

PRAGMATIC CONSIDERATIONS IN MAN-MACHINE DISCOURSE

Walther v. Hahn

Research Unit for
Information Science and Artificial Intelligence

UNIVERSITY OF HAMBURG
D-2000 HAMBURG 13, West-Germany

Introduction

This paper presents nothing that has not been noted previously by research in Artificial Intelligence but seeks to gather together various ideas that have arisen in the literature. It collects those arguments which are in my view crucial for further progress and is intended only as a reminder of insights which might have been forgotten for some time.

Research on discourse has achieved remarkable results in the past decade. The standard has been raised from simple question answering to dialogue facilities; a fact which, as we all know, implies much more than only extending the borderline of syntactic analysis to the level of more than one sentence and more than one speaker.

However, at the same time we all know that the reality of discourse is nearly as far away as before from what we are able to model now. It's certainly not worth enumerating all the deficiencies of current models and to list what the real goals are. It's a matter of every-day experience to see the "blatant mismatch between superficial human ease and theoretical mechanical intractability" (Berwick (2)27).

The situation seems similar to that of modern linguistics, "which has tried its best to avoid becoming entangled in the complexity of conversation, but has been gradually forced in this direction by uncooperative data" (Power / dal Martello (23)).

Though it is one of the so called "good old" traditions of science to eliminate a lot of the most difficult questions by saying 'this is not our job', in AI, however, from the cognitive point of view we must realize that everything is our job. "Its dirty work, but somebody's got to do it" (Israel (18)).

This seems rather contrary to what Berwick (2) shows in his 'Cook's tour' around the geography of dialogue, where everything fits together in an overall map and where modularity is a virtue: Our knowledge of natural discourse processes is highly insular without bridges in between; and: the reality of discourse is complex in that

everything is contingent with everything else; in fact, nothing is 'modular' in this sense (see Fodor (11)).

What I will do in this paper is to show all this as the patchwork it is and to encourage to approximate the alternatives seen so far. In a lot of fields of dialogue research the discussion often is characterized by an 'either-or'-view whereas we should try to find a 'as-well-as'-solution or even another new path of research.

Looking at the results of our work we have to accept at least three lines of progress which all have their own merits, namely

(1) developing new concepts, based on new integrating ideas, even if only limited implementation or other proofs of feasibility might be possible at time,

(2) the unfolding of these ideas by theoretical background work and experiment in all detail. The result of this work could show the intractability of such approaches or prove that this approach can be mapped onto a known solution (as e.g. Johnson-Laird (19) has tried to show for meaning postulates and decompositional semantics).

(3) the exploitation of the ideas in constructing working systems which may show whether or not the idea passes the feasibility test.

The general feeling in Artificial Intelligence now seems to be either resignation or particularization of the problem of discourse. The alternative after all is not doing everything at the same time in one single ultimate system or doing nothing, but we must go on to fit together the great puzzle even if there are a lot of missing pieces in areas which we have already attacked.

The first Challenge: Integration of Perception and Function

The practical view tends to be restrictive in its approach to perception because this is the world of the naive user of practical systems. People involved in natural dialogue

obviously hear or read only words, move objects or manipulate symbols. They know that their intuitions about the role and function of the symbols might be wrong, but all cognitive actions are triggered by more or less physical objects. And, what is even more important, naive users are sure that the visible words or even cursor positions coincide with the function intended by them. It is always difficult, e.g. to demonstrate users ambiguities in their utterances. People are surprised when you explain indirect speech acts to them.

Scientists, on the other hand, have to reconstruct a series of hierarchical abstract levels and internal representations and often enough they get lost in their own symbolic maze and have to invent more and more artificial tricks to climb out of their constructions and still meet the surface of the utterance.

Of course, it is hopeless "to seek meaning in the physical properties of utterances and formal properties of language. However, the simple fact is, that speech is merely noise until its potential meaning is appreciated by the cognitive activity of a hearer" (Harris/Begg/Upfold (16)).

