Towards an Understanding of

Coherence in

Multimodal Discourse

Som Bandyopadhyay

January 1990

Deutsches Forschungszentrum für Künstliche Intelligenz

The German Research Center for Artificial Intelligence (Deutsches Forschungszentrum für Künstliche Intelligenz, DFKI) with sites in Kaiserslautern und Saarbrücken is a non-profit organization which was founded in 1988 by the shareholder companies ADV/Orga, AEG, IBM, Insiders, Fraunhofer Gesellschaft, GMD, Krupp-Atlas, Mannesmann-Kienzle, Nixdorf, Philips and Siemens. Research projects conducted at the DFKI are funded by the German Ministry for Research and Technology, by the shareholder companies, or by other industrial contracts.

The DFKI conducts application-oriented basic research in the field of artificial intelligence and other related subfields of computer science. The overall goal is to construct *systems with technical knowledge and common sense* which - by using AI methods - implement a problem solution for a selected application area. Currently, there are the following research areas at the DFKI:

- ❏ Intelligent Engineering Systems
- ❏ Intelligent User Interfaces
- ❏ Intelligent Communication Networks
- ❏ Intelligent Cooperative Systems.

The DFKI strives at making its research results available to the scientific community. There exist many contacts to domestic and foreign research institutions, both in academy and industry. The DFKI hosts technology transfer workshops for shareholders and other interested groups in order to inform about the current state of research.

From its beginning, the DFKI has provided an attractive working environment for AI researchers from Germany and from all over the world. The goal is to have a staff of about 100 researchers at the end of the building-up phase.

Prof. Dr. Gerhard Barth
Director

Towards an Understanding of Coherence in Multimodal Discourse

Som Bandyopadhyay

DFKI-TM-90-01

TOWARDS AN UNDERSTANDING OF COHERENCE IN MULTIMODAL DISCOURSE

S. Bandyopadhyay
German Research Center for Artificial Intelligence
Stuhlsatzenhausweg 3
D 6600 Saarbrucken 11
Federal Republic of Germany

ABSTRACT

An understanding of coherence is attempted in a multimodal framework where the presentation of information is composed of both text and picture segments (or, audio-visuals in general). Coherence is characterised at three levels: coherence at the syntactic level which concerns the linking mechanism of the adjacent discourse segments at the surface level in order to make the presentation valid; coherence at the semantic level which concerns the linking of discourse segments through some semantic ties in order to generate a wellformed thematic organisation; and, coherence at the pragmatic level which concerns effective presentation through the linking of the discourse with the addressees´ preexisting conceptual framework by making it compatible with the addressees´ interpretive ability, and linking the discourse with the purpose and situation by selecting a proper discourse typology. A set of generalised coherence relations are defined and explained in the context of picture-sequence and multimodal presentation of information.

TABLE OF CONTENTS

1. INTRODUCTION

A text cannot be characterised in terms of mere collection of sentences. A text has an organisational framework and the sentences in a text bind themselves together to constitute a ´unit´ that itself has a meaningful interpretation. Text linguistics, therefore, argue that texts constitute an object to be modeled differently from sentences in isolation /Van Dijk 85/. This is due to the occurance of a lingustic phenomenon, generally termed as **text coherence** /Hobbs 83, Reichman 78/ which establishes textuality above the sentence level.

The same analysis is true for a sequence of pictures, for example in silent movies or cartoons, where the pictures bind themselves together in order that a definite meaningful interpretation on the whole is possible. A sequence of pictures in silent movie is, therefore, different from a collection of pictures in that the former are ´coherently´ sequenced.

When we talk about multimodal presentation of information (in terms of a combination of text and pictures, or, audio-visuals in general), discourse segments are no longer structurally uniform. Each discourse segment in this mode is either a text segment (or, audio segment) or a picture segment (visual segment) or a combination of both. Again, the notion of ´coherence´ is the binding force which links the multimodal discourse segments together to generate an overall meaningful presentation i.e. a presentation which is *valid* (adjacent discourse segments are properly connected at the surface level), *structured* (having wellformed thematic organisation) and *effective* (capable of meeting the intended goal of the discourse from the perspective of the receptor of the discorse).

In this study, a proper characterisation of coherence is attempted in a multimodal framework which can ensure validity, thematic wellformedness and effectiveness of a multimodal discourse.

2. CHARACTERISING COHERENCE

2.1 What is Coherence?- Some General Observations

In /Rescher 73/, the coherence of a propositional set is understood as requiring not simply the obvious minimum of consistency but also the feature of being connected in some special ways. He argues that the *special* connection required links consistency

pragmatically with our affective life. Also in /Samet 84/, coherence is depicted as connectedness of various sort that plays a part in building the receptors´ mental representation.

In /Margolis 84/, the coherence is defined in terms of a model of the rationality of human thought and action in order to satisfy the condition for the special connection as proposed by /Rescher 73/. According to him, coherence is a function of the rationality of human thought and action with self-consistency as its necessary and minimal condition. Incoherence is a relative discrepancy of some sort in the speech, thought, behaviour or work of human being in terms of their products or acts or deeds. But at the same time, incoherence cannot be a mere formal property of sentences or statements or analogous elements detatched from the human activity by which they are generated; it must be *linked* to the intension and purposive life of human being. The idea of coherence is, therefore, context dependent : what is coherent in one context may appear incoherent in some other context /Margolis 84/.

From the above discussion, some general observations can be made as follows:

- A minimum requirement for coherence is self-consistency.

