ESSLLI workshop on

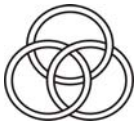
Coherence in Generation and Dialogue

Málaga, Spain

August 7th - 11th 2006

Workshop Proceedings

Rodger Kibble, Paul Piwek and Ielka van der Sluis (Eds.)



Part of ESSLLI 2006, the 18th European Summer School in Logic, Language and Information held from 31 July to 11 August 2006 at the University of Málaga.

Endorsed by SIGDIAL, the Special Interest Group on Discourse and Dialogue of the Association for Computational Linguistics.

Preface

We are pleased to present this volume which brings together the papers presented at the ESSLLI workshop on Coherence in Generation and Dialogue, August 7 – 11, 2006 in Málaga, Spain.

The aim of the workshop is to compare and contrast different models of coherence in natural language generation and dialogue that guide decisions such as: *What is an appropriate response at a given point in a dialogue?*, *What is the optimal ordering of propositions in a discourse?,* and *How should predicates, referring expressions and rhetorical relations be realised (verbally and/or non-verbally) so that the resulting utterance is fluent and can be interpreted naturally?* The papers that are collected in these proceedings present an overview of current approaches that address these and other questions.

The first set of papers on the programme (see page v) addresses coherence from the perspective of dialogue. Invited speaker Robbert-Jan Beun shows that it is possible to generate coherent dialogue sequences without the use of complex planning processes or speech act grammars. His framework consists of rules for updating the cognitive states of dialogue participants as a result of speech acts, and other rules for the generation of speech acts in a given cognitive state. Paul Piwek provides an account of the coherence of cooperative information-oriented dialogues in terms of the meaning of the expressions that are used and dialogue participants' discursive dispositions. His notion of meaning is fleshed out in terms of a calculus of Natural Deduction which is extended with rules for information transfer and observation. The paper by Rodger Kibble presents an approach to the generation of persuasive monologue based on a process of "inner dialogue" in which an author tries to anticipate objections and points of clarification that the addressee might raise. It is argued that certain RST relations embody strategies for pre-empting potential objections and requests for clarification. The paper by Andy Lücking, Hannes Rieser and Marc Staudacher builds on Nicholas Asher's Segmented Discourse Representation Theory (SDRT). Their presentation is preceded by an invited talk by Nicholas Asher. Lücking et al. propose a model for representing dialogue, including sub-sentential utterances and gestures. An extension of SDRT is described which makes it possible to treat meanings from different verbal and non-verbal sources as coherent wholes.

The following four papers address coherence in information ordering, content selection and modality use. Whereas the preceding papers look primarily at coherence from the perspective of dialogue and generation, now the emphasis is shifted to generation and monologue. In his paper on information ordering for automatic text production, Nikiforos Karamanis investigates whether the use of a centering-based metrics of entity coherence can be improved by the use of local rhetorical coherence. He shows that the simplest model performs best and that the use of local rhetorical coherence actually decreases performance. Clara Mancini and Donia Scott explore the construction of coherence in hypertext with the use of graphics and animation. In their cognitive approach, coherence is a characteristic of the mental representation that a reader constructs during text interpretation. The paper by Katja Filippova and Michael Strube presents a method for improving local coherence in German with implications for both automatically and human generated texts. They demonstrate that local

coherence crucially depends on which constituent occupies the initial position in a sentence. Finally, Albert Gatt and Kees van Deemter present and evaluate an algorithm that generates conceptually coherent plural descriptions. The algorithm is based on a notion of local coherence obtained from an empirical investigation of conceptual coherence in reference.

We believe that this volume brings together a collection of high quality papers − as witnessed by the very positive reviews we received for all the accepted papers − which illustrate the variety of approaches to and relevance of coherence in generation and dialogue. We are looking forward to a stimulating and fruitful workshop!