There is a good example which shows that there are even cases in which you cannot decide whether you are talking about objects or words or abstract constructions. Sidner (24) introduced the notion of "cognitive cospecification" for the following example to show that some anaphors cannot be replaced by a literal antecedent in any previous sentence:

"My neighbor has a monster Harley 1200
They are really huge but gas efficient bikes"

Another good example for the non-uniqueness of visual perception: Conclin and McDonald (7) tried to built their techniques of generating image descriptions on their observations what people found worth describing in photographs. But what these interviewee found salient was highly dependent of the context of the request for description. And this is a matter of pragmatics. "There is no salience in a vacuum" (7).

This discrepancy is reflected also in Butterworth's (5) 5. and 6. maxims for the linguistic study of conversation which state:

- "5. Let the theory do the work!
6. Let the phenomena guide the theory!"

In Artificial Intelligence it is not only the practical point that normally interaction is restricted to the screen, the keyboard and the mouse, that is, the surface of systems is the only visible link for the user as well as in principle for the knowledge engineer. Moreover, It is necessary to compare always the behaviour

of a system with what a user expects to see as an indication of the expected function, because it guides the intuition of the system's partner anyway. Careful concentration on what the user sees and expects to see even if the system fails to react properly is one of the best means of a pragmatically adequate treatment of discourse in Artificial Intelligence.

Some of the addressed problems can be reformulated on another level as

The 2nd Challenge: Integration of Intuition and Idealization

The representation of knowledge, especially the way logicians look at it, has often been the starting point of a long discussion, how natural, how plausible a specific representation is in comparison to underlying cognitive processes. Of course you can and should (at least to keep consistency) map all systematic representations onto a logical notation. But logicians and linguists all rely on their intuition in creating their significant examples and counterexamples for representation problems. Power and Martello (23) criticizing ethnomethodology say in their maxim (2): "There is no reason why intuitions about invented examples should be ruled out as a method of investigating conversation" .

Why do they argue with intuition in respect to what they represent but not in respect to how they represent it?

It is an accepted idealization among linguists, logicians and Artificial Intelligence researchers that input sentences must first be represented in an internal language. And then we are at home in our theories and can start our tricky algorithms. In syntactic analysis, e.g. we normally attach the syntactic categories to the input words by means of a "syntactic lexicon". It must always be clear that all this is highly counterintuitive for any naive speaker. We have in fact no ultimate reason for doing so except the argument that we see at the moment no way to proceed the input in another rule-guided way.

Normally we have no cognitive reasons for choosing exactly the representations we use. Consider the ideas of Langacker's (21) "cognitive grammar". His ideas, though he might be fallacious about drawing graphics being better than writing down predicates or operators, show that there are lots of plausible ways to talk about semantics and grammar. Doubtless we are often bound to topographical or space-oriented concepts in our linguistic intuition.

It goes without saying that we understand texts and sentences as the primary units, not words or morphemes or quantifications. Our understanding is supported by our visual memory, by acoustic memories, and by other

complex experience in the past. Only when we are forced to understand tricky linguistic or logical examples or must understand defective, illformed or mistyped utterances, (and we have no opportunity to initiate a clarification dialogue!), only then we will start up our analytic linguistic processor and check rules, endings, positions of words etc. (cf. experiments with garden-path-sentences).

There is no point in arguing with the incremental understanding of sentences by a listener as he hears each word. We certainly do not perform structural analysis word by word. At best we check structural constraints for our semantic/pragmatic hypotheses.

I would even go further. Presumably the intuitions do not contain any clear concept of understanding as long as there are no misunderstandings (I will come to this point later). And even then, as Goodman (12) shows, "people must and do resolve lots of (potential) miscommunication in everyday conversation. Much of it is resolved subconsciously - with the listener unaware that anything is wrong".

Ellman (8) even claims that "the classification of indirect speech acts is primarily for analytic purposes and it is not stated anywhere that this classification is essential to the understanding process".

To oversimplify: The basic intuition of a naive user refers to a SELF and a SYSTEM, which works by telepathy, superficially guided by the linguistic utterance of the user.

What you will object to is quite clear: "Intuition", as used here is a pre-scientific label for all the unsolved problems of complexity arising in every advanced implementation.