- Coherence is connectedness (or, togetherness) which reflects the rationality of an agent who tends to maintain coherence to fulfil his rational purpose. There are two implications of this. First, as coherence can be seen as connectedness, it is more often a matter of degree than a clearcut presence or absence. Second, as connectedness in this context is a consequence of rationality, the connections are often to be explained in terms of intention and purposive life of human beings.

- Coherence is context-sensitive. For example, the following three sentences are coherent in the context of the question "Tell me what thoughts are crossing your mind":
"It is snowing in Tibet. The moon is round. Cats eats rats."/Ziff 84/.

2.2 Coherence in Discourse : A Three-Level Description

2.2.1 Coherence in Discourse : Related Research

Some significant work has been devoted to the study of coherence in text (dialog or monolog) /Enkvist 78, Grimes 75, Hobbs 79, Hobbs 83, Reichmann 78 and others/. But unfortunately, no significant work has been done in the area of coherence in picture

sequence or coherence in multimodal discourse composed of text and pictures (or, audio-visuals, in general).

/Hobbs 83/ describes three kinds of coherence : the *global coherence* of an utterance is its relation to the speaker´s overall plan (/Litman 86/ also has the same basis of analysis); the *local coherence* of an utterance or larger segment of discourse is its relation to adjacent segments in the discourse in terms of satisfying certain textual goals; and *thematic coherence* is a coherence pattern observable, for example, in different writings of the same author. Hobbs defined a set of coherence relations and stated that a text will strike one as coherent to a degree that varies inversely with the degree of ´difficulty´ in inferring certain coherence relations.

/Lindeberg 85/ is an exploratory study designed to find and describe coherence patterns which distinguishes good expository essays from poor one. He emphasises on degree of coherence rather than coherence/incoherence in its absolute sense and showed that it is possible to arrive at coherence patterns, based on sequences of rhetorical functions, that are frequent in good essays and less frequent in poor essays. He segmented each text into functional units and identified the functional role of each unit (which expresses the rhetorical purpose of the unit in the discourse, e.g., assert, specify, contrast, result, cause and so on). Some combinations of functional roles generate patterns which are less coherent than other combinations. For example, a combination ´asert-asert-asert´ (i.e. three consecutive functional units are making a major assertions each by introducing a new topic or an aspect of a topic not covered before) will appear to be less coherent than a combination ´assert-specify-cause´ or ´asert-cause-specify´.

Samet and Schank /Samet 84/ analyse coherence in terms of connectivity of various sort that plays a part in the hearer´s evolving mental representation of the discourse. They identified two types of connectivities : *internal connectivity* which is the connectivity within the discourse among discourse segments; and, *external connectivity* which concerns the integratability of this unified representation into the hearer´s pre-existing conceptual framework. A text is coherent for a reader if and only if the reader has constructed an adequate conceptualisation for it which involves building a conceptualisation (internal connectivity) and integrating it into one´s own world-view (external connectivity). They defined a set of hierarchical connections similar to coherent relations as defined in /Hobbs 83/.

Most of these works are mainly concentrating on coherence detection in text in terms of identifying certain coherence relations or coherence patterns in the process of

understanding a text. Although the notion of global coherence /Hobbs 83/ and external connectivity /Samet 84/ are introduced to capture the role of readers' interpretive ability in detecting coherence, they received less attention in the overall discussion.

Efforts have also been made to generate coherent text based on rhetorical relations /Moore 88, McKeown 85/. McKeown did not try to characterise coherence, rather, coherence is a consequence of her methodology in a restricted sense. An interesting study would be to explore the possibility of using rhetorical relations as described in RST /Mann 87/ in characterising coherence in multimodal discourse. However, the present study is restricted in characterising coherence which is primarily motivated by /Hobbs 83/.

Next section is an attempt to characterise coherence in a more general framework of multi-modal presentation of information involving text and pictures (or, audio-visuals, in general), where a discourse segment is either a text (or, audio) segment, or a picture (or, visual segment) or a combination of both. Coherence is a function both of features in the discourse and of the addressees' interpretive ability which depends on various factors like knowledge of the subject, inferencing ability, purpose and motivation /Grice 75/. We, therefore, attempt to characterise coherence (i) *at the discourse level* which concerns the validity and thematic wellformedness of the discourse, and, (ii) *at the user level* which concerns the effectiveness of the discourse in the specific context.

2.2.2 A Three-Level Description of Coherence in Discourse

Coherence can be characterised at three levels : Coherence at the syntactic level (or, **syntactic coherence**), coherence at the semantic level (or, **semantic coherence**) and, coherence at the pragmatic level (or, **pragmatic coherence**). The first two levels is *discourse level* and the third level is the *user level* characterisation of coherence.

Syntactic coherence in a discourse is a surface-level phenomenon that deals with the immediate connectivity among adjacent segments using some rules or conventions of connectivity. Maintaining syntactic coherence, on the one hand, ensures validity of the presentation at the syntactic level, and, on the other hand, helps the addressee to capture the coherence at the semantic level. Syntactic coherence is the necessary and minimal condition for maintaining coherence but it alone does not ensure the coherence of the discourse. Its role in a discourse is similar to the role of grammaticality in a sentence in the sense that grammatical correctness of a sentence is necessary for easy understanding of the sentence but it alone does not ensure the meaningfulness of the sentence as such.

Semantic coherence concerns the content and global structuring of a discourse. It ensures a wellformed thematic organisation of a discourse so that the discourse can be conceived as a unified whole. The discourse segments are coupled together through some semantic ties (coherence relation /Hobbs 83/ or conceptual connectivity /Samet 84/) which in turns generate a thematic structure of the discourse. Semantic coherence presupposes syntactic coherence; but, in certain cases, even if the syntactic clues from a semantically coherent discourse are removed, one can still make a sense out of it.

