. (2003) and Anderson et al. (2004), alignment processes in different mediated communication scenarios shall be investigated: communication with restricted interlocutor visibility and video conferencing. In these studies, face-to-face communication will serve as a reference. Tasks like joint maintenance of a mechanical object, which are of a referential nature, appear to be particularly suited for investigation within the scope of these scenarios. In the further course of research, investigations will focus on details of the requirements in mediated communications such as the question which information is indispensable, facilitatory or detrimental (and to what degree), or redundant for successful communication.

Empirical observations of human eye movements shall be implemented in computer models and simulations. These will focus on the gaze-contingent and communication-driven information retrieval and information processing in the various communication modalities and their integration.

References

Anderson, A.H., Mullin, J., McEwan, R., Bal, J., Carletta, J., Grattan, E. & Brundell, P. (2004). Exploring why virtual team working is effective in the lab but difficult in the workplace. In H.Bromme, F. Hesse & H. Spada (Eds), *Barriers and biases in computer-mediated knowledge communication – and how they may be overcome.* Dordrecht: Kluwer (in press).

Kraut, R.E., Fussell, S.R. & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction, 18*, 13-49.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* (in press).

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**110,400**	**110,400**	**110,400**	**110,400**
Small equipment	2,000	—	—	—
Consumables	1,000	1,000	1,000	1,000
Travel	3,000	3,000	3,000	3,000
Office supplies	—	—	—	—
Other	1,500	1,500	1,500	1,500
Investment	30,000	—	—	—
Total	**37,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Investments: "EyeLinkII" eye tracking system. "EyeLinkII" is downward compatible to the extant "EyeLinkI" system but presents a substantial advance with respect to speed, accuracy, and versatility. In combination with the extant system, "EyeLinkII" will be used for simultaneous measurement of interlocutors' eye movements (as planned in project sections A2, A6, and B2). Costs given after recent price quote from the manufacturer, SR Research, Toronto, Canada.
- Equipment: High-speed DLP back projection system for large scale stimulus presentation
- Student assistant with 19 hours per week

Project area B: Intrapersonal aspects of alignment

The charm of the notion of alignment in communication lies in the fact that it addresses alignment between interlocutors, by addressing alignment within interlocutors – among the message to be conveyed, the relevant situation model and the representations on the phonetic, phonological, syntactic, lexical and semantic level. While inter-personal alignment is regarded by Pickering and Garrod (2004) as the application of routines holding throughout a dialogue, intra-personal alignment focusses on the transient short-term activation or 'priming' of linguistic representations used in both production and comprehension. In order to arrive at an alignment between interlocutors, it is thus necessary that information from diverse channels or of diverse formats and in both processing modes becomes aligned within an individual.

One of the proposed extensions of the notion of intrapersonal alignment is taking into account the relationships between verbal expressions and co-verbal deictic or iconic gesturing as a special case of a research direction studying production and/or comprehension processes with several input channels involved. A suitable field of investigation is provided by co-occurrences of speech and concomitant spontaneous gesture studied in project section B1 under the perspective of self-alignment. It is based on empirical results showing that speakers coordinate gesture stroke with phonetic stress and syntactic, semantic or pragmatic information. Project section B7 studies multi-channel processing in a more general way by measuring neurophysiological correlates of processes in which the information presented on one of the channels is more or less disturbed.

In order to work on a foundation of the intrapersonal aspects of the alignment paradigm central issues are discussed by several project sections of this project area. These are: compositionality and indirect interpretation in project section B2, and implicit common ground in project section B3. On the other hand, alignment will be related to a physiological substrate in project section B4 and B7, as well as and to classical Gestalt ideas in project section B5.

Project section B6, finally, focusses a single processing level, the acoustic phonetic one, and studies the alignment of representations for both production and comprehension.

In more detail the project sections are concerned with processes of intra-personal alignment in the following sense:

Project section B1 (Speech-gesture alignment; Kopp, Rieser & Wachsmuth) works on the construction of an integrative multi-modal theory of speech and gesture and the corresponding simulation in virtual reality (VR). It studies the temporal synchronisation, meaning and meaning composition of deictic/iconic gestures and accompanying verbal expressions in the context of the structure of dialogue (games) and the interaction of gesture, topic structure, focus and viewpoint.

Project section B2 (Alignment-based accounts of compositionality and indirect interpretation; Eikmeyer, Rieser & Sichelschmidt) starts from the fact that the interactive alignment account neither covers phenomena beneath the word level nor does it elaborate on problems of indirect interpretation. It aims at a major extension of the interactive alignment account in both these areas, including the treatment of metonymy and, in the further course, alternatives to Gricean explanations e.g. indirect speech acts or enthymemes, making use of the notion of routines.

Project section B3 (Construction of implicit common ground; Eikmeyer, Jäger, Rickheit & Rieser)
Implicit common ground is one of the central mechanisms of the interactive alignment. The establishment of implicit common ground as an information state, its semantic and cognitive properties and operations (e.g. default inferences) defined over it will be studied. Game-theoretical concepts and logics will be used for models of implicit common ground and preconditions for the realistic simulation of information states by VR will be elaborated.

Project section B4 (Neural synchronisation during language processing; Müller & Ritter) relates alignment to neurobiological notion of synchronisation. It will identify neural correlates of alignment during verbal communication by means of electroencephalography (EEG), i.e. eventrelated potentials and spectralanalytic methods such as coherence analysis and phase analysis. The starting hypothesis is, that alignment at the linguistic levels is accompanied by distinct correlational patterns observable in EEG traces of the recipient. To verify (or refute) this, a perturbation technique will be used in conjunction with the aforementioned.

Project section B5 (Gestalt approaches to alignment – from visual to verbal processing; Mehler, Rickheit & Sichelschmidt) starts from the observation that the interactive alignment approach to dialogue lacks specific ideas about the cognitive mechanisms that bring about alignment. Since some of the ideas addressed are reminiscent of the notion of Gestalt, it aims at exploring in greater depth the viability of utilising classical Gestaltist approaches in the study of alignment. Hereby special emphasis will be put on routinisation as a central notion.

Project section B6 (Alignment of acoustic-phonetic representations in speech perception and production; Fink & Eikmeyer) combines the study of intrapersonal alignment processes involved in both speech perception and production on the level of acoustic-phonetic representation with the realisation of a technical system for integrated speech recognition and synthesis. The latter will be used for simulating impaired systems and systems for language acquisition on the acoustic-phonetic level.

Project section B7 (Multimodal language processing: Dynamics and integration of parallel processes in verbal communication; Müller & Wörmann) takes the number and diversity of information sources in verbal communication as its challenge. A listener has to manage both the result of distinct parallel sub-processes, probably performed by distinct neuronal networks, and the multimodal integration of neuronal information streams. Stimuli demonstrating optimised, non-optimised and disturbed verbal behavior will be tested with brain imaging techniques.

Project section B1: Speech-gesture alignment

Stefan Kopp, Ipke Wachsmuth, Hannes Rieser

Motivation

When communicating naturally, humans frequently combine language and spontaneous gestures to form multimodal utterances. In such utterances, speech and gesture appear highly coordinated and closely intertwined—in other words, aligned to each other by the human speaker. These alignments concern the coordinated choices of the form of linguistic and gestural behavior, their relative timing, as well as their common organisation into a single phrasal utterance structure. Their effects are essential for how meaning is communicated by both modalities concertedly. The resulting confluence of language and gesture has led many researchers (e.g., McNeill, 1992) to believe that speech and gesture are products of the same generative process, starting from one ideational complex and comprising significant interactions between speech and gesture. Yet, it is still an open question as to *how* language and gesture align in producing a coherent multimodal utterance, and potential interactions between both modalities have been suggested at almost every stage of the production process, from memory, via conceptualising, down to formulating single behaviours, or even incessantly during the whole process. Researching this multimodal topic will substantially extend the Pickering and Garrod notion of self-alignment.

Objectives

Our goal is to systematically investigate the ways in which speech and gesture align within multimodal utterances in dialogue, and we aim to achieve an understanding of the underlying, intra-personal mechanisms allowing us to model the generation of coordinated language and gesture for embodied conversational agents (ECA). We will focus on deictic gestures, which directly point to a location or region in space, as well as on iconic gestures that impart visual information to the utterance depicting what is being referred to (including the fusion of both functions within single gestures). The main research topics to be taken up in this context are:

1. Temporal synchronisation of deictic/iconic gestures and verbal expressions,
2. Meaning of deictic/iconic gestures, and its composition along with the composition of the meaning of accompanying verbal expressions,
3. Form of deictic/iconic gestures and the accompanying verbal expressions, and their derivation from the meaning(s) to be conveyed,
4. Deictic/iconic gesture phrases and the structure of dialogue (games),
5. The packaging of information into deictic/iconic gesture phrases in dialogue, i.e. the interaction of gesture, topic structure, focus and viewpoint.

Concerning (1) to (5), we will study alignment phenomena arising at various levels: The synchronisation of the co-expressive phases in gesture and speech serves as a precondition for much else, and it constrains both the timing of gesture within single phrases as well as the temporal course of speech and gesture across subsequent phrases (Kopp & Wachsmuth, 2004). Deictic gestures must be in the scope of "their" referring expressions and the same holds good for iconic gestures. The distribution of meaning across speech and gesture leads to gestures that—depending on both the speaker's communicative goals as well as the current discourse context—are found either to be redundant with speech, or to contribute information complementary to what is expressed in the verbal modality (McNeill, 1992). Pointing gestures establish reference to external objects by complementing the referring verbal expressions or by independently identifying the object referred to. In contrast, iconic gestures are intimately

bound up with predication. They either "echo" or complement verbal content and are fully interpretable only in the context of simultaneous speech. The form of complementation, then, can even be quite subtle and first studies suggest that elementary meanings can be lexicalised such that concomitant gestures can be used to add, e.g., directional or aspectual information to verbal content (Kopp et al., 2004). This shows that a tight compositional relation between gestural and verbal meaning must exist, and that, in most cases, gestures interact with verbal content in a systematic way. That is, there is alignment between gesture meaning and the syntactic and semantic structure of the verbal expression, involving situational alignment (see Rieser, 2004). If a gesture is used for emphasis, expressing a viewpoint, or as a focus marking device, it serves a pragmatic function and must interact with Gricean conversational postulates. Furthermore, as statistical investigation of annotated corpus data has shown (see Lücking, Rieser & Stegmann, 2004), gesture supports intra-personal turn-taking regularities and the delimitation of dialogue moves. Hence, it is firmly embedded in and aligns with dialogue structure.

Methods

Our research will encompass the empirical study and analysis of human behavior (linguistic project part), as well as the conception of computational models of the processes involved and their realisation in virtual humans (informatics part). Both the linguistic and the informatics parts, will be based on the same grammar formalisms (LTAG, AVMs). The comparison of the producible synthetic behavior with the originally observed real-human data, then, allows us to evaluate our models and to inform their iterative revisions. This will include new, refined questions to be answered by further empirical analysis, ultimately yielding a better understanding of the intra-personal processes under investigation.

Work programme

Experimental studies on deictic and iconic gesture will be carried out using VR technology such as data gloves and motion tracking. In addition, eye-tracking facilities will be used to specify phases of attention as well as to analyse human reaction to artificial speech-gesture productions. A major target with respect to deictic gestures is to delimit the pointers' "pointing cones", i.e. the domains singled out by pointing gestures, that will serve as a basis for constructing rigid models of pointing. Similarly, experiments will be conducted involving the use of iconic gestures as nonverbal predication. All empirical studies are expected to elicit sets of dialogue games, which will then be annotated according to an extended HCRC coding schema and investigated using various statistical methodologies for pattern extraction.