In a sense you are right, because a lot of the unsolved problems might perhaps arise from the fact that the solutions known so far are counterintuitive. But to be serious, the notion of intuition alone is too vague. We must at least define: the intuition of whom? Grosz (13) showed that any application oriented natural language interface must regard the intuitions (diverging on different levels) of a potential user as well as of the database expert (knowledge engineer).

Much more general is the objection that intuitions concerning plausibility of a system's surface exposed to the user is not a static affair. In the course of the work with a specific system a user will change his intuitions about the appropriateness of its behaviour and its interpretation of the user's utterances. As far as I know there is no comparative research on the dynamic pragmatics of long-term use of a system.

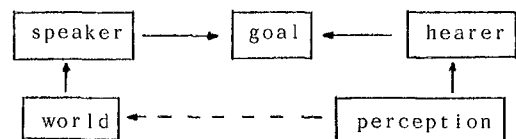
A weighty reproach, however, comes from a methodological point of view, expressed by Carroll and Bever (6). In experiments of semantic adequacy ratings one group of test persons were heavily biased in their intuition by the fact of sitting in front of a mirror. This mere matter of the setting changed the ratings so much that the result of one group would fit well to hypotheses of general semantics whereas the other group's result would rather back generative syntax. Such are intuitions.

But in any case listening to what people think they are doing and the system is doing is one of the most surprising heuristics and we definitely always need this corrective instance to construct systems which are pragmatically more adequate.

Let us now have a closer look to the process of scientific idealization: we normally do not only start with the translation of the data in a form which we can handle, but we also divide the whole problem of human discourse into subproblems and sub-subproblems. This is, similar to the translation paradigm, another "good old" tradition which we tacitly accepted; of course, we cannot do everything at the same time. But this technical routine has been internalized in an extremely strong way and is not longer only a crutch of science. What I address here is the opposition of particularism and holism.

Israel (18) criticizes the ideal of modularity as a concept being imported from traditional linguistics and psychology. Their conceptions of correctness are modular, perhaps because of the lack of procedural theories and the lower degree of formal complexity in their models, because of the lack of procedural representations of their models by means of implementations. In Israel's view the main fallacy in discourse models is that "modularists" try to solve syntactic and semantic processing first and than see what they can do for pragmatics additionally. Even syntax in the theories of these hopeless "syntacto-semantic imperialists" (18) is clearly divided into sentence-by-sentence and level-by-level processes. And once we have cut the problem into pieces we forget even to try to fit it together again afterwards.

In my opinion we are moreover too accustomed to boxes and arcs for illustrating of our ideas in AI. Figures as the following corrupt clear communication:



In opposition to these simplistic views of language, neurolinguistics has shown that understanding is a sort of pattern (re)construction working freely through different levels of abstraction between the level of physical perception and understanding or the reaction respectively.

We can apply holistic ideas anywhere: Appelt's (1) arguing for unification in grammars as a very elegant way to pass pragmatic features through different levels of a language processing system is a good example.

Another example might be the opportunistic planning by Hayes-Roth and Hayes-Roth (17) which, from a cognitive point of view, can model human planning behaviour in a very convincing way. The fact that they start with isolated tasks and then put together chunks of pre-planned actions is no argument for modularity because there is no intermediate built-in level of completed substructures. So the incremental strategy of the HEARSAY II-architecture fits much better to the holism of the understanding process.

Though there might be other good reasons for preferring modular implementations in today's work: Let us try to achieve again a holistic and intraintuitive model of human dialogue processes.

The 3rd Challenge: Integration of Different Sources of Plausibility

The main process of idealizing the data is to evaluate the phenomena in respect to their importance for further treatment. But where do the criteria of this evaluation process come from? One possibility is to rely on the background sciences e.g. linguistics, psychology, sociology etc.

In comparison to the 70's there is indeed much more cooperation with what I called the background sciences. As Brady (3) remarks, Artificial Intelligence has overcome the first years in which we thought that the very specific view and the methodological implications of Artificial Intelligence were so extremely different from everything in the past, that we had better start again thinking about language and cognition in our own paradigm.