Last but not least we would like to thank everyone who helped to make the workshop happen. First of all, thank you to this year's ESSLLI programme and organizing committees for their smooth cooperation and assistance. Thanks are also due to SIGDIAL for their endorsement and support. Furthermore, we thank our reviewers Rieks op den Akker, Harry Bunt, Raquel Fernandez, Jonathan Ginzburg, Dirk Heylen, Erik Krabbe, Emiel Krahmer, Peter Kühnlein, Jill Nickerson, Massimo Poesio, Richard Power, Matthew Purver, Hannes Rieser, Donia Scott, Wilbert Spooren, Matthew Stone, Mariët Theune and Sandra Williams for their time and expert judgement. Finally, special thanks to our invited speakers Robbert-Jan Beun and Nicholas Asher for agreeing to participate in the workshop.

Rodger Kibble, Paul Piwek and Ielka van der Sluis (co-chairs)

Programme Committee

Rieks op den Akker	Twente University, The Netherlands
Harry Bunt	Tilburg University, The Netherlands
Raquel Fernandez	University of Potsdam, Germany
Jonathan Ginzburg	King's College London, United Kingdom
Dirk Heylen	Twente University, The Netherlands
Rodger Kibble	Goldsmiths College, United Kingdom (*co-chair*)
Erik Krabbe	University of Groningen, The Netherlands
Emiel Krahmer	Tilburg University, The Netherlands
Peter Kühnlein	Bielefeld University, Germany
Jill Nickerson	Ab Initio Software Corporation, United States
Paul Piwek	The Open University, United Kingdom (*co-chair*)
Massimo Poesio	University of Essex, United Kingdom
Richard Power	The Open University, United Kingdom
Matthew Purver	CSLI, Stanford University, United States
Hannes Rieser	Bielefeld University, Germany
Donia Scott	The Open University, United Kingdom
Ielka van der Sluis	University of Aberdeen, United Kingdom (*co-chair*)
Wilbert Spooren	Vrije Universiteit Amsterdam, The Netherlands
Matthew Stone	Rutgers University, United States
Mariët Theune	Twente University, The Netherlands
Sandra Williams	The Open University, United Kingdom

Table of Contents

Papers

Workshop Programme

MONDAY, AUGUST 7

17:00 *Rodger Kibble, Paul Piwek and Ielka van der Sluis*: Introduction

17:45 **Invited Speaker**: *Robbert-Jan Beun*: "A Simple Dialogue Game for the Generation of Coherent Speech Act Sequences"

TUESDAY, AUGUST 8

17:00 *Paul Piwek*: "Meaning and Dialogue Coherence: A Proof-theoretic Investigation"

17:45 *Rodger Kibble*: "Generating Coherence Relations via Internal Argumentation"

WEDNESDAY, AUGUST 9

17:00 **Invited Speaker**: *Nicholas Asher*: Title TBA

17:45 *Andy Lücking, Hannes Rieser & Marc Staudacher*: "SDRT and Multimodal Situated Communication"

THURSDAY, AUGUST 10

17:00 *Nikiforos Karamanis*: "Entity versus Rhetorical Coherence for Information Ordering: Initial Experimentation"

17:45 *Clara Mancini & Donia Scott*: "Hyper-Document Structure: Representing Cognitive Coherence in Non-linear Documents"

FRIDAY, AUGUST 11

17:00 *Katja Filippova & Michael Strube*: "Improving Text Fluency by Reordering of Constituents"

17:45 *Albert Gatt & Kees van Deemter*: "Conceptual Coherence in the Generation of Referring Expressions"

Papers

A Simple Dialogue Game for the Generation of Coherent Speech Act Sequences

Robbert-Jan Beun
Universiteit Utrecht /Padualaan 14, de Uithof
3584CH Utrecht, The Netherlands
rj@cs.uu.nl

Abstract

A dialogue game is presented that enables us to generate coherent elementary conversational sequences at the speech act level. Central to the approach is that the cognitive states of players change as a result of the interpretation of speech acts and that these changes provoke the production of a subsequent speech act. It will be shown that the structure and the coherence of conversational units do not necessarily have to be the product of a complex planning process or a speech act grammar. Although simple in its basic form, the framework enables us to produce abstract conversations with some properties that agree strikingly with dialogue properties found in Conversation Analysis.