The patterns found during data analysis will inform the technical modelling of the underlying generative processes. Speech and gesture are supposed to run through a multi-stage production process, and we will break down the possible aligning interactions on different levels along the pipeline commonly conceived for the generation of natural language:

- Content planning (selection of information and its arrangement in a rhetorical structure),
- Microplanning (encoding the content plan into linguistic terms and gesture morphology),
- Surface realisation (producing the final speech and gestures).

For the various linking points between speech and gesture along this line, i.e. information packaging, dialogue moves, semantics of constructions, lexicon and syntax, and dialogue games, interfaces will be constructed. The techniques used for the content planning stage and the microplanning stage will encompass constraint-based grammar, LTAG informed systems, and dynamic semantics based on DRT-derivatives (including SDRT). Surface realisation is responsible for synchronous speech and gesture and will be achieved using the embodied conver-

sational agent MAX. In extending MAX, conceived models will then be put into action and tested in real-time VR simulations.

References

Kopp, S., Wachsmuth, I. (2004). Synthesizing Multimodal Utterances for Conversational Agents, *Journal Computer Animation and Virtual Worlds, 15*(1), 39-52.

Kopp, S., Tepper, P., & Cassell, J. (2004). Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. *Proceedings of ICMI'04*, to be published.

Lücking, A. Rieser, H. Stegmann, J. (2004). Statistical Support for the Study of Structures in Multi-Modal Dialogue: Inter-Rater Agreement and Synchronization. *Proceedings of Catalog'04*, pp. 56-64.

McNeill, D. (1992). *Hand and Mind*. Chicago and London: The University of Chicago Press.

Rieser, H. (2004). Pointing in Dialogue. *Proceedings of Catalog'04*, pp. 93-101.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Student assistant with 19 hours per week

Project section B2: Alignment-based accounts of compositionality and indirect interpretation

Hans-Jürgen Eikmeyer, Hannes Rieser, Lorenz Sichelschmidt

Motivation

The interactive alignment account (Pickering & Garrod, 2004) does not cover phenomena beneath the word level; it also does not elaborate on problems of indirect interpretation. The proposed research project section aims at a major extension of the interactive alignment account in both these areas. To begin with, we will investigate interlocutors' alignment on the interpretation of nominal compounds. This will include the treatment of metonymy, and hence address matters of indirect interpretation. In the further course, we attempt to develop alignment-based alternatives to Gricean explanations e.g. indirect speech acts or enthymemes, making use of the notion of routines as put forward by Pickering and Garrod (2004).

Objectives

- In-depth investigation of the non-compositional constitution of meaning and of the interface of compositional and non-compositional meaning.
- Implementation of interfaces between Gricean and non-Gricean explanations.
- Development of non-Gricean explanations for metonymy, indirect speech acts, and enthymemes based on routinisation.
- Development of a grammar for the description of various types of expressions such as compounds (in particular, nominal ones).
- Underspecified descriptions for the diverse relations that hold in nominal compounds of the same pattern.
- Adaptation to German of current theories of conceptual combination and empirical assessment of their respective explanatory power.
- Development of routines which can serve as shortcuts, and assessment of their cognitive adequacy.

Methods

Research in the proposed project section will be guided by the extended empirical-simulative approach.

The empirical investigation will proceed from the generation of an expedient dialogue corpus (within the CRC framework). Dialogues in this corpus will be annotated and analysed as to lexical, syntactic, and semantic characteristics as well as to dialogue moves, games, and transactions. For annotation, an extended HCRC schema will be used that embraces sub-word level compound structures and paradigms of indirect interpretation such as metonymy, indirect speech acts, and enthymemes. In addition, various statistical tools will be employed to extract patterns not easily amenable to intuitive inspection.

In parallel with the corpus-based approach, psycholinguistic experiments will be conducted in order to test specific hypotheses on regularities in the production and comprehension of non-compositional expressions. The experiments will make use of processing time measurement in combination with suitable priming techniques (for instance, to study the degree of activation of specific relationships that hold between the elements of a nominal compound). If so required, these data will be supplemented by spatiotemporal data from eye movement studies.

The findings will partially be incorporated in computer simulations of the probability of arriving at a specific interpretation and of the cognitive effort that goes along with it. Such simulations could be viewed as variants of the current 'landscape model' of comprehension (a connectionist model which uses activation landscapes to capture the relations between online processes and offline representation of comprehension), however, tailored to the requirements of indirect interpretation. The simulation models will, in turn, be evaluated against new empirical data.

Work programme

Compositionality principles, originally traced back to Frege, were one of the main driving forces in the semantic theories proposed in the 70s and 80s by Montague, Lewis, Cresswell, and others. About the same time Grice made it clear that a whole field of interesting semantic and pragmatic relations need explanatory devices of a non-compositional sort. He based these on rationality principles and general maxims. In essence, Gricean explanations can attribute non-compositionally derived content to expressions such as so-called indirect speech acts.

In the proposed project section, we will study the non-compositional constitution of meaning and the interface of compositional and non-compositional meaning. Focusing initially on nominal compounds, we will start out with the constitution of meaning below the word level and then proceed to metonymies tied to words and constructions, to indirect speech acts, and finally to enthymemes within dialogue games. Interpretation of expressions will in general be viewed as a joint project in the sense of Clark (1996). Non-compositionality on the morphological level is different from non-compositionality in the other fields mentioned. Seemingly identical formation patterns for nominal compounds show that they have to be interpreted along different lines; e.g., *tablecloth* vs. *playwright* (Costello & Keane, 2000; Gagné, 2003). This makes underspecification theories a viable starting point for description. Actually, compounds are double-faced, since metonymies frequently enter into compound formation, cf. *propeller bolt* (Propellerschraube) in the CRC corpus.

With the domain described so far, alignment problems are tied up in various ways. With respect to compounds, addressees have to choose the appropriate relation in an underspecified construction. For the other phenomena mentioned, elaborate Gricean derivations require too much time and cognitive effort. However, the cognitive mechanisms that enable recipients to arrive – seemingly without much effort – at an interpretation of conceptual combinations (which, in German, typically surface as compound words) are not yet well understood. Alternatives to fully fledged Gricean derivations will have to resort to defaults, routines, or other shortcuts. Of course, there is always the problem of establishing a routine for the first time – say, with respect to metonymies like *propeller bolt* (Propellerschraube). In addition, one has to consider interfaces among lexicalisation, routines, Gricean explanations and standard compositionality which might function as models for expressions of a hybrid nature.

References

Clark, H.H. (1996). *Using language.* Cambridge, UK: Cambridge University Press.

Costello, F.J. & Keane, M.T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science, 24*, 299-349.

Gagné, C.L. (2000). Relation-based combinations versus property-based combinations: A test of the CARIN theory and dual-process theory of conceptual combination. *Journal of Memory and Language, 42*, 365-389.

Meyer-Fujara J. & Rieser H. (2001). Depiktionsmetonymien und ihre Auflösung im Diskurs. In L. Sichelschmidt & H. Strohner (Eds.), *Sprache, Sinn und Situation* (pp. 91-109). Wiesbaden: Deutscher Universitäts-Verlag.

Meyer-Fujara J.H. & Rieser H. (2003). Semantics-Pragmatics interface for metonymy resolution. In F. Schmalhofer, R. Young & G. Katz (Eds.), *Proceedings of EuroCogSci 03, the European Cognitive Science Conference 2003* (p. 418). Mahwah: Erlbaum.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* (in press).

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	1 SHK	1 SHK	1 SHK	1 SHK
Total	**98,400**	**98,400**	**98,400**	**98,400**
Small equipment	2,000	—	—	—
Consumables	1,000	1,000	1,000	1,000
Travel	3,000	3,000	3,000	3,000
Office supplies	—	—	—	—
Other	1,000	1,000	1,000	1,000
Investment	—	—	—	—
Total	**7,000**	**5,000**	**5,000**	**5,000**

Remarks:

- Equipment: Portable experimentation system ("Dothan" hardware with "NBS Presentation" programming environment)
- Student assistant with 19 hours per week

Project section B3: Construction of implicit common ground

Hans-Jürgen Eikmeyer, Gerhard Jäger, Gert Rickheit, Hannes Rieser

Motivation

Implicit common ground is one of the central mechanisms in Pickering and Garrod's "Mechanistic psychology of dialogue" in order to secure resource-sensitive, swift and fine-grained alignment among agents in dialogue. Concerning this concept we will investigate the following research issues: Which of the agents' (sub-turn) exchanges are used to establish implicit common ground? Taking implicit common ground as an information state, which are its semantic and cognitive properties and which operations (e.g. default inferences) can be defined over it? Investigation of implicit common ground will make use of a hybrid system consisting of game-theoretical concepts and logics. Setting up implicit common ground via agents' sub-turn exchanges will serve as a precondition for the realistic simulation of information states by VR techniques. Modelling exchange on a sub-turn level will in the long run improve speed and robustness of man-machine-interaction.

Objectives

- Empirical investigation of agents' inferring by means of the implicit common ground,
- Cognitive status of agents' inference mechanisms,
- Theoretical models of inferring by means of the implicit common ground,
- Taxonomy of coordination devices for constructing common ground,
- Interface of implicit common ground and full common ground.

Methods

Theory:

Theoretical modelling will be based on fine-grained hypothesis-driven corpus annotation. In order to detect and evaluate structural patterns, for example concerning default grounding, statistical methods will be used. Models of implicit common ground will be set up using modern methods of semantics and pragmatics such as default theory, underspecification approaches and current theories of dialogue such as PTT and SDRT (Heydrich, Kühnlein & Rieser, 1998). Roughly, the aim is to devise a hybrid model for common ground and the package of methods tied to it, using the theory of games for the "automatic" processes and some BDI-architecture for the deliberative and negotiative ones (Jäger, 2004).

Psycholinguistic Experiments:

Implicit common ground can be considered as a possible explanation why interlocutors in an interactive communicative situation are able to build up easily corresponding information states. Based on the initially set up models of implicit common ground a series of experiments will be conducted to examine how theoretically relevant factors influence the establishment of aligned representations. Therefore, experiments on language production and understanding will be conducted in different settings (*e.g.* construction tasks) and communicative constellations (*e.g.* with only imagined vs. real interacting partners) using eye tracking (Hanna, Tanenhaus & Trueswell, 2003) and dual task paradigms. Having identified triggering mechanisms for implicit common ground, in a further step of experimental work also the interface of implicit and full common ground will be considered in more detail. Full common ground is needed in cases of misalignment when the proposed interactive repair mechanisms fail. Based on the empirical and simulative results dealing with implicit common ground factors will be varied to find out when explicit and costly processes of negotiation in dialogue will be necessary to guarantee communicative success.

VR-Simulation:
Robust and swiftly operating dialogue competence is crucial for a lifelike multi-modal interaction with embodied conversational agents (ECA) in real-time Virtual Reality (VR). In contrast to traditional uni-modal, e.g. text based, dialogue systems, the ECA in VR is involved in a face-to-face interaction using both natural language and gestures. Resource-sensitive models of implicit common ground can be the key to extending existing approaches to dialogue handling to a sub-turn level. As a consequence, interactivity and robustness of ECAs under real-time conditions will be increased. Therefore, the main research interest of the computer science part is to simulate and evaluate the theoretical models in cooperation with the ECA MAX. In the beginning we will concentrate on the domain of construction tasks, where we can employ existing work such as an incrementally operating reference resolution engine.

Work programme

Fully fledged theories of common ground are usually based on BDI architectures. This makes them "expensive" systems. In addition, they are non-decidable, hence they cannot be used as models of real mental processes in a straightforward way. Pickering and Garrod have developed an alternative point of view with respect to speedy coordination processes: Agents establish a so-called implicit common ground; full common ground and the BDI-architecture are only used on demand. No separate full representation of the agent's private beliefs, the mental states of the other agent and the mutually believed common ground is assumed. Implicit common ground is established and maintained by help of an interactive repair mechanism mapping a model of the input onto one's own current representation of the situation. If misalignment has occurred, an utterance is re-formulated or repaired, in order to restore the implicit common ground. Hence, we may regard the implicit common ground as the interactively monitored representation of the situation model valid by default for every discourse participant.