This has become better now even though I think that there is too little cooperation with sociology e.g. in questions of partner modelling, or multi-user effects.

There is also a growing interest in AI from the other sciences in AI. Walton (27) explicitly states, that there is a new interest of logicians in a logical theory of discourse because of the representational work done in AI. There is hope that this contact will influence the disadvantageous tradition of logics to eliminate everything which is not regular enough as some sort of pragmatic pollution.

Cognitive psychology, after decades in the declarative and microexperimental paradigm (at least in Europe), is trying again to sketch more general and broader cognitive models.

However, there are fields in which discourse analysis cannot rely on linguistics because of the missing explicitness concerning procedural aspects of language (see the Dresher/Hornstein controversy in Cognition 4,1976 ff). E.g. modern linguistics is just starting to discover language generation.

But we need even sketchy procedural models of understanding, of generation, of anaphora, or of spatial perception and description today.

And there is the same holistic reason why we cannot simply take the results of linguistics or psychology and program them: linguists are not used to constructing integral models. In their paper-and-pencil work there is no need for explicitly relating e.g. the view of page 20 to that of page 200. Implementation of discourse understanding processes, on the other hand, produces systems in which everything must fit together.

A third argument, however, hits linguistics as well as AI: We have no well-developed linguistics of natural language man-machine-communication. This means: no theories about language acquisition, generation, understanding, partner model, pragmatics, etc. of man-machine-communication.

Evidence from mock-up systems, simulated by persons, is methodologically vague and mostly too isolated from real application. Besides this it is restricted to short-term results. Nobody will play the mock turtle for months with hundreds of test persons.

Of course, linguists concerned with man-man-interaction have another interest in cognition. They do not implement their theories, or they do so for methodological reasons and not for the construction of working integrated software-systems. This has another result, namely that empirical work in linguistics is concentrated more on very general types of discourse (informal dialogues, party small-talk etc.) and not so much on dialogues in the fields of application in which practical AI needs natural dialogue examples.

Kittredge and Lehrberger (20) brought together linguists and AI people under the notion of "sublanguage". This volume could have referred, however, to all the research on "technical language" or "registers" done in Europe since the early Prague School. Meanwhile there are available a lot of detailed studies, some highly developed, though largely informal theories and a lot of statistic material about communication in non-social contexts and among experts (for a survey see v.Hahn (15)).

This research investigates what in AI is sometimes neglected: The semantic and syntactic restrictions in technical languages, the differences between written and spoken language or the effects of communication with non-individual addressees.

Wynn's PhD thesis (28) seems to be one of the few empirical studies for the American office setting.

Empirical work in this field is necessary for plausible performance of application oriented systems. McKeown et al. (22), although she did not invent the linguistic characteristics of their system but based it on transcripts of actual student advising sessions, admits, that "it would be desirable to have much larger set of plans, knowledge about their base rates and importance, and additional criteria for tracking their relevance and likelihood during the interaction".

In the long run we need such research for practical systems even in the starting phase of designing a system. We will be forced to start work with very clear functional specifications and will apply much more of the techniques of software engineering.

Let me close this paragraph with a more heuristic remark. Some remarkable progress in procedural modelling of human language abilities has been achieved by looking at the problems from the opposite side.

I will give some examples of this figure-ground heuristics:

Falzon et al. (9), investigating the conditions of "natural" technical communication, did not look at the understanding process of a hearer but at the techniques of communicative experts, how they guide the the partner in restricting his or her linguistic activities.

Wachtel (26) recommends looking at ellipsis as the unmarked linguistic form whereas explicit full sentences are to be motivated by a specific context.

Webber and Mays (25) as well as Goodman (12) started to do research on misunderstandings and misconceptions to get an idea of proper understanding; instead of the flow of continuous coherent interchanges Hayes-Roth and Hayes-Roth (17) Grosz and Sidner (14) scrutinized interruptions as "a salient feature of cognitive processing in general" (17).