1 Introduction

The central goal of this paper is to present a computational framework that enables us to generate coherent elementary conversational sequences at the speech act level. For that, I will embrace the notion of a *dialogue game* in which two players produce speech acts or 'moves' to transfer relevant information with respect to their goals. Central to the approach is that the cognitive states of the players change as a result of the interpretation of the speech acts (see e.g. Bunt, 1989) and that these changes provoke the production of a subsequent move. The game works roughly like this: A speech act is generated on the basis of preconditions formed the cognitive state of the sender, but changes the cognitive states of both sender and addressee after it has been manifested. In the next turn the addressee adopts the sender role and, subsequently, the changed mental constructs of his or her state function as the new preconditions for the next speech act. Conversational relevance of subsequent speech acts is established by the initial cognitive state of the participants and the rules for cognitive update that change a particular cognitive state and the rules for cooperative behaviour that dictate the performance of a particular speech act. As in realistic conversational situations, it is assumed that the relevant information can be distributed among the participants. Consequently, the structure of the dialogues may become rather complex, and may result in subdialogues and the generation of counter-questions.

2 Dialogue Coherence

In its basic form, a dialogue can be conceived as a linear alternating sequence of symbolic elements – or utterances – between two participants (see e.g. Hamblin, 1971). The various contributions in the dialogue have a meaning and a purpose – i.e. there is a relation between the symbolic elements and particular mental constructs that result from the interpretation process. In general, the utterances do not form independent segments of speech, but show a coherent structure of conversational units like words in a single sentence. Hence, there not only is a relation between the revealed symbols and the mental constructs, but also between the various constructs themselves – so-called *coherence relations*. Sometimes the words in a sentence explicitly refer to this type of relations (e.g. anaphoric reference, temporal and rhetorical phrases), but often, explicit verbalisation of these structural units is left out from the surface structure of the sentence (see also Givón (1995), Sanders & Noordman (2000)).

Coherence relations can be described on a syntactic and a semantic level. Syntactically, most models of conversation include both a *linear* and a *hierarchical* conception of coherence relations. Linearity is established by a notion of pairing – two utterances that for some reason seem to be related to each other at the same level. In Conversation Analysis, for instance, the

fundamental pairs of conversational organisation are sequences called 'adjacency pairs': a question is followed by an answer, a greeting by a countergreeting, etcetera (e.g. Levinson, 1983). From speech act theory we know the notion of 'uptake' (Austin, 1962), being the dependency of a successful performance of an illocutionary act on the reaction of the addressee. Hierarchy, on the other hand, is established by embedded structures that may appear between paired units. In conversation this can be created by so-called 'insertion sequences' – i.e. deviations from the main point that are usually expressed by the first part of an adjacency pair. Through the embedding structures of adjacency pairs, the *recursive* organisation of conversation becomes apparent. Similar structures can be found in, for instance, Power (1979), Grosz & Sidner (1986), Polanyi (1988) and Longacre (1996).

In line with, for instance, Redeker (1990) and Bateman & Rondhuis (1997), I will assume that the semantic nature of the coherence relations is twofold: a. *informational*, where the relation between the units corresponds to an existing relation in the world that is described by the discourse (co-reference, spatio-temporal relations, causation, and the like) and b. *intentional*, where the relation is not between the state of affairs described in the units, but between the mental constructs of the dialogue participants in terms of attitudes, such as beliefs and intentions (the illocutionary and perlocutionary relations, such as question-answer, offer-rejection and threat-defence). The informational view often refers to discourse produced by single speakers, while the intentional view is often used in connection with dialogue situations. Clearly, coherence in a dialogue could not be established without the continuity or recurrence of informational elements. Since this paper is about dialogue, I will concentrate on the latter type of relation.