In this project we investigate which mechanisms participate in the construction of common ground. We will try to answer the following questions: On which information is the implicit common ground based – direct perception, domain knowledge, linguistic knowledge, grounded dialogue moves? Under which type of (presumably default) inferences is the implicit common ground closed? Some closure assumption will be necessary in order to consider implicit common ground as reliable information at all. At the same time, we will investigate which mechanisms language provides to establish implicit common ground. Close scrutiny of the SFB 360 "Baufix"-corpus and comparison with comparable corpora such as TRAINS (University of Rochester) and Map-Task (HCRC, Edinburgh) has shown that *inter alia completions*, *sub-turn accepts, overlaps* and *re-phrasings* are candidates for such mechanisms (Poncin & Rieser, 2000).

Further central aims of the project will be to examine the nature of the alignment processes for grounding and the cognitive status of the resulting implicit common ground. In a later phase psycholinguistic mechanisms of switching from automatic inferences tied up with implicit common ground to controlled inferences associated with full common ground (under knowledge, belief, supposition etc.) will be investigated.

Co-operations:
We will co-operate with project A1 on problems concerning the interface among implicit common ground and full common ground. The same holds true with respect to the simulation by ECA MAX.

References

Hanna, J. E., Tanenhaus, M. K. & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43-61.

Heydrich, W., Kühnlein, P. & Rieser, H. (1998). A DRT-style modelling of agents' mental states in construction dialogue. In J. Hulstijn & A. Nijholt, A. (eds.). *Formal Semantics and Pragmatics of Dialogue – Proceedings of the Twente Workshop on Language Technology 13 (Twendial '98)*. Enschede, Universiteit Twente, Faculteit Informatica, 1998.

Jäger, G. (2004). Common grounds, questions and relevance. In *Proceedings of the Workshop Questions under discussion*, ILLC, Amsterdam, 1997.

Poncin, K. & Rieser, H. (2000). Multi-speaker utterances and co-ordination in task-oriented dialogue. In M. Poesio & D. Traum, D. (eds.), *Proceedings of Götalog 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue*. Göteborg University. 15-17 June 2000. Gothenburg Papers in Computational Linguistics 00-5.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	2,500	—	—	—
Consumables	1,500	1,500	1,500	1,500
Travel	4,000	4,000	4,000	4,000
Office supplies	—	—	—	—
Other	1,000	1,000	1,000	1,000
Investment	—	—	—	—
Total	**9,000**	**6,500**	**6,500**	**6,500**

Remarks:

- Equipment: Workstation for experiments (3D-capable, large screen), server for simulations (dual processor)
- Other: Costs for subjects in experiments
- Student assistant with 19 hours per week

Project section B4: Neural synchronisation during language processing

Horst M. Müller, Helge Ritter

Motivation

The aim of this project is to identify neural correlates of alignment during verbal communication by means of electroencephalography (EEG). The time course of meaning constitution during scenario-embedded sentence comprehension can be investigated by analysing the event-related potentials (ERP analysis). The electrophysiological dissociation of parallel processes within neuronal cell assemblies can be achieved by using spectralanalytic methods such as EEG coherence analysis and phase analysis. Our hypothesis is: Alignment at the different levels postulated e.g. by Pickering and Garrod (phonetic/syntactic/semantic/pragmatic) is accompanied by distinct correlational patterns observable in EEG traces of the recipient. To verify (or refute) this hypothesis, we propose a perturbation technique in conjunction with a combination of spectralanalytic methods and machine learning techniques for spectral analysis of EEG data.

Objectives

The investigation of alignment processes by means of non-invasive electrophysiological techniques. Further, the development of combined machine-learning and coherence analysis techniques for EEG data analysis. Third, the study of processes of neuronal synchronisation which allows monitoring of intrapersonal alignment processes during verbal communication and thus allows the determination of interpersonal alignment between two communicating partners.

Methods

- Neuro-/psycholinguistic experiments with auditory and visual stimuli.
- Neuronal activity will be measured by electroencephalography (32 to 64 channels), using two different techniques: Analysis of the event-related potential (ERP-analysis) and spectralanalytic techniques (power and coherence analysis, phase synchronisation) (Weiss & Müller, 2003).
- In the past, we were successful in classifying single trial EEG data for Brain-Computer Interfacing (Kaper et al., 2004) as well as for identifying different types of nouns from auditory stimuli (Meinicke et al., 2003). In the future, we want to further improve these techniques to be able to differentiate prominent stages during meaning constitution, e.g. the way of "getting the point" listening to complex, metaphoric, or idiomatic sentences from just a single trial. For that purpose we will utilise coherence and phase correlation information for feature extraction for a state-of-the-art machine learning classifier (SVM, MCC, FDA).
- Cluster analysis (LVQ, kNN, SOM) shall be used to identify prototypical subgroups of coherence patterns.

Work programme

In order to obtain empirical data on the way utterances enable comprehension in a scenario-based verbal communication, subjects shall listen to several sentences in context. We expect to:

- Identify/characterise EEG correlates of alignment at different levels (phonetic/syntactic/semantic/pragmatic),
- In this way, open up an objective, brain signal-based approach to the study of alignment in various communicative situations,
- Study self-alignment as a special case.

Specifically, we will present naturally spoken sentences in context to subjects. An undisturbed listener is expected to develop a state of normal alignment with the presentation of material, spanning all four levels of the P & G model. We refer to this as condition A.

We then create variations of condition A by modifying the presentation material such that a selective disturbance of the alignment condition A results. To this end, we will devise presentation variants designed to selectively provoke misalignments restricted to a single level (phonetic/syntactic/semantic/pragmatic). We will analyse the resulting EEG-patterns against the baseline condition A. We expect deviations to reflect the break-down of alignment at specific levels in the processing chain. A critical discussion point might be that disturbances at the same level (e.g., phonetic) can be caused in a variety of ways. We would only be interested in the common part of all these variations. To what extent such a common part can be extracted from the EEG patterns will be a critical element of our study.

A particularly interesting case will be the processing of metaphors: from the viewpoint of alignment theory, interpretation of the metaphor should be indicative of existing alignment at a high semantic/pragmatic level. Absence of such alignment should lead to a literal (non-interpreted) reception of the metaphor. By comparing the resulting EEG patterns for both cases (using "balanced conditions", where the probability of the interpretation of a metaphor is near 50%) we should be able to study the EEG-correlate of alignment at a very high level of processing.

To analyse the behaviour of neural assemblies from the EEG data and step forward to single-trial analysis and gain further insights into prototypical patterns, we aim to fuse coherence and phase analytic methods with machine learning techniques.

References

Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T. & Ritter, H. (2004). BCI Competition 2003 – Dataset IIb: Support Vector Machines for the P300 speller paradigm. *IEEE Transactions Biomedical Engineering* 51.

Meinicke, P., Kaper, M., Weiss, S., Müller, H.M. & Ritter, H. (2003). Using Maximum Contrast Classifiers for EEG data analysis. In: F. Schmalhofer, R. M. Young & G. Katz (Eds.) *Proceedings of the European Cognitive Science Conference*. Lawrence Erlbaum.

Pickering, M.J. & Garrod, S. Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences (in press).

Weiss, S. & Müller, H.M. (2003). The contribution of EEG coherence to the investigation of language. *Brain and Language* 85: 325-343.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	11,000	—	—	—
Consumables	800	800	800	800
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	1,200	1,200	1,200	1,200
Investment	—	—	—	—
Total	**16,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Equipment: High-end PC, Software (Matlab and Toolboxes), EEG electrocap, Face-mounted Display Olympus EyeTrek
- Other: Payment for subjects
- Student assistant with 19 hours per week

Project section B5: Gestalt approaches to alignment – from visual to verbal processing

Alexander Mehler, Gert Rickheit, Lorenz Sichelschmidt

Motivation

The interactive alignment approach to dialogue (Pickering & Garrod, 2004) postulates that representations at various levels are aligned, but lacks specific ideas about the cognitive mechanisms that bring about alignment. Some of the ideas addressed are reminiscent of the notion of Gestalt that European psychology featured some 70 years ago.

The reason for studying the correspondence between the notions of alignment and of Gestalt goes back to a comment by Johnson-Laird (1988: 19) on traditional Gestaltists: "They were right, but they lacked a language in which to make their ideas clear, and they sought laws of structure rather than an explanation of mental processes". One might wonder about the justification for stating that "they were right". Aren't we better off now, that we might have an appropriate language; now, that we are capable of studying mental processes on line, and now, that we have been warned to look for Gestalten inside the human mind rather than inside the material?

Objectives

Altogether, this project section aims at exploring in greater depth the viability of utilising classical Gestaltist approaches in the study of alignment in communication.

At that, the objective of the project section is twofold. On the one hand, Gestalt can be taken as an explanatory concept. This would suggest applying classical Gestaltist ideas – as typically developed in the field of (visual) perception – in the linguistic domain (specifically, with respect to the maintenance of interlocutors' conceptual representations). Such an endeavour would put the following research questions into focus:

- Is the notion of 'alignment in communication' – in particular, the routinisation aspect – expressible in Gestaltist terms, and if so, to what extent?
- Is it feasible to transfer particular Gestaltist ideas to the domain of verbal communication, and what are the prospects and limits of such an enterprise?
- Which alignment-related phenomena (e.g., abstraction or inference in language processing) can successfully be reconceptualised on a Gestaltist basis, and which cannot?

On the other hand, Gestalt is undoubtedly a concept that requires further explanation. From this point of view, the proposed project section provides a means to investigate classical 'laws of Gestalt' anew – on the basis of more sophisticated on line research techniques that are available now. Such an endeavour would put the following research questions into focus:

- Is it possible to reconceptualise classical Gestalt principles in terms of cognitive processes, and what are the prospects and limits of such an enterprise?
- What are the cognitive procedures at the bottom of Gestaltist regularities, and to what extent can they be empirically substantiated?

Methods

Research in the proposed project section is methodologically guided by the idea of an integrated experimental-simulative approach.

Experiments: Psycholinguistic experiments shall be carried out on diverse Gestalt-related phenomena (organisation, similarity, abstraction, inference, etc.) which lend themselves to study by on line and off line measurement techniques such as reading or naming, retention, and priming. The priming paradigm is particularly suited to investigate the conditions for and the time course of activation of particular concepts.

Formalisation and Simulation: One of the major drawbacks of applying Gestalt approaches in alignment research relates to the lack of formalisation. What is needed is a formal reconstruction and integration of Gestalt principles in order to account for compositional as well as non-compositional effects of Gestalt formation and to empirically test their predictions by means of corpus-based studies. This formalisation will be worked out in the framework of Latent Semantic Analysis (Kintsch, 1998). Furthermore, it will be used as a starting point for the development and evaluation of a toolbox for the implementation of Gestaltist routines in simulation models of text comprehension and production.

Work programme

While to date there is increasing interest in holistic, noncompositional approaches to language processing, the few attempts to apply Gestaltist ideas to language (e.g., Dhande, 2002; van Lambalgen, 2003) – in the first place, the idea of a "sentence gestalt" (e.g., Rohde & Plaut, 2003) – have remained half-hearted. We believe, though, that the Gestaltist approach bears in itself still unexploited potential, in particular with respect to explaining the cognitive mechanisms at the bottom of those routines that bring about alignment in communication.

For instance, analysis of dialogue suggests that Gestalt notions might be fruitfully applied at various levels of grammar. A case in point are so-called continuations and completions produced by interacting dialogue participants. For example, in a construction dialogue from the Bielefeld ‚toy aircraft' corpus, the instructor starts with *Well, now you take* and the constructor continues with *a bolt*. This also nicely demonstrates the link between Gestalt and alignment processes: The utterance fragment is completed to yield a declarative sentence, which, functionally speaking, is an indirect speech act, namely a request. The trigger for the completion in turn is the existence of a bolt in the situation, hence we have situational alignment adding up to lexical and semantic alignment.