While coherence at the semantic level concerns what to communicate and how to organise it, it does not address the issue as to what extent to communicate in order to suit the addressees´ interpretive ability. In other words, semantic coherence deals with the relevance of a discourse segment with respect to the entire discourse, but it does not deal with the adequacy of presentation. As a result, a semantically coherent discourse may appear incoherent to an addressee whose interpretive ability does not match with that assumed in the discourse. Instead of calling it incoherent, we will call it *pragmatically* incoherent with respect to that addressee.

Pragmatic coherence, therefore, concerns the *effectiveness* of a discourse. A discourse is pragmatically coherent to an addressee or a group of addressees if it is compatible with the addressees´ interpretive ability. It presupposes semantic coherence.

The purpose of discourse is communication with a goal of being understood. The generator of the discourse generates some messages and the receptor applies an active inference process to understand the message and to link what is new in it to what he already knows. It is thus part of the generator´s job to provide the necessary linkage and to try to manipulate the receptor´s inference process to lead him to the proper interpretation /Hobbs 79/. **Pragmatic coherence** deals with the manipulation of the receptor´s inference process. A discourse which is pragmatically incoherent to a receptor can be sometimes made pragmatically coherent by adding more information in terms of clarification, elaboration, examples and so on which makes the presentation adequate for that receptor, but at the same time it has to be relevant to preserve the semantic coherence.

Another aspect of pragmatic coherence is the discourse typology. A discourse should be coherent with the purpose and situation in order to make the communication effective. An effective love letter should look different from an effective businessletter. A discourse should, therefore, be context dependent and be capable of *linking* the discourse to the purpose and situation for effective communication.

To summarise, *syntactic coherence* concerns the linking mechanism of the discourse segments at the surface level in order to make the presentation valid; *semantic coherence* concerns the linking of discourse segments through some semantic ties in order to generate a wellformed thematic organisation of the entire presentation; and, *pragmatic coherence* concerns effective presentation through the linking of the discourse with the addressees´ preexisting conceptual framework by making it compatible with the addressees´ interpretive ability, and linking of the discourse with the purpose and situation by selecting a proper discourse typology.

2.3 Coherence in Text

2.3.1 Syntactic Coherence

Syntactic coherence in text concerns the immediate connectivity among adjacent text segments through some surface markers or cohesive devices or overt linguistic mechanism such as anaphora and cataphora and ellipsis, conjunction and the like which serve to link text segments to each other. This surface-level, language dependent linking mechanism is often termed as *text cohesion* /Halliday 75/. Hobbs pointed out that cohesion does not ensure text coherence /Hobbs 79/. Lindeberg also showed that a text may have a fair amount of cohesive tightness but may not be coherent /Lindeberg 85/. Still, if a text is to be wellformed, it must have semantic coherence as well as sufficient signals of surface cohesion to enable the reader to capture the coherence. /McKeown 85/ indicated that certain choices at the surface level are critical in order to produce a coherent text. She pointed out that the generator must be able to make reasoned decisions about when to use pronominal reference and about the syntactic construction that should be used. An example of lexical choice (bought vs. sold) is given below /McKeown 85/ where (b) is a text sequence *incoherent* at the *syntactic level:*

(a) Jane was in a hurry to finish her shopping. It was a chore she particularly despised. First, Jane *bought* $3.00 worth of bobby socks from Michael.

(b) Jane was in a hurry to finish her shopping. It was a chore she particularly despised. First, Michael *sold* $3.00 worth of bobby socks to Jane.

2.3.2 Semantic Coherence

Text coherence at the semantic level concerns the global structuring of text in terms of wellformed thematic organisation. Instances of global text structuring through text coherence phenomena are given by regular pattern of thematic progression in a text /Danes

74/ or by various additional functional coherence relation such as elaboration, parallel, contrast, occasion, compatibility and the like /Hobbs 79, Hobbs 83, Reichman 78/.

In /Danes 74/, text coherence pattern is described which either follow the principle of constant theme, linear thematization of rhemes, or derived themes. According to /Werlich 76/, coherence is created whenever the generator uses some linguistic unit as a text base unit with a theme and then expands this in linear progression in conventionally ordered and completed sequences of linguistic units. /Hahn 86/ is an attempt towards creating a text graph using this notion of text coherence pattern in expository text.

Coherence relations describe a set of semantic ties which in turn defines a coherent structure of a text. Although /Samet 85/ has shown that these coherence relations alone cannot ensure semantic coherence of a text (especially, the notion of causal connectivity plays an important role beyond coherence relation in a narrative text), coherence relations are an important first step towards describing semantic ties among text segments. /Tucker 86/ is also based on a set of coherence relations in analysing coherence in expository text.

2.3.3 Pragmatic Coherence

The notion of pragmatic coherence in text takes care of the readers´ interpretive ability and adequecy of presentation in order to make a text effective. In other words, a text would be pragmatically coherent if it is compatible with the readers´ interpretive ability and is adequate to satisfy the reader.

As an example, suppose that a computer salesman describes a product of his company as : " The PC2000 is equipped with a 8086 cpu as opposed to the 8085 of the previous model. The standard amount of dynamic RAM is 256 Kbytes. One of the two RS-232C ports also serves as a higher speed RS-422 port...". This description is coherent and effective to a computer engineer and the like having sufficient background in computer hardware. But the same description will appear incoherent to a person outside the computer engineering profession. A more effective description to him would be : " The personal computer PC2000 is more powerful than its previous model. It has standard amount of memory and appropriate mechanism to link it with other computer for communication at a higher speed...". Again, this description is not effective enough to satisfy a computer engineer because it is *inadequate*.