Harris/Beg/Upfold regard semantic understanding not as a reconstruction process: "the hearer does not construct a message from components extracted from speech but rather narrows down and refines a message by successively rejecting an inappropriate information from a general message" (16).

By the way, this heuristics holds even for the style of publications: It is a good tradition esp. in American reports to discuss the limitations and the shortcomings of one's own approach, which is not often heeded in European papers.

The 4th Challenge: Integration of Finding Procedures, Representations, and Evaluation Processes

In this last paragraph I will follow another line of the holism argument: In contrast to linguistics, in AI every process must be defined on at least three levels.

- (1) how to find in the data those features addressed by the theory,
- (2) how to represent them
- (3) how to infer on them or to evaluate the representation

In the intuition of the speaker/hearer this is in fact one simple process. Meta-utterances of speakers never will refer to only one of these processes.

Too much work in discourse analysis lacks one of these three levels. Of course, specific work may concentrate on one aspect without elaborating the others. But the arguments for the approach must come from all three processes.

Some examples:

You can represent the process of running a car (a similar example was first introduced by Faught (10) as a sequence of choices, because one can observe all these actions and objects:

- foot: left / right
- hand: left / right
- movement: put on / release / move
- device: clutch / accelerator / gearshift / brake

But in real driving actions you will never find a moment, when a driver has to choose between, say, the brake and the clutch directly. There are patterned sequences representing the plans of "go faster" or "go slower" etc. in which the elementary actions occur on different places, but everything seems to be compiled in some way.

Theoretical work often starts with statements like "Let (x(y)z(a)) be the representation of of the sentence (7c)". It is nowhere explained by which detection procedures this representation can be obtained or whether there is even the slightest chance of defining an analysis algorithm which maps (7b) onto (x(y)z(a)). Is cognitively plausible reasoning possible on this structure?

Empirical work often starts with statements like "The speaker is here slightly influenced by the fact that ..." Does that mean to introduce some sort of predicate SLIGHTLY_INFLUENCED (x,y)? How can this specification be found in the linguistic data and how can you infer on that (Following Butterworths (5) 4th maxim "Remember that conversationalists talk".)

The tight connection of analysis, representation and evaluation is necessary, among others, because every explanation of the system must be based on some sort of self-inspection of the system. But a system cannot answer to a request for clarification: "I could find a discourse constituent unit but I was not able to construct a discourse unit out of it".

It is not reasonable to address features of data which cannot be represented in a tractable way and cannot be evaluated for plausible processes on higher levels. Or to invent representations for which you cannot find a mapping from the data.

What is the use of an inference mechanism for an natural language interface, if it cannot handle vague natural language quantifiers detected by the parser?

We criticize all these partial views to discourse understanding processes also for another reason:

We must show the plausibility of the detection procedures, the representation and the inferences also under the natural conditions of mass data, that means e.g. multiple views on a subject, or remembering and forgetting. Most of the proposals for dialogue structures never have occupied with the mass phenomena. What will happen, when all the heterogeneous details are represented, when you will have several thousand non-uniform inference rules?

Of course we ever will discuss thoroughly the very features of natural dialogues which we cannot handle today, and start with fragments. But to propose e.g. any arbitrary representation without connection forwards and backwards is only a tiny step towards the solution of the discourse problems. Our knowledge of discourse processes is at least so that we cannot any longer design isolated structural fragments of the analysis and generation process.

Let me summarize: Cognitively sound approaches to discourse processes must start once again to take seriously the user and his intuitions about man-machine-interaction. We must free our general concepts from the shortcomings of modularity, that means to accept the equal importance of discovery procedures, representations, and evaluation. The reliability of one of these processes can only be justified by arguments of both others. We should exploit the results of the background sciences linguistics, psychology

and social science as far as they support a pragmatic and procedural view of discourse. All this to set out a new pragmatic and holistic view of our natural, flexible, efficient and " whatsoever " way of communication.

Acknowledgements

I am grateful to Tom Wachtel for essential discussions and for revising the English version of this paper. An invitation to the 'maison des sciences de l'homme' at Paris gave me the time to write the paper. The preparation of the paper was supported by the ESPRIT project LOKI.