With the intention of providing a holistic alternative to elementarist approaches, Gestaltists maintain that:

- In the course of processing, stimuli lend themselves to organisation in terms of Gestalten.
- Gestalten are distinctive, finite 'figures' (as opposed to ground) with an internal structure.
- The internal structure of Gestalten develops in the course of processing.
- Gestalten are processed as meaningful units which does not require decomposition.
- The imperfection of 'bad' Gestalten tends to be reduced in representation (Prägnanz).
- The perfection of 'good' Gestalten tends to be retained across situations.

These principles, originally developed for (visual) perception, easily transfer to the domain of language processing. They bear a close resemblance to psycholinguistic conceptions of cognitive routines in context integration, perspective taking, and "good enough" processing. In this sense, they constitute an agenda for the proposed project section.

On looking at some of the key principles of Gestalt psychology – the so-called 'Gestalt laws' – it becomes clear that these lend themselves to application and empirical testing in the domain of language production and comprehension.

- The Similarity Principle says, in broad terms, that elements which are similar in appearance are likely to be grouped. Likewise, we can observe facilitating effects of similarity in language: In syntax, for instance, pronoun resolution is facilitated by a match in gender, and in semantics, it has been shown that verbal information is integrated within but not across different scenarios.
- The Proximity Principle says, in broad terms, that elements which are similar in location are likely to be grouped. Likewise, in discourse processing, there are effects of surface distance on pronoun resolution as well as effects of stimulus onset delay on priming – which could be regarded as ‚semantic proximity'.
- The Principle of Common Destiny says that elements which vary in a similar way are likely to be grouped. Likewise, in discourse, it has been observed that dynamic change over time of a situation model affects all of its elements in the same way.
- The Principle of Prägnanz says that a set of elements is readily supplemented to form a 'good' Gestalt. The same is true for discourse processing: If it were not for error correction, noise reduction, and inference, interlocutors could hardly operate or cope with fragmentary or underspecified verbal expressions.
- The Principle of Meaningfulness says that information is readily interpreted as meaningful. A linguistic analogue to this can be found in Gricean maxims as well as in ways of overriding 'Donnellan'-type misreference and 'Moses'-type semantic illusions.

There are quite a few more phenomena that could be directly mapped from Gestaltist terms onto alignment. Contrast would be one class of phenomena, ambiguity or incompleteness would be another. It stands to reason that the abovementioned factors are not the only ones, and that they can occur in combination and, possibly, interaction.

Besides, it would also be worth an attempt to reformulate early Gestaltist ideas in contemporary Alignmentese.

References

Dhande, S.S. (2002). *A Computational Model to Connect Gestalt Perception and Natural Language.* Unpublished thesis; Cambridge, MA: MIT. Available from URL: http://web.media.mit.edu/~sheel/publications/

Johnson-Laird, P.N. (1988). *The computer and the mind.* London: Fontana.

Kintsch, W. (1998). *Comprehension – a Paradigm for Cognition.* New York: Cambridge University Press.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* (in press).

Rohde, D.L.T. & Plaut, D.C. (2003). Connectionist models of language processing. *Cognitive Studies, 10*, 10-28.

van Lambalgen, M. (2003). *Logical construictions suggested by vision.* Unpublished manuscript. Available from URL: http://staff.science.uva.nl/~michiell/docs/

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**110,400**	**110,400**	**110,400**	**110,400**
Small equipment	2,000	—	—	—
Consumables	1,000	1,000	1,000	1,000
Travel	3,000	3,000	3,000	3,000
Office supplies	—	—	—	—
Other	1,000	1,000	1,000	1,000
Investment	—	—	—	—
Total	**7,000**	**5,000**	**5,000**	**5,000**

Remarks:

- Equipment: Stationary experimentation system ("Longhorn" platform with "NBS Presentation" programming environment)
- Student assistant with 19 hours per week

Project section B6: Alignment of acoustic-phonetic representations in speech perception and production

Gernot Fink, Hans-Jürgen Eikmeyer

Motivation

In current automatic systems for perception or production of speech fundamentally different formalisms are used to represent the knowledge about the realisation of acoustic-phonetic speech events, e.g. phones or words (Huang, Acero & Hon 2001). For the purpose of speech recognition first a local analysis of speech is performed, which produces sequences of feature vectors. Then, statistical models of such sequences are trained from large databases of speech. These models – usually Hidden-Markov models (HMMs) – capture the variation in realisation for speech events in highly optimised feature spaces. However, there are hardly any efforts for exploiting this knowledge for production purposes. Instead, current speech synthesis systems basically – with only a few notable exceptions, e.g. (Tamura, Masuko, Tokuda & Kobayashi 1998) – completely rely on the concatenation of appropriate speech samples combined with the generation of a fundamental frequency envelope derived from the text to be synthesised.

Following the theory of alignment as put forward by Pickering and Garrod (2004) an interlocutor also implicitly aligns representations for speech production with those of perception in order to monitor the production of his own speech. Such an interaction can currently not be modelled in an artificial system due to the abovementioned diversity in formalisms used. Furthermore, as the alignment of representation can be viewed as a key element in learning of representations, "compatible" representations of speech events in both perception and production are of fundamental importance for the acquisition of acoustic-phonetic knowledge.

Objectives

The goal of the project is, therefore, twofold: On the one hand, a model for the intra-personal alignment processes involved in speech perception and production on the level of acoustic-phonetic representations will be developed. On the other hand, this model will be implemented in a technical system for integrated speech recognition and synthesis. The core element of this implementation will be a framework for the uniform representation of acoustic-phonetic knowledge. Moreover, this technical system will be used for simulating (i) impaired systems, both for language production and perception processes, and (ii) systems for language acquisition on the acoustic-phonetic level.

Methods

An important aspect of alignment is that is forms the basis of learning at the subconscious level. Therefore, representations suitable for the modelling of alignment phenomena should also be suitable for modelling learning effectively. The latter is not the case for the template based methods currently applied in speech synthesis systems. However, the statistical methods used for speech recognition purposes allow for an effective learning of model parameters from databases of speech samples. Consequently, intra-personal alignment at the acoustic-phonetic level will be modelled within a similar statistical framework. It will be derived from classical speech recognition models based on Hidden-Markov models by complementing these with the necessary capabilities for supplying controlled synthesis of speech events from the internal representations.

Work programme

The project will start in parallel with psycholinguistic experiments to identify the relevant alignment effects on the acoustic-phonetic level and the development of a statistical modelling framework that can be used both for speech recognition and synthesis purposes. In contrast to existing approaches that use HMMs for speech synthesis we will aim at a seamless integration of segmental and supra-segmental (e.g. prosodic) aspects of phonological knowledge. By means of the realised acoustic-phonetic alignment model effects of intra-personal alignment on the level of speech events will then be studied in a technical system and compared to evidence for corresponding alignment effects found in psycholinguistic studies. Finally, for demonstrating the effectiveness of the proposed model methods will be developed, that exploit the alignment of perception and production of speech for learning of new speech elements.

References

Huang, X., A. Acero, and H.-W. Hon (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Englewood Cliffs, New Jersey: Prentice Hall.

Pickering, M.J. and S. Garrod (in press). Toward a mechanistic psychology of dialogue.

Tamura, M., T. Masuko, K. Tokuda, and T. Kobayashi (1998). Speaker adaptation for HMM-based speech synthesis system using MLLR. In *Proc. Third ESCA/COCOSDA Workshop on Speech Synthesis*, pp. 273–276.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	1 SHK	1 SHK	1 SHK	1 SHK
Total	**98,400**	**98,400**	**98,400**	**98,400**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- 1 IIa BAT: computer scientist with background in pattern recognition, speech recognition & synthesis
- 1 IIa/2 BAT: linguist with background in psycholinguistics, speech perception & production
- Student assistant with 19 hours per week

Project section B7: Multimodal language processing: Dynamics and integration of parallel processes in verbal communication

Horst M. Müller, Friedrich Wörmann

Motivation

Verbal communication is not only based on the processing of morpho-syntactic, or semantic-pragmatic aspects, but also on the simultaneous analysis of prosodic, gestic, mimical or contextual information. In a given scenario-based communication, the listener's comprehension is both the result of distinct sub-processes parallelly analysed and the multimodal integration of discrete, neuronal information streams (Müller, 2003). Each of these discrete sub-processes is probably performed by distinct neuronal networks and may be observed by neurolinguistic techniques. A successful multimodal integration is necessary for the speed, the extent, and the robustness of alignment processes between speaker and listener. Stimuli demonstrating optimised, non-optimised and disturbed verbal behaviour will be tested with brain imaging techniques in order to study how such sub-processes will contribute to the alignment of different aspects of verbal communication.

Objectives

This project aims at analysing alignment processes and their dependence on multimodal integration with brain imaging techniques. The process of multimodal integration will be analysed by varying conditions with matched and/or unmatched parts of stimuli.

Methods

- Neurolinguistic experiments with multimedia presentation of stimuli, measurement of behavioural data (e.g., go/nogo or reaction time).
- Measurement of neuronal activity by functional magnetic resonance imaging (fMRI).
- Simultaneous application of fMRI and electroencephalography (EEG).
- Combined analysis of fMRI- (location), EEG coherence (parallel activity), and event related potentials (ERP) data (time).

Work programme

The main aim is to obtain empirical data on the integration of multimodal information during a scenario-based communication. For a given scenario the interaction between parsing of spoken language with matched and mismatched prosodic, gestic, and mimical information will be investigated. The influence of the successful multimodal integration on the speed and certainty of the listener's comprehension will be analysed. Subjects will watch and listen to a person speaking several sentences in context, accompanied by gestic and mimical information. The supporting or disturbing effect, e.g. of the speaker's pointing finger, can be analysed by varying the intensity or the duration of the stroke.

Besides using functional magnetic resonance imaging (fMRI) physiological data will be collected by applying a combination of fMRI and an electrophysiological spectralanalytic technique (EEG coherence analysis) for the first time. This combination would allow information to be obtained on parallel acting language-related processes, in terms of their temporal and local dynamics. The ability to distinguish different processes sharing the same brain area is necessary for the investigation of multimodal integration from a neurolinguistics point of view.

Thus, the following aspects of meaning constitution can be investigated:

1) Parallel processing: Which neuronal synchronisation processes can be separated during scenario-based verbal communication, indicating distinct cognitive processes in certain modalities?
2) Time course of parallel processes: After separating different sub-processes the exact time course of each sub-process can be measured. This would allow single steps of alignment during the scenario-based verbal communication to be monitored.
3) Localisation: Which brain areas are involved both in multimodal integration and alignment? Does alignment in communication share brain areas with other cognitive processes? Is alignment a common principle in terms of neuronal synchronisation?
3) Comprehension: In which way could different speaker strategies (e.g. using cross-modal information) influence the time course of comprehension in complex verbal communication?

By investigating physiological processes such as neuronal activity (fMRI) and neuronal synchronisation of cell assemblies (EEG coherence) (Weiss et al., 2004; Wörmann et al. 2003), we expect to provide empirical data, which should have a strong impact on the theoretical approaches to alignment in communication (Rickheit, 2004).

References

Müller, H.M. (2003). Neurobiologische Grundlagen der Sprache. In: G. Rickheit, T. Herrmann & W. Deutsch (Hrsg.) *Psycholinguistik: Ein internationales Handbuch*. Berlin: de Gruyter, pp. 57-80.

Rickheit, G. (2004). Alignment und Aushandlungsprozesse im Dialog. Zeitschrift für Psychologie (im Druck).

Weiss, S., Müller, H.M., Schack, B., King, J.W., Kutas, M. & Rappelsberger, P. (2004). Increased neuronal synchronization accompanying sentence comprehension. International Journal of Psychophysiology (in press).