Another aspect of pragmatic coherence concerns the pragmatic appropriateness of a text /Hovy 85/ in terms of its suitability with the characteristic of the hearer, the conversational

setting and their interpersonal goal. Selection of proper text typology or stylistics and use of proper pragmatic particle /Enkvist 85/ can make a communication far more effective. On the other hand, improper use of those can render the communication inappropriate and incoherent at the pragmatic level.

2.4 Coherence in Picture-Sequence

2.4.1 Syntactic Coherence

The pictures in a picture-sequence are not autonomous; their task is to contribute to the flow of information transmitted by the picture-sequence, to link up with what went before and with what comes after. Each picture represents a set of objects or event and contains a single focus of attention. The syntactic coherence in picture-sequence concerns the immediate connectivity of adjacent pictures. The conventions of this connectivity at the surface level is termed as *continuity*. Three types of continuities are identified:

i) Continuity in Perspective

ii) Continuity of Action

iii) Continuity in Focus of Attention

To illustrate this, we will take examples from the instruction manual for an *espresso coffee machine* (adopted from Philips Espresso Machine HD 5649).

i) Continuity in perspective: Examples of continuity in perspectives are **Spatial Continuity, Continuity in view-point, Continuity in Color**, etc.

Spatial Continuity is said to be maintained when there is no *sudden* change in the relative position *or* relative size of the objects (or, parts of objects) between two consecutive pictures in a picture-sequence having *same focus of attention*. For example, the pictures A and B in figure 1 are coherent illustrating the process of filling water in the water-container of a coffee machine. A *spatial discontinuity* is observed if one moves from picture B to picture C in the sequence, where the position of the hole is shifted from left (in B) to right (in C). A spatial discontinuity due to sudden change in size of the switch is also observable from B to C.

Continuity in View-point is said to be maintaied when there is no change in view-point between two consecutive pictures in a picture-sequence having *same focus of attention*. For example, a *continuity* in view-point is observed in fig. 1. If the same machine is

viewed from a different angle as compared to that in A in the sequence, a discontinuity in view-point is observable.
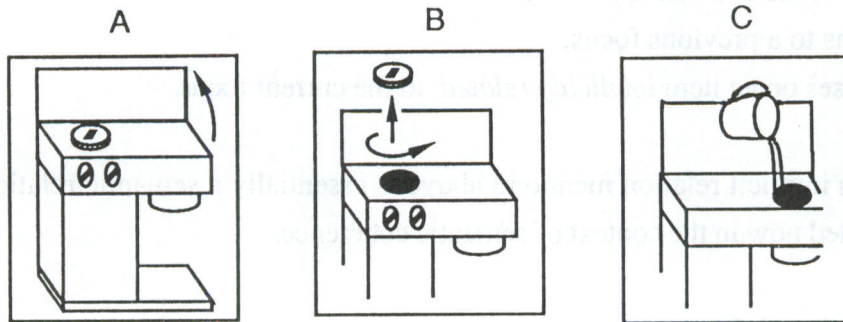

Fig.1. Filling up the water-container of a coffee machine

Continuity in Color is said to be maintained when there is no *sudden* change in the relative colors of the objects (or, parts of objects) between two consecutive pictures in a picture-sequence having *same focus of attention.*

ii) Continuity of Action , also called *temporal continuity,* is said to be maintained when temporally sequenced pictures in a picture-sequence are continuous in the time-domain. In other words, when a picture-sequence depicts certain action or sequence of actions, a picture in the sequence needs to maintain continuity of action with respect to its previous picture in the sequence, unless there is a total shift in the focus of attention. For example, let us refer to fig.2 where pictures A, B and C illustrate the process of filling up the water-container. A temporal *discontinuity* is observed, since the cover is shown in its position in C even after its removal in B.
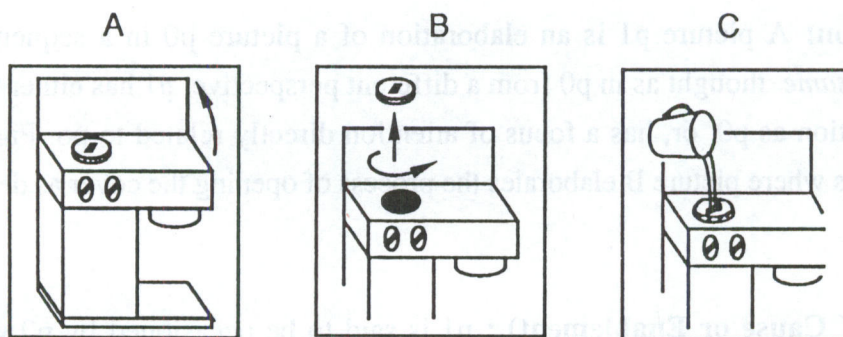

Fig.2 The process for filling up the water-container: A temporal discontinuity

iii) Continuity in Focus of Attention : /Sidner 79/ showed how immediate focus can shift or be maintained in a text. Same constraints can be used to understand the continuity in focus of attention in a picture-sequence. In a picture-sequence, a picture either a) maintains the current focus,

    or, b) shifts focus to an item just introduced,

    or, c) returns to a previous focus,

    or, d) focusses on an item *implicitly related* to the current focus.

Of course, this implicit relation mentiond above is essentially a semantic relation which will be illustrated now in the context of semantic coherence.