References:

- (1) Appelt, D.E., TELEGRAM. A Grammar Formalism for Language Planning. In: Proc. IJCAI 1983. 595-599.
- (2) Berwick, R.C., Computational Aspects of Discourse. In: (4). 27-87.
- (3) Brady, J.M., Foreword in (4).
- (4) Brady, J.M. / Berwick, R.C., Computational Models of Discourse. Cambridge (Mass.): MIT Press 1983.
- (5) Butterworth, B., Maxims for Studying Conversation. In: Semiotica 24 (1978) 3/4. 317-339.
- (6) Carroll, J.M. / Bever, Th.B., The Non-Uniqueness of Linguistic Intuitions. IBM Research Division Report RC 6938. Oct 1978.
- (7) Conclin, E.J. / McDonald, D., Saliency: The Key to the Selection Problem in Natural Language Generation. In: Proc. 20th ACL Meeting 1982. 129-135.
- (8) Eelman, J., An Indirect Approach to Types of Speech Acts. In: Proc. IJCAI 1983. 600-602.
- (9) Falzon, P. / Amalberti, R. / Carbonell, N., Dialogue Control Strategies in Oral Communication. INRIA Report 377. Centre de Rocquencourt Versailles March 1985.
- (10) Faught, W.S., Conversational Action Patterns in Dialogs. In: Waterman / Hayes-Roth (eds.), Pattern-Directed Inference Systems. N.Y 1978. 383-397.
- (11) Fodor, J.A., Modularity of Mind. An Essay on Faculty Psychology. Cambridge (Mass.): MIT Press 1983.
- (12) Goodman, B.A., Communication and Miscommunication. BBN Report 1984.

- (13) Grosz, B.J., Transportable Natural Language Interfaces: Problems and Techniques. In: Proc. 20th ACL Meeting 1982. 46-50.
- (14) Grosz, B.J. / Sidner, C.L., Discourse Structure and the Proper Treatment of Interruptions. In: Proc. IJCAI 1985. 832-839.
- (15) v.Hahn, W., Fachkommunikation. Berlin: de Gruyter 1983.
- (16) Harris, G. / Begg, J. / Upfold, D., On the Role of the Speaker's Expectations in Interpersonal Communication. In: Journal for Verbal Learning and Verbal Behaviour 19 (1980). 597-607.
- (17) Hayes-Roth, B. / Hayes-Roth, F., A Cognitive Model of Planning. In: Cognitive Science 3 (1979). 275-310.
- (18) Israel, D., Preface in (4).
- (19) Johnson-Laird, P.N., Mental Models of Meaning. In: Joshi / Webber / Sag (eds.), Elements of Discourse Understanding. Cambridge (Mass.): MIT Press 1981. 106-126.
- (20) Kittredge, R. / Lehrberger, J. (eds.), Sublanguage. Studies of Language in Restricted Semantic Domains. Berlin: de Gruyter 1982.
- (21) Langacker, R., An Introduction to Cognitive Grammar. In: Cognitive Science 10 (1986) 1-40.
- (22) McKeown, K. / Wish, M. / Matthews, K., Tailoring Explanations for the User. In: Proc. IJCAI 1985. 794-798.
- (23) Power, R. / dal Martello, M.F., Methods of Investigating Conversation. In: Semiotica 53 (1985) 1/3. 237-257.
- (24) Sidner, C., Focusing in the Comprehension of Definite Anaphora. In: (4). 267-330.
- (25) Webber, B. / Mays, E., Varieties of User Misconceptions: Detection and Correction. In: Proc. IJCAI 1983. 650-652.
- (26) Wachtel, Tom, Discourse Structure. LOKI-Report NLI 1.1 Research Unit for Information Sc. and AI, Univ Hamburg. April 1985.
- (27) Walton D. N., New Directions in the Logic of Dialogue. In: Synthese 63 (1985). 259-274.
- (28) Wynn, E.H., Office Conversation as an Information Medium. PhD Univ. of California Berkeley 1979. Ann Arbor: UMI 1985.