Wörmann FG, Jokeit H, Luerding R, Freitag H, Schulz R, Guertler S, Okujava M, Wolf P, Tuxhorn I, Ebner A. (2003). Language lateralization by Wada test and fMRI in 100 patients with epilepsy. Neurology 61:699-701.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	20,000			
Consumables	1,300	1,300	1,300	1,300
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	1,500	1,500	1,500	1,500
Investment	—	—	—	—
Total	**26,300**	**6,300**	**6,300**	**6,300**

Remarks:

- Equipment: Video beamer for stimulus projection into the fMRI chamber, audio presentation system for fMRI, special EEG electrodes for use in fMRI, artefact rejection software for EEG during fMRI, transportable presentation computer with a fMRI interface, transportable mass storage unit and media
- Other: Payment for subjects
- Student assistant with 19 hours per week

Project area C: Systemic aspects of alignment

Alignment in communication is a central part of human social interaction. Research in many disciplines of social science has shown that in most cases social interaction is goal-driven and to a high degree partner and context dependent. Communicative conventions and strategies, the personalities of the communication partners, and the purpose of their interaction affect their mutual adaptation and their expectations for future interactions. In communicative alignment research the impact of the social situation on alignment processes must be taken into account, not only as an additional factor but as a necessary basis and precondition for many cognitive processes.

As Michael Schober and Susan Brennan (2003) put it:

> *"Given the complexities, it seems naive to imagine that we can determine, across discourse context, and for a particular aspect of language use (say word selection in production), what sorts of adaptations always occur. Presumably the adaptations vary according to conversationalists' (chronic or momentary) attentional capacities, discourse goals, interest in taking their partner's perspective, and so on)."* (Schober & Brennan, 2003, p. 155)

In Project section C, one main focus is on the combined effects of some of the factors mentioned by Schober and Brennan. The conceptual basis of this attempt is related to a systemic view of alignment in communication. According to this view, alignment is seen as a complex interaction of cognitive systems emerging in a more comprehensive social system with new capabilities. In Project section C, some of the most important factors governing these systemic processes of alignment will be investigated. The factors covered include *task orientation*, *time course*, *transfer* to novel situations and in learning contexts:

- As goal directedness is a general feature of social actions, many communicative activities are performed to serve certain functions. *Task orientation* in human communication will influence verbal and nonverbal communication and other social behaviour at different levels of the action hierarchy. In some cases, the interactions may not result in social convergence, but in misunderstanding and divergence. In addition, many communicative interactions rely on social problem solving in institutional settings (Giles, Coupland & Coupland, 1991). The management of communication includes many highly challenging multi-tasking and task switching efforts, which are supported by automatic and routinised cognitive mechanisms (Garrod & Pickering, 2004).

- Communicative alignment consists of processes of different duration. The *time course* of processing words, utterances, conversation turns, and conversation sequences comprise various cognitive and emotional processes including conceptual priming, model construction, strategies, and expectations. Short-term discourse processes may be more automatic than long-term processes, some of which are highly controlled and thoroughly planned (Snyder & Stukas, 1999).

- As present empirical research shows that communicative alignment is highly situation-dependent, the problem arises of how to handle the necessary *transfer* to novel situations in research and practical application. Recently, this crucial problem has been handled by closer consideration of the original situation (Anderson, Reder & Simon, 1996). In many learning contexts transfer of novel linguistic and conceptual knowledge is of high impor-

tance, e.g. in first-language acquisition (Tomasello, 2003), second-language learning and language therapy (Ferguson, 1999).

Considering the theoretical problems of research into situated processes of alignment – what are the adequate methods to deal with them? A plausible approach is the control and systematic variation of the most important situational factors as proposed by Schober and Brennan:

> "The evidence so far suggests that adaptation doesn't seem to be an all-or-nothing phenomenon at any level; people can be shown to adapt under some circumstances and not to adapt under others at virtually every level of language use – from higher discourse-level functions to articulation. Thus, we propose, the more fruitful agenda is to explore the factors that affect conversationalists' adaptations in particular circumstances – the sorts of tasks, individual ability differences, discourse goals, and so on that affect when and how partners can adapt to each other." (Schober & Brennan, 2003, p. 155)

In agreement with the research agenda proposed by Schober and Brennan, in Project section C the situational scenarios of communicative alignment will be carefully designed, controlled, and systematically varied. The resulting picture of alignment in communication will not only show the high flexibility of this type of social interaction, but also as converging evidence some basic and general cognitive strategies, which may be relevant for many situations of human communication. To reach this goal, a close cooperation between formal, empirical and simulative methods will be helpful (Eikmeyer, Kindt & Strohner, in prep.)

In project sections C1 (Wachsmuth, Sagerer, Hielscher-Fastabend), C2 (Metzing, Sagerer & Wrede) and C3 (Jäger) some fundamental questions of situated alignment will be addressed: How can alignment be supported by the interaction space constructed by the communication partners? What are the influences of alignment on the overall style of a specific discourse? Which processes take place in small group interactions?

The second group of project sections including C4 (Hielscher-Fastabend & Strohner) and C5 (Eikmeyer, Kindt, Strohner & Weingarten) will focus on specific problems and goals of human dialogue: 1. strategies to communicate emotions and emotive connotations and 2. routines to handle semantic understanding problems in repair sequences.

Finally, the third group of project sections C6 (Ritter, Steil & Sagerer), C7 (Ritter, Sagerer & Mehler) and C8 (Ritter, Sagerer & Jäger) addresses important problems of human interaction with computer programmes and artificial agents: search for verbal information, search for visual information and collaboration in sensory-motor activities.

References

Anderson, J.R., Reder, L.M. & Simon, H.A. (1996). Situated learning and education. *Educational Researcher, 25*, 5 – 11.

Eikmeyer, H.-J., Kindt, W. & Strohner, H. (in prep.). The extended experimental-simulative method.

Ferguson, A. (1999). Learning in aphasia therapy: It's not so much what you do, but how you do it! *Aphasiology, 13*, 125-132.

Garrod, S. & Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Science, 8,* 8-11.

Giles, H., Coupland, G. & Coupland, N., (Eds., 1991). *Contexts of accomodation: Developments in applied sociolinguistics.* Cambridge: Cambridge University Press.

Schober, M. & Brennan, S. (2003). Processes of interactive spoken discourse: The role of the partner. In A.C. Graesser, M.A. Gernsbacher & S.R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123-164). Mawah, N.J.: Erlbaum.

Snyder, M. & Stukas, A.A. (1999). Interpersonal processes: The interplay of cognitive, motivational and behavioral activities in social interaction. *Annual Review of Psychology*, 50, 273-303.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Project section C1: Interaction space

Ipke Wachsmuth, Gerhard Sagerer, Martina Hielscher-Fastabend

Motivation

In face-to-face communication, interlocutors make use of the space between them in various respects. They may look into each other's eyes to initiate or re-establish communication, direct attention by gaze and pointing gestures to objects present in the situation (e.g., on the table), or use iconic gestures to describe shapes and orientations of objects or spatial layouts while giving explanations. A particular observation is that interlocutors are likely to mutually align by sharing their gesture space with that of the other (i.e. the space in front of a person's body in which hand and arm movements are carried out while speaking), while they intimately converse about individual spatial features of joint current interest. In a scene where interlocutors are proximate to each other, this can result in an alignment of individual ways of referring to physical objects.

Objectives

Our aim is to understand instances of gestural alignment by way of "reinstantiations" of (deictic or iconic) gestures. An interaction model of alignment for such processes will draw on theories of imitation and postural mirroring between conversants (e.g., Rotondo & Boker, 2003) and include gaze. Such a model will enable a more natural interaction between an artificial agent and a human that share the same interaction space. In order to identify primitives for modelling the cognitive processes underlying the alignment of gestures used in shared interaction space, the comparison of normal communicative behaviour with aphasic/apractic use of gesture and pantomime is a further objective. Only few studies have tried to systematically analyse gestures of aphasics in a communicative setting, even less of them taking into account a frequent co-occurrence of aphasia with asymbolic and apraxic symptoms (Rose & Douglas, 2003). Altogether, our more far-reaching goals are

- to conceive and understand in detail the notion of interaction space and its role for alignment in communication,
- to investigate usage of interaction space in human interlocutors who verbally communicate about a spatial layout or the like,
- to enable (virtual or physical) artificial agents to observe and utilise coverbal gestures and gaze in the interaction space aligned with that of a human interlocutor.

Methods

Research methods will encompass both the empirical study of human behaviour in its "normal" and impaired functional levels with specific patient groups and its analysis to devise models of the processes involved, as well as the realisation and application of findings for interactions of a human with an artificial agent. Comparing the behaviour, producible with an operational model, with the originally observed real-human data will allow us to evaluate the model and inform its revisions in an expanding spiral, yielding a progressive understanding of the interpersonal processes under investigation. The imitation of iconic gestures by an embodied conversational agent (Kopp, Sowa & Wachsmuth, 2004) serves as a starting point in gestural alignment modelling.

The evaluation of the realised models will be performed with two different instantiations of an artificial interlocutor. The first evaluation system uses a virtual human ("MAX") to concentrate on studying the models underlying alignment phenomena that have been derived from empirical studies. The second evaluation system will make use of a robotic agent, with an artificial

human head with torso and arms, to focus on interactions in real physical environments that are different in nature from a virtual reality environment. The physical settings will require the extraction of human behaviour (gestures and gaze) from non-intrusive sensors, like cameras, and will first require the realisation of adequate recognition methods. An added value of these two settings will be the possibility of comparatively studying the differences of the alignment of humans with a virtually or physically represented artificial communication partner.

Work programme

The work programme will be centered around the idea of understanding instances of gestural alignment as reinstantiations of (deictic or iconic) gestures. Individual types of reference, e.g. by eye gaze, head movements, or pointing in "non-responsive" settings (i.e. with no interlocutor present) will be compared to reference in a face-to-face setting. Critical variables to be considered are time of reinstantiation, verbal context of reinstantiation, similarity of gesture in terms of orientation, speed, etc. Roughly, the methods applied are to

- empirically study human-human gesture alignment behaviour,
- devise a model from empirical findings,
- apply the model in two instantiations of artificial agents,
- evaluate and iteratively revise the model.

References

Kopp, S., Sowa, T., & Wachsmuth, I. (2004). Imitation games with an artificial agent: From mimicking to understanding shape-related iconic gestures. In A. Camurri & G. Volpe (Eds.): *Gesture-Based Communication in Human-Computer Interaction (pp. 436-447)*. Berlin: Springer (LNAI 2915).

Rose, M. & Douglas, J. (2003). Limb apraxia, pantomine, and lexical gesture in aphasic speakers: Preliminary findings. *Aphasiology, 17*(5), 453-464.

Rotondo, J. L. & Boker, S. M. (2003). Behavioral synchronization in human conversational interaction. In Stamenov, M. & Gallese, V. (Eds.): *Mirror Neurons and the Evolution of Brain and Language (pp. 151-162)*. John Benjamins.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	3 SHK	3 SHK	3 SHK	3 SHK
Total	**181,200**	**181,200**	**181,200**	**181,200**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Student assistant with 19 hours per week

Project section C2: Modelling of routines, discourse strategies and conversational styles

Dieter Metzing, Gerhard Sagerer, Britta Wrede

Motivation

Routines are focussed upon because they play an important role in language use (and its development) as well as in the development of advanced dialogue systems. From a linguistic point of view routines are a key element in the development of social communication, of common practices and coordinated behaviour among speakers/actors. From a natural language systems point of view current dialogue systems suffer from a lack of flexibility and adaptation to the communication partner. There are first approaches to include motivational and politeness strategies into artifical dialogue systems for tutelage. However, the application of such strategies generally follows static rules based on task-oriented goals. It would be desirable to provide a dialogue system with the capability to dynamically adapt to the communication patterns of the interaction partner (e.g. by using similar politeness strategies or routines).