2.4.2 Semantic Coherence

The discourse structure of a picture-sequence can be described by defining certain coherence relations. These coherence relations, derived from /Hobbs 79, Hobbs 83/, are the semantic ties that can exist between two pictures in a sequence, not necessarily adjacent. These relations are :

a) Elaboration

b) Occasion (Cause or Enablement)

c) Expansion

    c.1) Specification

    c.2) Generalisation

    c.3) Parallel

    c.4) Contrast

d) Temporal Relation

    d.1) Parallel

    d.2) Sequential

a) Elaboration: A picture p1 is an elaboration of a picture p0 in a sequence if p1 expresses the *same* thought as in p0 from a different perspective. p1 has either the same focus of attention as p0, or, has a focus of attention directly related to po. Fig. 3 is an example of this where picture B elaborates the process of opening the cover as depicted in picture A.

b) Occasion (Cause or Enablement) : p1 is said to be occasioned by p0 when the event depicted in p0 produces a state that is consumed /Balzer 77/ by the event depicted in p1. When the first event causes the second event to occur, they are causally related; when

the first event sets up the situation for the second, p1 is said to be enabled by p0. An example is given in figure 4, where changing the switch-position causes the steam to come out of the nozzle (i.e. A causes B in fig.4), and, this enables one to hold the coffee-cup at the exhaust (i.e. C is enabled by B in fig.4).
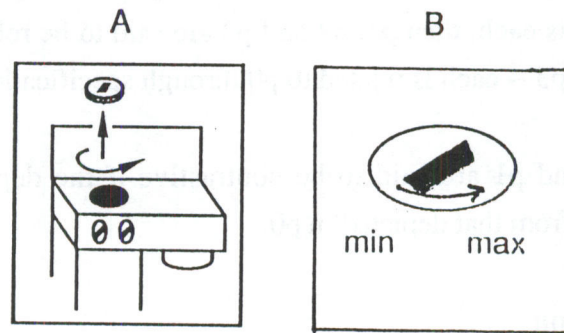
Fig. 3 Process of Removing the Cover in a Coffee Machine
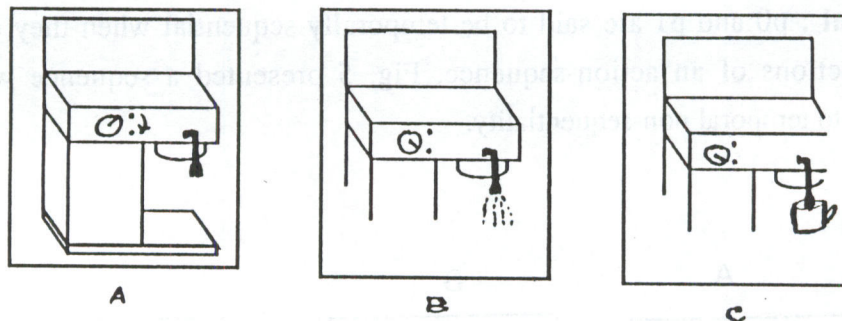
Fig.4 Preparing Espresso Coffee.

c) Expansion

c.1) Specification : When p1 specifies something illustrated in p0 in more detail, it is said to be a specification relation. Use of close-up or zoom is a method for this in a picture-sequence. For example, if p0 shows an object and p1 zooms in a part of the same object to show it in greater detail, they are said to be related through specification relation.

c.2) Generalisation : When p0 indicates some entity or a set of entities and p1 indicates the class to which they belong, p1 is a generalisation of p0. It is sometimes a reverse of specification relation.

c.3) Parallel : The parallel relation involve moving from a predication about some set of entities illustrated in p0 to the *same* predication about a similar set of entities illustrated in p1. For example, if p0 shows a machine consisting of three parts and p1, p2 and p3 show one of those parts each, then p1, p2 and p3 are said to be related parallelly. At the same time, p1, p2 and p3 -- each is related to p0 through specification relation.

c.4) Contrast : p0 and p1 are said to be contrastive if the depiction in p1 is either dissimilar or opposite from that depicted in p0.

d) Temporal Relation

d.1) Parallel : p0 and p1 is said to be temporally parallel when the event depicted in p0 and that depicted in p1 occurs simultaneously. Two simultaneous events are shown in a picture-sequence through a mechanism known as inter-cut where two simultaneous events are broken into sub-events and presented in an alternate sequence of pictures.

d.2) Sequential : p0 and p1 are said to be temporally sequential when they depict a sequence of actions of an action-sequence. Fig. 5 presented a sequence which is incoherent due to temporal non-sequentiality.
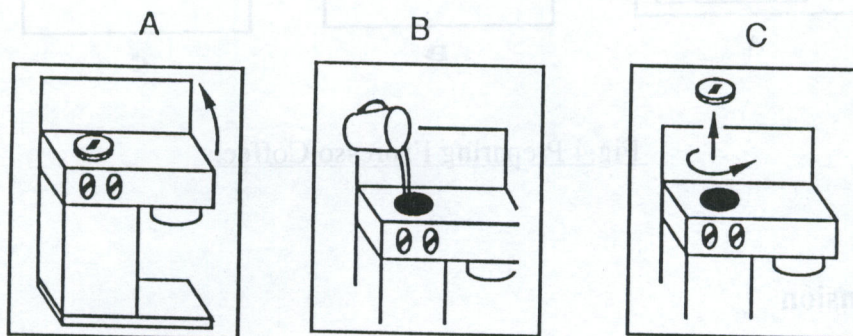

Fig.5 Filling up the water container : an incoherent sequence

2.4.3 Pragmatic Coherence

A massage can be conveyed through a coherent sequence of pictures in a multiple number of ways. But all of them may not be pragmatically coherent with respect to the intended addressees. The number of pictures and the content of each picture in a picture-sequence are the determining factors in generating a pragmatically coherent picture-sequence.