In our research perspective there are connections between routines and rationality in everyday activities. Routines are chosen or modified by discourse strategies or conversational styles. Routines pre-structure the choice, realisation and coordination of activities (utterances, gestures, actions). We will analyse routines in terms of types of alignment and alignment as 'alignment of preferences'.

Objectives

The goal of the project is: modelling aspects of conventionalised communicative behaviour, reconstructed as ‘alignment of preferences’ with respect to two levels, one perceptual (routines of utterances, behavior and actions), the other one abstract (discourse strategies, conversational styles). ‘Alignment of preferences’ is related to different types of ‘common ground’ among the participants (ranging from ‘implicit common ground’ to ‘full common ground’), and it refers to preferences in the production as well as in the interpretation of communicative acts.

Methods

Different methods have to be developed for recognition/generation of directly observable routines on the one hand and for recognition/generation of discourse strategies and conversational styles on the other. These methods will be based on empirical analyses of structured multimodal corpora, either annotated on several levels and layers or without annotation, corpora for queries, statistic analyses or automatic learning methods. Research in formal pragmatics will be relevant for the reconstruction of discourse strategies and conversational styles. Sociolinguistic parameters of variation will be included.

Work programme

The project requires qualitative and quantitative modelling:
(1) Descriptive techniques for the empirical analysis of multimodal data and corpora (XML-based information modelling, i.e. methods for multi-level, cross-level, cross-modality annotation and retrieval) and representations for discourse strategies and socio-linguistic variation;
(2) Dialogue modelling techniques in order to model the alignment of preferred discourse patterns. This requires the ability to detect, recognise and predict such patterns. A very promising way of modelling this alignment process is to use statistical models such as Hidden Markov Models and n-gram language models. It has been shown that these techniques provide good results in the recognition of dialogue acts based on lexical and prosodic cues (Stolcke et al.

2000). It can be expected that additional cues such as gesture and mimic information can enhance the performance of a dialogue act recognition system. The recognition of dialogue patterns can then be used to drive the response generation process in an artifical dialogue system to produce utterances that are aligned to those of the communication partner.

Approaches in formal pragmatics may become relevant in two respects:
(a) By their contribution to modeling discourse strategies and
(b) By their contribution to an explanation of 'aligment preferences', routinisation and language use (cf. the project of Gerhard Jäger on dialogue games and group dynamics).

References

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., Meteer, M. (2000). Dialog Act Modeling for Automatic Tagging and Recognition of Conversational Speech. In: Computational Linguistics, Vol. 26, 3, pp. 339-373.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**110,400**	**110,400**	**110,400**	**110,400**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Student assistant with 19 hours per week

Project section C3: Dialogue games and group dynamics

Gerhard Jäger, Hans Strohner.

Motivation

Extant applications of game theory to linguistic pragmatics adopt the classical game theoretic assumption of perfectly rational players. This is in line with the standard picture of linguistic pragmatics. The alignment model of communication (Pickering & Garrod, in press) entails that these rationality assumptions are wrong in at least two respects:

- Language users construct an *implicit common ground* rather than a full common ground in most situations. This means that "players" think strategically in the game theoretic sense. This undermines standard solution concepts like the notion of a *Nash equilibrium* that play a central role in the mentioned models.
- If alignment is mediated via priming, frequency and recency of linguistic variants are important determinants in the choice of strategies. So the utility of a strategy depends on its susceptibility to priming – a factor that is not even accessible to conscious deliberation on the part of the language users.

The project is based on the working hypothesis that the actual dynamics of linguistic conventions is determined both by rational and by mechanistic factors. The investigations will focus on the dynamics in larger groups of people.

Semantic conventions: We will concentrate on short term conventions that are created within a conversation and do not necessarily persist beyond it. Candidates are metonymic expressions which are semantically specialised during the conversation (like *Dreier* for a building block with three holes in it) or semantic routines where the usage conditions for a complex phrase are more restricted than the compositional semantics determines (like *warming up* in the context of athletics, meaning "doing light physical exercises").

Group dynamics: Strategically acting agents will take their knowledge about their partner into account. Thus (non-)rationality in communication can be tested in conversations between more than two participants. Adaptation to the current addressee is an indicator of rationality (and vice versa). We envisage two setups. In communication in larger groups with possibly changing sets of participants, every agent has in principle sufficient information about the information state of the other participants. In iterated dialogues between pairs of individuals from a larger population (see Garrod & Doherty, 1994), rational vs. mechanistic interaction models predict different macro-dynamics for the entire population.

Objectives

The project aims at an investigation of the *dynamics of semantic conventions in conversations* and its modelling by means of game theoretic tools.

Methods

- Game theoretic modelling,
- Psycholinguistic experiments (priming, dialogue dynamics),
- Computer simulations,
- Mathematical analysis.

Work programme

The project will consist of three stages (which may overlap):

- Adaption of models of bounded rationality from economics to the application at hand. Obvious candidates are Evolutionary Game Theory – especially its stochastic variety (see Jäger, 2003 for a pilot study of an application of Stochastic Evolutionary Game Theory to linguistic problems) and evolution among agents that form expectations – and Behavioural Game Theory. These models mainly focus on *games in normal form* (so called "one-shot games", where all participants act simultaneously and only once), while *extensive games*, where players act consecutively and alternatingly, are better models for dialogue games. The evolutionary dynamics of extensive games is as yet poorly understood. In this stage, precise parametrised models will be developed.
- In the second stage these theoretical models will be tested empirically. We will conduct a series of psycholinguistic experiments where competing potential semantic conventions are implicitly offered to the subjects via priming, and their evolution within and across conversations in varying group setups are studied.
- In the final phase, the findings from the second stage will be theoretically evaluated. The hypotheses from the first phase will be adjusted and, if necessary, revised. Furthermore, the consequences of the empirically obtained models of micro-dynamics will be extrapolated to the macro-dynamics of the entire language community via mathematical modelling and computer simulations.

References

Garrod, S. & G. Doherty (1994). Conversation, coordination and convention: am empirical investigation of how groups establish linguistic conventions. *Cognition* 53: 181-215.

Jäger, G. (2003). Evolutionary Game Theory and Linguistic Typology: a Case Study. In P. Dekker & R. van Rooij, eds., *Proceedings of the 14th Amsterdam Colloquium,* ILLC, University of Amsterdam, 21-26.

Pickering, M. J. & Garrod, S. (in press). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**110,400**	**110,400**	**110,400**	**110,400**
Small equipment	2,500	—	—	—
Consumables	1,000	1,000	1,000	1,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	1,500	1,500	1,500	1,500
Investment	—	—	—	—
Total	**8,500**	**6,000**	**6,000**	**6,000**

Remarks:

- Equipment: PC for psycholinguistic experiments. Other: Costs for subjects
- Student assistant with 19 hours per week

Project section C4: Emotional alignment. Handicap and optimisation of intra- and interpersonal comprehension of emotional information

Martina Hielscher-Fastabend, Hans Strohner

Motivation

It is often stressed, that aphasics communicate better than they talk. This is obviously true for many contexts, in which aphasics only wish to express basic needs, or inform their partners about simple facts in a concrete and familiar context.

But many aphasics experience their expressive and comprehensive competences as very limited as soon as complex and abstract information has to be discussed. This limitation should also concern a differentiated expression and understanding of personal evaluations and emotional reactions. A consequence is incongruity of emotional utterances on the verbal and nonverbal channels. Because prosody, facial and gestural expression are often better presumed, communication partners have to deal with heterogeneous messages and may inadequately react to the verbal content. In addition, it has been shown (Hielscher-Fastabend, 2001) that aphasics do not interpret emotional information specifically correct, but catch only a raw idea of the basic emotion, that is expressed by their communication partner. Deficits clearly concern the verbal content and inferences, but also nonverbal information (Hielscher-Fastabend, 2004). No empirical data exist up to now, that focus on the communicative consequences of such emotional "misunderstandings", which may result in emotional "misalignment" of communication.

Associated with these problems may be problems some aphasics report on, which occur in defining their interpersonal relations in the way they did before stroke. Some aphasics lose dominance in their partnership or with their children. Problems of emotional alignment and of balanced processes of leading and pacing may play a major role for a satisfactory interpersonal relation.

Objectives

The project aims especially at the following topics. 1. The *relations between verbal and nonverbal competence in aphasics* have to be analysed in more detail. Deficits in intra-personal alignment of information from verbal and nonverbal channels are of high relevance for successful communication. Especially information about emotions and connotation depend on the correct integration of verbal, prosodic and facial information. In this respect, a close cooperation with the projects A7 and A2 will be fruitful. 2. The *influence of emotional processes on communicative success* in experimental settings as well as in normal communication will be of central interest in this project. Emotional alignment in communication is preliminarily defined here as an implicit and spontaneous mirroring of emotional style in facial expression, voice and prosodic aspects, and in emotional verbal form if agreement on the connotation of contents is reached by the interacting partners. A cooperation with project B4 on this second topic will be of special importance. Alignment processes and routines for emotional expression and connotation for instance may be differentially impaired in aphasic patients with additional conceptual or asymbolic symptoms. Another very frequent complication in aphasic patients is depression, which will have a special impact on communicative behavior. 3. *Developing framing strategies for close relatives of aphasics* means a logical consequence and a fruitful application of the results from studies on topic 1 and 2. Since emotions play a central role in close family relationships, the planned project aims at a better understanding of the role of emotions in communicative interactions between aphasics and their close relatives. The preferred theoretical framework is related to an integrative approach to communicative optimisation (Strohner &

Brose, 2002). If natural and automatic processes of modelling partners do no longer result in reliable interpretations, new routines have to be established.

Methods

As basic methodological tools we will use:

- Development of a screening for multi-modal emotion comprehension and expression,
- Psycholinguistic experiments,
- Therapy evaluation for communicative competence.

Work programme

In the first stage our project will focus on some intra-personal aspects of combinatory processing emotion information from different channels. Aphasics' competence in comprehension and production of emotion information has to be assessed, which can be done only in part by standardised methods. A screening for emotional communicative competences will be compiled (Hielscher-Fastabend, 2004) and extended by some new tasks and materials. Reaction-time data and eye-tracking experiments will add quantitative information about the time course of these emotion comprehension processes.

In the second stage, emotional utterances of aphasics and control persons will be analysed and rated concerning their affective and connotative content. Performance in isolated production and comprehension tasks will be compared to performance in communicative settings. A cooperation with the projects XY on nonverbal gestural and facial expression is important. In dialogue settings the thematic paradigm can be the interior equipment of a fictive holiday apartment as well as the interpretation and valuation of abstract modern art objects or emotional relevant video sequences. Emotion expressions, emotional connotation and their empathetic reactions will be considered in dialogue between unimpaired controls and will be compared to dialogue between aphasics and control subjects (confidante, therapist, relative). Only with a confidante are specific experimental variations possible. The confidante will have to show specific *emotional patterns* (extreme pacing, extreme leading, confronting) varying predominantly the *verbal, the prosodic or the facial expression*. Many studies show, that these experimental conditions are quite difficult to realise and to control. At least, post hoc ratings have to be conducted for the confidantes' utterances. An alternative may be seen in using a computational interaction partner with "emotional competences" which can be programmed to react in a specific way, as it is used in projects B4 and C6. Of course, differences of human communication and man-machine-communication will have to be considered.

In addition, the transfer of these results to training concepts of "Emotional Alignment Strategies" will be a main goal of the whole project. Many efforts of describing the processes in aphasia therapy mainly have relied on the therapists´ aims and the interaction sequences between therapists and clients. Most of these approaches focused on certain specific cognitive and behavioural learning processes, neglecting to a certain degree emotional alignment aspects between therapists and aphasic clients. In controlled intervention studies with aphasics and their close relatives, several training methods of emotional alignment strategies are compared. The results of these intervention studies are explained in terms of the emotional alignment processes investigated in the other parts of the project.

References

Hielscher-Fastabend, M.H. (2001). *Emotionskonzepte und Prozesse emotionaler Sprachverarbeitung.* Habilitationsschrift, Universität Bielefeld.

Hielscher-Fastabend, M.H. (2004). Comprehension of emotional information in patients with aphasia. *Folia Phoniatrica, 56,* 14-26.
Strohner , H. & Brose, R. (Eds., 2002). *Kommunikationsoptimierung.* Tübingen: Stauffenburg.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT	1 IIa/2 BAT
	3 SHK	3 SHK	3 SHK	3 SHK
Total	**122,400**	**122,400**	**122,400**	**122,400**
Small equipment	1,500	—	—	—
Consumables	1,000	1,000	1,000	1,000
Travel	5,000	5,000	5,000	4,000
Office supplies	—	—	—	—
Other	2,500	—	—	—
Investment	—	—	—	—
Total	**10,000**	**6,000**	**6,000**	**5,000**

Remarks:

- 1 IIa BAT: planning and conducting experiments with aphasic patients. The potential assistant has to have a double qualification in Clinical Linguistics (diagnosis and therapy) and empirical experience in Psycholinguistics.
- 1 IIa/2 BAT: planning and conducting psycholinguistic experiments, statistical analyses.
- 3 SHK: assisting experiments, corpus transcriptions etc., 19 hours per week
- Equipment: 1 Camcorder for outdoor video recordings in clinical contexts
- Other: psychometric tests and emotion stimuli

Project section C5: Repairs and reformulations in dialogue

Hans-Jürgen Eikmeyer, Walther Kindt, Hans Strohner, Rüdiger Weingarten

Motivation

Repairs and reformulations are syntactical patterns by which communicators solve their local problems in formulating and understanding quickly and successfully in a routinised way. In our project section we are especially interested in other-repairs and other-reformulations, i.e. cases in which the repair or reformulations are initiated not by the interlocutor who has run into a communication problem but by his partner. But we will also treat the intra-personal variants of self-repairs and self-reformulations.

Repairs and reformulations can be seen on the background of a scale of process properties ranging from negotiation to alignment. The leading hypothesis of the project is, that they are to be located more on the alignment side, but this has to be substantiated by empirical evidence. Since both formal and semantic-pragmatic properties of linguistic expressions give rise to repairs or reformulations, the position on the above-mentioned scale has to be determined for all relevant cases. Complicated cases of repair and reformulation rely on the alignment of cognitive representations, categorisations, expectations and plans of the communication partners. At the same time, some repairs and reformulations are not possible without complex semantic and pragmatic inference processes. Thus, the project connects alignment and inference research in the specific area of repairs and reformulations.

Objectives

One goal of this project section is the investigation of the alignment processes which form the basis of repairs and reformulations in dialogues with two or more communication partners. Specifically, we want to explain how communication partners reach the required adjustments of meaning by using minimal variants or minimal modifications of a formulation. A previous project has shown that in those cases an alignment of representations, and especially of plans of action and expectations, plays an important role (Kindt & Rittgeroth, in prep.).

The second goal is to explain how fast and resource-saving repairs really are. This leads, e.g., to questions about the information that is necessary for starting a repair and the control exerted on the language production process in order to carry out the repair (Schade & Laubenstein, 1993). This research will provide detailed insight into the delineation of alignment from explicit negotiation.

Methods

We will use an extended experimental-simulative method, which combines corpus-based, experimental and simulative methods. Using this method allows both theoretical modelling and experimental research.

Work programme

After the preparation of an initial corpus (see project section X1) existing theories of reformulations and repairs will be analysed in order to clarify the relationship between repairs and reformulations. Detailed corpus analyses as to which unfulfilled expectations initiate the use of other-repairs and other-reformulations and which special strategies are applied will be carried out. These will lead to the formulation of hypotheses for the conduction of psycholinguistic experiments.

In particular, we want to investigate the efficiency of specific strategies of repair and reformulation. Some of the relevant variables are the properties of the interaction partners and their interaction history and the task to be carried out. Since influences of media have been shown to be highly important for repair strategies, our research scenarios will include not only face-to-face-communication but also computer-mediated interaction. With on-line internet communication the interaction times of text production and repair interventions can also be measured exactly (Nottbusch, Weingarten, Sahel, in press).

In addition, as an important pragmatic influence on repair processes, the function of the dialogues will be varied. In one condition, the communicative scenario will be internet chat, and in an alternative condition subjects interact in an instructional setting.

As a result of the empirical research a theoretical model will be developed which explains processes of alignment in repair and reformulation interactions. Central aspects of the model will be evaluated by computer simulation. This work programme will be cyclically carried out in several instances.

References:

Eikmeyer, H.-J. et al. (1995). Coherence regained. In: Rickheit, G. & Habel, C. (eds). Focus and coherence in discourse processing. Berlin, New York: de Gruyter. p. 115-142.

Kindt, W. & Rittgeroth, Y. (in prep): Strategien der Verständigungssicherung.

Nottbusch, G., Weingarten, R. & Sahel, S. (in press): From written word to written sentence production. In: D. Galbraith, L. van Waes. & M. Torrance, M. (Eds.), Recent developments in writing process research. (Vol. 1). Dordrecht: Kluwer.

Schade, U. & Laubenstein, U. (1993). Repairs in a Connectionist Language-Production Model. In: Rieger, B. & Köhler, R. (Eds.) *Contributions to Quantitative Linguistics.* Dordrecht: Kluwer. p. 79-90.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	3 IIa/2 BAT	3 IIa/2 BAT	3 IIa/2 BAT	3 IIa/2 BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**106,800**	**106,800**	**106,800**	**106,800**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- 1 IIa/2 BAT: communication analysis
- 1 IIa/2 BAT: psycholinguistics
- 1 IIa/2 BAT: computational linguistics
- Student assistant with 19 hours per week

Project section C6: Adaptive alignment in human-robot cooperation

Helge Ritter, Jochen Steil, Gerhard Sagerer

Motivation

The project is based on an extension of the concept of alignment as formulated by Garrod and Pickering for the language domain into the domain of action modalities and starts from two initial hypothesis:

Firstly, we assume that processes of alignment are not restricted to support language and dialogue functions only but that similarly structured mechanisms as well support the *cooperative execution of common actions*. Conceptually this can be regarded as a logical extension of alignment from the field of verbal-communicative actions into the space of non-verbal cooperative action and an opportunity to investigate processes bridging the areas of language processing and production and action processing and execution.

Secondly, we have the hypothesis that the formation of alignment is *adaptive*: Repetitions of cooperative actions *facilitate the appearance of and raise the level of* the achievable alignment. This process of adaptation is an important component of the acquisition of *team expertise*.

Objectives

While testing the above hypothesis in Human-Human cooperation is an empirical issue, for Human-Robot cooperation it yields a *conceptually new approach* for the *dynamic structuring of a synergetic Human-Machine interaction*. The development of a Human-Robot interaction architecture, which is based on these ideas, its implementation and practical evaluation in a real world scenario constitute the main focus of the proposed project.

Methods

The main approach is to transfer the multi-level based concept of alignment developed by Pickering and Garrod to a multi-level hierarchy of sensori-motor actions. While language based alignment progresses from phonology up to semantics an analogous action hierarchy has to be identified which contains different control and planning levels, which represent different levels of increasingly abstract representations (eg: low-level compliance control, grasp configurations with respect to objects, hand positions in space, task level).

In this context, "Priming" will have to be realised in the form of coupled short term memories on the different levels for the robot to facilitate the reactivation of motor and action patterns which were traversed in previous episodes of cooperation. We regard such facilitated reactivation based on similarity of situative context as "action alignment" with the human partner.

Work programme

As a concrete real world scenario we propose the cooperation of a human, which uses one arm only, with an anthropomorphic dextrous hand of a 7-DOF robot arm. Such a situation is of practical relevance, if only one hand of a human is available for task execution, for instance due to a handicap. Furthermore, bimanual manipulation is important for many every day tasks, such that the cooperative endorsement of a single hand by a robot assistant is an important and scientifically interesting task.

Because of the high structural analogy to language based alignment we expect that the proposed multi-level architecture will be especially well suited for integration with a (parallel operating) language processing path of alignment. This provides the perspective for close cooperation with corresponding projects in the speech processing and understanding domain, which starts on a more conceptual level in the beginning, and can increasingly progress towards integration of developed modules and opens the possibility to investigate inter-modal self-alignment.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Student assistant with 19 hours per week

Project section C7: Semantic alignment for keyword-based database front-ends

Helge Ritter, Gerhard Sagerer, Alexander Mehler

Motivation

Search engines, for instance Google, have recently become very useful and important every-day tools mainly by means of a keyword-based indexing to immense amounts of data. From an alignment-based viewpoint, however, this keyword-based access is very rudimentary and ignores other modalities and more complex interactions a user could have with the database engine. Contemporary dialogue systems act at a complementary level of complexity: they use highly developed speech processing interfaces, which, however, typically provide access to very small knowledge databases.

From a cognitive perspective, access to large amounts of data seems to be the more important factor to establish alignment: developing children already have a significant world knowledge at their disposal even while they use language in a more rudimentary, keyword-based way. This motivates the search for ways to incrementally improve the alignment of a user with a database engine by means of recognising and maintaining simple verbal context and semantically meaningful categories.

Goals

The main focus of the proposed project is to develop a search engine front-end, which allows verbal and language based alignment to interact with a keyword-indexed database engine. As a platform for such a dialogue system we will use a humanoid robot head, which allows to provide the user with intuitive feedback. The ultimate goal is to enable the machine to pass a "broken-language Turing test": the system must be able to conduct a natural language-based dialogue, whose degree of perfection is measured in terms of semantic plausibility rather than in grammatic correctness.

Methods

The main elements to enhance a standard search interface with language-based and semantic elements will be: (i) inclusion of temporal language context, (ii) thereupon-based assembly and maintenance of situational state information on several sematic levels of abstraction, (iii) an intentionally simple dialogue interface design, which is oriented to early child language.

Methodically, new combinations of statistically oriented approaches originating in data- and text mining and of computer linguistic methods for dialogue systems have to be explored. To implement "semantic priming" to establish the necessary alignment we intend to use WordNet.

Work programme

The final implementation of the proposed dialogue systems has to proceed in several steps and first has to be tested on a controlled reference system, which allows the development of proper evaluation measures. Concrete intermediate steps are

- definition and setup of an index-based reference database and processing by WordNet,
- implementation of a multi-level hierarchy of abstraction for word, similarities, expressions, and simple context up to sematic categories,
- implementation of temporal "priming" dynamics,
- design of a simple dialogue system oriented at child language,

- evaluation in form of "broken language" Turing test,
- application and evaluation of the frontend for web-based large databases.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Student assistant with 19 hours per week

Project section C8: Visual alignment dynamics for feedback in cooperative search

Helge Ritter, Gerhard Sagerer, Gerhard Jäger

Motivation

Starting from the hypothesis that alignment is not only a fundamental process for effective, and resource efficient coordination of human verbal communication but also affects the visual modality it must be considered an *important factor in designing visuo-verbal Human-Machine cooperation*. Therefore, a visual- and dialogue-based search in image databases is a natural setting to study the potential of visual alignment. In particular, content-based search in image databases is difficult to achieve with more traditional approaches mainly because it is difficult to organise a semantically interpretable and meaningful feedback from the user to the machine. Here, the evaluation of the degree of alignment of user and machine, measured in terms of attention and visual behaviour with respect to the presented search results, offers innovative methods to improve the human-machine communication. Complementary, the synthesis of machines oriented to requirements and functional attributes of alignment provides (because of the privileged "accessibility" of an artificial system) an interesting approach to *investigate the process of alignment* itself.