Moreover, a picture can be taken from different angle (e.g. high angle, low angle, etc.), from different view-point (e.g. front view, rear view, side view and so on) and from different distance (e.g. close-up, medium or long distance). The effectiveness often depends on proper selection of angle, view-point and distance while presenting a picture in a picture-sequence.

The purpose and situation also determine the style of presentation of a picture-sequence. For example, building up a fast,suspense sequence in a movie is different from building up a slow, romantic sequence. The first one usually comprises of more number of shots each with a shorter duration whereas the second one usually uses less number of shots each with a longer duration.

2.5. Coherence in Multimodal Discourse

In multimodal discourse, two adjacent discourse segments can both be text segments or picture segments, or, one of them is text segment and another is picture segment. Here, we will concentrate on the coherence pattern between a text segment and a picture segment, since the coherence pattern within adjacent text segments and within adjacent picture segments have already been discussed.

2.5.1 Syntactic Coherence

Linking mechanism between a text segment and a picture segment at the surface level is rather a loose concept, as any text segment can be coupled with any picture at the syntactic level. But one important consideration in the presentation of multimodal document is the positioning of picture with respect to the text segment referring that picture. If the picture is too far away from the relevant text segment or comes after some other pictures, it will lead to a surface-level incoherence.

The notion of syntactic coherence is more important in a movie where this linking mechanism between audio segment with corresponding visual segment is termed as *synchronisation*. Synchronisation of voice with picture (with lip and movement) is one example of maintaining syntactic coherence in a movie.

2.5.2 Semantic Coherence

A set of coherence relations similar to that described in section 2.4.2 with an additional relation termed as *equivalence* can be described to illustrate the semantic tie between a text segment and a picture segment or vice versa. The definition of those relations are already given in sec. 2.4.2. Here, *equivalence* is defined and some examples are presented.

Equivalence : Truly speaking, a picture and a text segment can never be equivalent in the formal sense of the term. However, a picture is said to be equivalent with a text segment if they convey conceptually identical thought from same perspective with reference to a common focus of attention. The simplest example is the relation between a picture and its caption (usually).

Fig. 6 gives an example of the coherence relations between a picture and text segments.

"Filling up the water-container."   **equivalence**

"Fill the container with *cold* water"   **elaboration**

"A *similar* mechanism is followed in *all* the coffee-machine in this series."   **generalisation**

"The previous model has a different mechanism ; it uses a separate container."   **contrast**

"The previous model also has the *same* mechanism for filling the container"   **parallel**

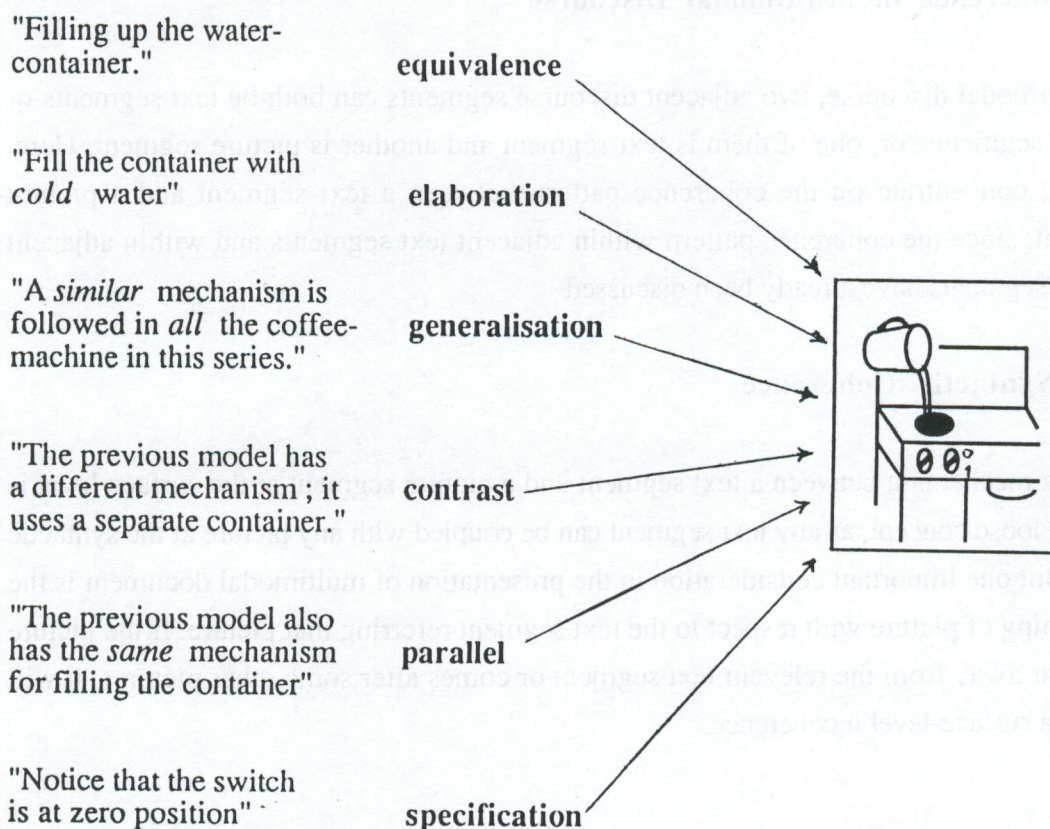"Notice that the switch is at zero position"   **specification**

Fig.6 Illustrating Coherence Relations between Text and Picture segment

2.5.3 Pragmatic Coherence

Pragmatic coherence in multimodal discourse concerns the factors illustrated in pragmatic coherence in text (sec.2.3.3) and pragmatic coherence in picture sequence (se.2.4.3). Apart from those, it also concerns the *selection of modality* of presentation, that is, the decisions such as when to use picture and when to use text and to what extent. At the same time, the selection of the most suitable graphical presentation for a given object among several possibilities needs to be done in a specific context to make the presentation more effective.