Objectives

The proposed project aims at treating *alignment as a new, enhanced concept of classical relevance feedback* to support a cooperative, content based search in image databases. Evaluating the dynamics of alignment while a search is progressing, provides additional cues for the system, whether or not it is improving the presentation of search results. Adaptation of the machine with respect to the user's query is, by online-learning, based on the feedback signal generated from measuring the degree of alignment. The main goal is to enable the search engine to present a subset of search results, which is with respect to the user, better filtered.

Methods

Classical relevance methods employ user feedback about the degree of appropriateness of search results from the image database so far found and presented. The multi-stage theory of alignment provides a suitable framework to fundamentally ground the concept of relevance feedback from the perspective of communication, which was previously mainly based on statistical and pragmatic arguments. It furthermore allows us to improve our understanding of cooperative search in large databases along a cognitive dimension. Recording and evaluating eye movements as an indicator and constituting factor for alignment as well as new multimodal forms of feedback can thus significantly support cooperative search.

Work programme

- Transfer of the multi-stage theory of alignment to the situation of cooperative, content-based search in image data,
- Development and implementation of a multi-level representation of alignment between the human and the image database system,
- Conceptualisation and implementation of an suitable *alignment dynamics* in this framework,
- Online-adaption based on alignment feedback for iterated search queries,
- Evaluation of the framework in concrete search tasks.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	2 IIa BAT	2 IIa BAT	2 IIa BAT	2 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**141,600**	**141,600**	**141,600**	**141,600**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Student assistant with 19 hours per week

Project section X1: Development of XML-based multimodal alignment corpora

Alexander Mehler, Hans-Jürgen Eikmeyer, Dieter Metzing, Hannes Rieser

Motivation

The project focuses on the empirical resources of the CRC projects and on the documentation of their analyses. This concerns the theory-driven multilevel categorisation, the cross-level and cross-modal linking of multimodal alignment data and the output of the data analyses performed in the SFB projects.

Since the annotation of alignment data demands the integration of information concerning several levels of description, different modalities, actors, and dynamically reorganising interaction processes, very powerful annotation schemes and tools using standardised representational formats are needed. Although current corpus technology can profit from experiences and solutions obtained in research efforts in this field (1), corpus-processing of empirical and experimental multimodal alignment data and their analyses is a rather ambitious goal which will lead to new types of corpora, tools and methods.

The goals mentioned require close cooperation between this and other CRC projects. This will introduce an additional level of exchange, coordination, sharing of resources or results and of integration. The role of the project will be to inform the other projects about existing methods and tools, to contribute to their development as well as to carry out research on information modelling.

Objectives

The project focuses on the development of extensible, flexible formats for the representation of alignment corpora, which allows the modelling of the multimodal manifestations of alignment processes, their standardised annotation and statistical analyses. This relates to the following sub-goals:

- In order to make use of standardised schemes and tools for the analysis, annotation and retrieval of multimodal data, the representational formats to be developed have to be based on the eXtensible Markup Language (XML).
- Further, these formats should allow the extension of the number of description levels (and their visualisation, for example, as tiers) to be accounted for in the projects.
- The formats should also allow the interrelation, ordering and visualisation of units belonging to the same or different description levels according to arbitrary scales (e.g. timeline).
- The formats to be developed need to support the annotation of several synchronised data streams (e.g. videos).
- The system of categories defining a single description level should be extensible. It should also be possible to enhance the set of attributes assigned to a single category in a flexible way.
- It should be possible to merge several compatible annotations covering the same data, for example, if gesture and speech of a video film were annotated separately and one wants to get at the combined information.
- It should be possible to distinguish between core dimensions of annotating multimodal data interesting for all SFB projects and project specific extensions of the description levels and annotations.

- A uniform system should be used for the management of alignment corpora which allows easy data interchange between the projects. This also relates to the provision of flexible, document-oriented views on the alignment data.

In order to approach these goals, existing annotation schemes and tools have to be evaluated and tested. Because of the need for flexible and extensible data modelling and analysis it will be necessary to develop an *alignment corpus management system* which integrates the different annotation tools to be used and the annotation data they produce.

Methods

The development and testing of the formats for representing multimodal data will be based on methods of XML-based information modelling. The alignment corpus management system will be implemented in accordance with the principles of information systems development.

Work programme

The project is divided into four stages: The requirement analysis (collecting and systematising the requirements of the CRC projects with respect to the representation, annotation, storage and retrieval of alignment data) occurs parallel to the evaluation and testing of existing annotation schemes and tools. These two stages will be followed by the development of representational formats which account for the several aspects of flexible and extensible data modelling and annotation enumerated above. After this phase, the design and implementation of the alignment corpus management system and finally its evaluation and testing are performed. Concurrent to all four stages, the project will instruct and advise the usage of selected/developed annotation schemes and tools.

(1) cf. research related to :
the HCRC Map Task Corpus (http:www.hcrc.ed.uk/maptask/)
the NITE Project (Natural Interactivity, Tools and Engineering) (http://nite.nis.sdu.dk/) and the research group 'Texttechnologische Informationsmodellierung' (http://www.text-technology.de)

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
	2 SHK	2 SHK	2 SHK	2 SHK
Total	**82,800**	**82,800**	**82,800**	**82,800**
Small equipment	25,000	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- Equipment:
  Audio equipment for recordings:
  1 pc. 8 channel mixer e.g. MOTU Audio 896HD, firewire; price 2,000 EUR
  4 pcs. Sennheiser microport sets, sender, receiver, headset; price: 4 x 2,000 EUR
  Video equipment for recordings:
  2 pcs. Digital progressive scan camcorder, 1.2 mpix e.g. JVC GR-PD1, price 2 x 2.500 EUR
  Server computer for corpus data:
  1 pc. PC-Server, Win64, 16 GB RAM, 480 GB disc space, Tamino XML Server; price 10,000 EUR
- Student assistant with 19 hours per week

Project section X2: System evaluation

Stefan Kopp, Jochen Steil, Britta Wrede

Motivation

In the general framework of the proposed CRC we regard alignment a complex multi-level, multi-modal, and multi-party phenomenon and, naturally, different facets have to be studied in depth in separate projects. One essential step of the experimental-simulative methodology advocated for our CRC is the evaluation and the measurement of the degree of alignment achieved in these various settings. There are a number of evaluation approaches available, for instance, from dialogue study (Bernsen & Dybkjær, 2004), eye-tracking, or statistics, and these methods would normally be applied independently and tailored to the specific research issues (cf. Dagstuhl seminar, March 2004). However, there is also a need for systematic and common cross-level and cross-modal measures that allow us to evaluate and account for the degree of alignment achieved, to compare our results, and, ultimately, to facilitate knowledge transfer and cross-fertilisation among the project sections.

At least two groups of new questions must be tackled to this end: (1) *within a modality*: what are suitable indicators to measure degrees of alignment from observable interaction data, especially for a technical system in an interaction with a human user (in domain-oriented dialogues, social aspects, etc.); (2) *across modalities:* how to combine and weight modality or level specific measures of alignment to generate a more comprehensive account of the alignment taking place between the interlocutors?

To ground such evaluation theoretically, advanced overarching notions of alignment that relate and combine specific aspects and partial results have to be developed. This will lead to new benchmark experiments and, thereby, promote the coherence of the proposed CRC as a whole. To implement this overall important aspect, the project will be structured in a special way: in an initial phase of theoretical and conceptual work the project will be supported by a CRC-wide working group. In the years 3 & 4, when single project sections have reached the evaluation stage of their work programme, to tackle this step cooperatively, it is planned to partially reassign people from the particular projects to the System Evaluation project (4x1 IIa/2 BAT from A5, B1, B3, C6).

Objectives

Current techniques mainly address the evaluation problem as measuring system performance with respect to a given task. A measure of alignment has to go beyond this approach because alignment as a property of communication deals inherently with *how* a goal is achieved as opposed to only *whether* (obviously, a high degree of alignment can be present even when failing to complete a task). Furthermore, measures of a system's alignment capabilities need to be specified both for different kinds of systems (for recognition, generation, motor-interaction, natural dialogue, etc.) working in different domains/scenarios and as for different kinds of alignment (intra-personal, at different levels, inter-personal, situation-dependent, etc.). For these different kinds of situations, one goal of our project is to come up with working definitions of alignment (e.g., as adjusting frames of representation that structure the abstract interaction spaces/dimensions), and to develop comparable measures for the corresponding reference frames. The ultimate goal will be to find overarching definition(s) that integrate both local and global aspects (situation models, lexical processing, syntactic alignment/priming, articulation), both in production and perception/comprehension as well as across different modalities.

Methods

Definitions and methods developed in the projects for examining specific kinds of alignment in specific kinds of systems (e.g., eye-tracking for intra-personal alignment; methods of micro- and macro-level analysis for inter-personal alignment) will be collected and re-conceptualised to become integrated and comparable in a common framework. This will enable the evaluation group to define cross-modal methods and settings to develop benchmark problems that allow for assessing progress within single projects as well as comparing it between projects. Of special importance here will be to learn from statistical evaluation from experiments with humans (reference to corpus). To integrate this with machine-oriented techniques, we will draw on experiences with multi-dimensional frameworks used in machine learning design. Finally, in close cooperation with the respective projects, we will conduct comparison and evaluation experiments with their first prototypes.

Work programme

Theoretical and conceptual work has to be conducted to adapt existing definitions and evaluation approaches from dialogue theory to multi-modal settings, and to identify suitable alignment measurement dimensions including quantitative (temporal, spatial, similarity or keywords) as well as qualitative (semantic similarity, success of goal-directed interaction, etc.) ones. This work will be supported by the project spanning working group. The more practical evaluation work will be conducted together with the collaborators reassigned from the projects in the years 3 & 4. This results in the following work programme:

- Identify alignment measurement dimensions employed in projects,
- Identify missing dimensions and develop suitable methods to close gaps,
- Identify further situation-dependent factors,
- Establish evaluation methods for cross-modal integration,
- Establish a common evaluation framework for project-specific definitions and approaches,
- Define benchmarks and experiments,
- Perform an array of evaluations.

To this end, a number of experiments will be conducted in close cooperation with the respective projects and with the help of the respective additional staff. The results of these evaluation studies from the project sections will also be compared to find additional, situation-depending factors that influence alignment and system performance more generally.

References

Dagstuhl seminar on "Evaluating Embodied Conversational Agents", 15-19 March, 2004, http://homepages.cwi.nl/~zsofi/eeca

N.O. Bernsen, L. Dybkjær (2004). Evaluation of Spoken Multimodal Conversation, Proc. *ICMI'04*, to be published.

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 IIa BAT	1 IIa BAT	1 IIa BAT	1 IIa BAT
Total	**58,800**	**58,800**	**58,800**	**58,800**
Small equipment	—	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	3,500	3,500	3,500	3,500
Office supplies	—	—	—	—
Other	—	—	—	—
Investment	—	—	—	—
Total	**5,500**	**5,500**	**5,500**	**5,500**

Remarks:

- 4 IIa/2 BAT funding from other projects sections in the 3rd and 4th year

Project section Z: Central services

Spokesman for the SFB

Auxiliary funding

	2006	2007	2008	2009
Funding for staff	1 Vb BAT	1 Vb BAT	1 Vb BAT	1 Vb BAT
Total	**42.000,00**	**42.000,00**	**42.000,00**	**42.000,00**
Small equipment	3,000	—	—	—
Consumables	2,000	2,000	2,000	2,000
Travel	2,500	2,500	2,500	2,500
Office supplies	4,000	4,000	4,000	4,000
Visiting researcher funds	10,000	10,000	10,000	10,000
Colloquia	15,000	10,000	10,000	15,000
Total	**36,500**	**28,500**	**28,500**	**33,500**

Remarks:

- Equipment: PC for the service project section
- Travel: Costs for spokesman's journeys within Germany
- Colloquia, researchers: International colloquia and researchers, "Klausurtagung"