3.CONCLUSION

An attempt is made to characterise coherence in multimodal framework . Coherence is characterised at three levels: syntactic, semantic and pragmatic level. Maintaining coherence at these three levels ensures validity, thematic wellformedness and effectiveness of a multimodal discourse. A set of coherence relations are defined and explained in *multimodal* framework. However, this is an initial attempt and no claim is made regarding the completeness of this set of relations.

REFERENCES

/Balzer 77/
Balzer, R., Goldman, N. & Wile, D. (1977): Informality in program specification. Proceedings, IJCAI 1977, Cambridge, Mass.

/Danes 74/
Danes, F. (1974): Functional sentence perspective and the organisation of the text. In: Danes (Ed.). Papers on functional sentence perspective. Academia.

/Enkvist 78/
Enkvist., N.E. (1978): Coherence, pseudo-coherence and non-coherence. In: Östman (Ed.). Cohesion and semantics. Meddelanden fran Stiftelsens för Abo Akademi forskningsinstitut Nr.41.

/Enkvist 85/
Enkvist, N.E. (1985): Coherence, composition and text linguistics. In: Enkvist (Ed.). Coherence and composition. Abo Akademi Foundation. Sweden.

/Grice 75/

Grice, H.P. (1975): Logic and conversation. In: Cole & Morgan (Eds.). Syntax and Semantics, vol.3. Academic press. New York.

/Grimes 75/

Grimes, J. (1975): The thread of discourse. The Hague. Mouton.

/Hahn 86/

Hahn, U. & Reimer, U.(1986). Topic essentials. Proceedings. COLING 1986.

/Halliday 75/

Halliday, M. & Hasan, R.(1976): Cohesion in English. Longman.

/Hobbs 79/

Hobbs, J.(1979): Coherence and coreference. Cognitive science, 3 (1).

/Hobbs 83/

Hobbs, J.(1983): Why is discourse coherent ? In: Neubauer (Ed.). Coherence in natural language texts. Buske.

/Hovy 87/

Hovy, E.H.(1987): Some pragmatic decision criteria in generation. In: G. Kempen (Ed.). Natural language generation: New results in artificial intelligence, psychology and linguistics. Nijhoff :Dordrecht. Boston.

/Lindeberg 85/

Lindeberg, A.C.(1985): Cohesion, coherence pattern and EFL essay evaluation. In: Enkvist (Ed.). Coherence and composition. Abo Akademi, Sweden.

/Litman 86/

Litman, D.J.(1986): Linguistic coherence: A plan based alternative. Tech. report no. 311305-0299. AT & T Bell laboratories, MH 3C-408A.

/Mann 87/

Mann, W.C. & Thompson, S.A.(1987): Rhetorical Structure Theory: A theory of text organisation. In: L.Polany (Ed.). Discourse Structure. Norwood, N.J.: Ablex.

/Margolis 84/

Margolis, J.(1984): The locus of coherence. Linguistics & Philosophy, 7(1).

/McKeown 85/

McKeown, K.R.(1985): Discourse strategies for generating natural language text. Artificial Intelligence, 27.

/Moore 88/

Moore, J.D. & Paris, C.L.(1988): Constructing coherent text using rhetorical relations. Proceedings. Tenth annual conference of the Cognitive Science Society.

/Reichman 78/

Reichman, R.(1978): Conversational coherency. Cognitive Science, 2(4).

/Rescher 73/

Rescher, N.(1973): The coherence theory of truth. Clarendon. Oxford.

/Samet 84/

Samet, J. & Schank, R.(1984): Coherence and connectivity. Linguistics & Philosophy, 7(1).

/Sidner 79/

Sidner, C.L.(1979): Towards a computational theory of definite anaphora comprehension in english discourse. Ph.D. Dissertation. MIT, Cambridge. Mass.

/Tucker 86/

Tucker, A.B., Nirenburg, S. & Raskin, V.(1986): Discourse and cohesion in expository text. Proceedings. COLING 1986.

/Van Dijk 85/
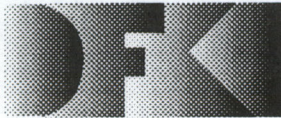
Van Dijk, T.A.(Ed.)(1985): Handbook of discourse analysis. Academic Press. U.K.

/Werlich 76/

Werlich, E.(1976): A text grammer of english. Quelle & Meyer. Heidelberg.

/Ziff 84/

Ziff, P.(1984): Coherence. Linguistics & Philosophy, 7(1).

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

DFKI Publikationen

Die folgenden DFKI Veröffentlichungen oder die aktuelle Liste von erhältlichen Publikationen können bezogen werden von der oben angegebenen Adresse.

DFKI Publications

The following DFKI publications or the list of currently available publications can be ordered from the above address.

DFKI Research Reports

RR-90-01
Franz Baader
Terminological Cycles in KL-ONE-based Knowledge Representation Languages
33 pages

Abstract: Cyclic definitions are often prohibited in terminological knowledge representation languages, because, from a theoretical point of view, their semantics is not clear and, from a practical point of view, existing inference algorithms may go astray in the presence of cycles. In this paper we consider terminological cycles in a very small KL-ONE-based language. For this language, the effect of the three types of semantics introduced by Nebel (1987, 1989, 1989a) can be completely described with the help of finite automata. These descriptions provide a rather intuitive understanding of terminologies with cyclic definitions and give insight into the essential features of the respective semantics. In addition, one obtains algorithms and complexity results for subsumption determination. The results of this paper may help to decide what kind of semantics is most appropriate for cyclic definitions, not only for this small language, but also for extended languages. As it stands, the greatest fixed-point semantics comes off best. The characterization of this semantics is easy and has an obvious intuitive interpretation. Furthermore, important constructs – such as value-restriction with respect to the transitive or reflexive-transitive closure of a role – can easily be expressed.

RR-90-02
Hans-Jürgen Bürckert
A Resolution Principle for Clauses with Constraints
25 pages

Abstract: We introduce a general scheme for handling clauses whose variables are constrained by an underlying constraint theory. In general, constraints can be seen as quantifier restrictions as they filter out the values that can be assigned to the variables of a clause (or an arbitrary formulae with restricted universal or existential quantifier) in any of the models of the constraint theory. We present a resolution principle for clauses with constraints, where unification is replaced by testing constraints for satisfiability over the constraint theory. We show that this constrained resolution is sound and complete in that a set of clauses with constraints is unsatisfiable over the constraint theory iff we can deduce a constrained empty clause for each model of the constraint theory, such that the empty clauses constraint is satisfiable in that model. We show also that we cannot require a better result in general, but we discuss certain tractable cases, where we need at most finitely many such empty clauses or even better only one of them as it is known in classical resolution, sorted resolution or resolution with theory unification.

RR-90-03

Andreas Dengel & Nelson M. Mattos

Integration of Document Representation, Processing and Management

18 pages

Abstract: This paper describes a way for document representation and proposes an approach towards an integrated document processing and management system. The approach has the intention to capture essentially freely structured documents, like those typically used in the office domain. The document analysis system ANASTASIL is capable to reveal the structure of complex paper documents, as well as logical objects within it, like receiver, footnote, date. Moreover, it facilitates the handling of the containing information. Analyzed documents are stored by the management system KRISYS that is connected to several different subsequent services. The described integrated system can be considered as an ideal extension of the human clerk, making his tasks in information processing easier. The symbolic representation of the analysis results allow an easy transformation in a given international standard, e.g., ODA/ODIF or SGML, and to interchange it via global network.

RR-90-04

Bernhard Hollunder & Werner Nutt

Subsumption Algorithms for Concept Languages

34 pages

Abstract: We investigate the subsumption problem in logic-based knowledge representation languages of the KL-ONE family and give decision procedures. All our languages contain as a kernel the logical connectives conjunction, disjunction, and negation for concepts, as well as role quantification. The algorithms are rule-based and can be understood as variants of tableaux calculus with a special control strategy. In the first part of the paper, we add number restrictions and conjunction of roles to the kernel language. We show that subsumption in this language is decidable, and we investigate sublanguages for which the problem of deciding subsumption is PSPACE-complete. In the second part, we amalgamate the kernel language with feature descriptions as used in computational linguistics. We show that feature descriptions do not increase the complexity of the subsumption problem.

RR-90-05

Franz Baader

A Formal Definition for the Expressive Power of Knowledge Representation Languages

22 pages

Abstract: The notions "expressive power" or "expressiveness" of knowledge representation languages (KR-languages) can be found in most papers on knowledge representation; but these terms are usually just used in an intuitive sense. The papers contain only informal descriptions of what is meant by expressiveness. There are several reasons which speak in favour of a formal definition of expressiveness: For example, if we want to show that certain expressions in one language *cannot* be expressed in another language, we need a strict formalism which can be used in mathematical proofs. Though we shall only consider KL-ONE-based KR-language in our motivation and in the examples, the definition of expressive power which will be given in this paper can be used for all KR-languages with model-theoretic semantics. This definition will shed a new light on the tradeoff between expressiveness of a representation language and its computational tractability. There are KR-languages with identical expressive power, but different complexity results for reasoning. Sometimes, the tradeoff lies between convenience and computational tractability. The paper contains several examples which demonstrate how the definition of expressive power can be used in positive proofs – that is, proofs where it is shown that one language can be expressed by another language – as well as for negative proofs – which show that a given language cannot be expressed by the other language.

DFKI Technical Memos

Susan Holbach-Weber

Connectionist Models and Figurative Speech

27 pages

Abstract: This paper contains an introduction to connectionist models. Then we focus on the question of how novel figurative usages of descriptive adjectives may be interpreted in a structured connectionist model of conceptual combination. The suggestion is that inferences drawn from an adjective's use in familiar contexts form the basis for all possible interpretations of the adjective in a novel context. The more plausible of the possibilities, it is speculated, are reinforced by some form of one-shot learning, rendering the interpretative process obsolete after only one (memorable) encounter with a novel figure of speech.

Som Bandyopadhyay

Towards an Understanding of Coherence in Multimodal Discourse

18 pages

Abstract: An understanding of coherence is attempted in a multimodal framework where the presentation of information is composed of both text and picture segments (or, audio-visuals in general). Coherence is characterised at three levels: coherence at the syntactic level which concerns the linking mechanism of the adjacent discourse segments at the surface level in order to make the presentation valid; coherence at the semantic level which concerns the linking of discourse segments through some semantic ties in order to generate a wellformed thematic organisation; and, coherence at the pragmatic level which concerns effective presentation through the linking of the discourse with the addressees´ preexisting conceptual framework by making it compatible with the addressees´ interpretive ability, and linking the discourse with the purpose and situation by selecting a proper discourse typology. A set of generalised coherence relations are defined and explained in the context of picture-sequence and multimodal presentation of information.