In the intentional view, the participants of the dialogue are usually assumed to generate and execute a particular plan, and the utterances are considered actions that achieve some sort of communicative effect. Plans can be viewed as sequences of mental activities based on some type of reasoning mechanism designed to accomplish some goal state, and are therefore considered as a prerequisite to action. The idea is that the participants understand language, not only when they understand the informational part, but especially when they successfully infer each other's plans and goals. In Allen & Perrault (1980), for instance, agents generate meaningful responses on the basis of a recognised plan of the other – in their model, they particularly focus on responses to implicit requests and indirect speech acts. The structure of the plan reflects the coherence between the mental constructs, which are expressed in terms of goals, shared intentions and nested beliefs.

Although the planning model has had an enormous impact on artificial intelligence approaches to natural language processing and communication, there seem to be valid reasons why it can be rejected in a first approach to the generation of conversational units (c.f. Suchmann, 1987). A problem with most of the planning work is that the belief and intentional models are overly complex and unattractive from a computational point of view. For instance, the closure of the belief and intention axioms generates an infinite set of belief and intention constructs, which among other formulae contains the preconditions for the next turn. It is unclear, however, how the next speaker selects the relevant mental constructs from a possibly infinite set of potential preconditions. Humans have only finite information resources and limited reasoning capabilities, and it is highly unlikely that they take into account the almost infinite amount of pre- and postconditions that have to be calculated prior to almost any action. Moreover, humans live in an extremely dynamic and complex world that has to be monitored constantly to avoid unexpected occurrences. Consequently, even if all the necessary calculations are made to perform a particular action, communication models must incorporate sensitivity to local circumstances and resources for the redundancy and correction of unexpected outcomes.

Another problem is the choice of the proper type of mental constructs. Whether we have to include such constructs as shared intentionality (Searle, 1992) or mutual beliefs (e.g. Clark & Marshall, 1981) probably depends on the type of dialogue phenomenon one wants to describe. It has been shown, for instance, in Taylor, Carletta & Mellish (1996) that particular mental constructs such as nested beliefs of the third level – 'A believes that B believes that A believes' – and beyond are simply unnecessary to model the properties of a co-operative dialogue.

In this paper, I will try to show that the coherence of the speech acts is tied to local interactions contingent to the agent's particular situation and that the coherence relations can be described in a strictly situated sense, entirely driven by the history of the speech acts and the dynamics of

the mental constructs of the participants. In other words, coherence is not considered as an intrinsic property of a text or a dialogue, but mainly as a mental phenomenon (c.f. Gernsbacher and Givón, 1995).

In what follows, a dialogue game and its underlying communication model will be described that enable us to generate linear and hierarchical speech act sequences. A particular instance of the model will be chosen in which participants have no access to the outside world and only receive information based on the exchange of conversational units. Describing the properties and the dynamics of the mental constructs in relation to the various dialogue contributions is an essential part of this work. In order to develop such a framework and to avoid the problems of infinity in the planning approach, the following questions will be addressed:

1. What type of mental constructs should be included to model a dialogue's basic structural properties?
2. How do the various dialogue contributions change the existing mental constructs?
3. How do these changes influence the generation of new contributions?

3 The dialogue game

The dialogue game is divided into two parts (for a similar approach, see Piwek (1998) or Amgoud, Maudet & Parsons (2000)): a. the *game-board* that contains information about a particular state of the game, and b. the *dialogue rules* that control the behaviour of the participants (generation rules) and that prescribe how the game-board changes (update rules). The game-board represents the participants' cognitive state and typically changes because of the participants' communicative actions. In line with the basic model, the participants' cognitive state roughly contains two types of information: a. information about the participants' beliefs, and b. information about their intentions. The second type of information is related to the agents' commitments to perform a specific action and gives relevance to the individual moves of the participants.

'Moves' or information flows between the two participants are composed of two elements: a. plain information about the domain of discourse (the semantic content), and b. information about the way the different mental constructs should be updated (the communicative function). Every move is completely determined by the cognitive state of the participant who has the turn to act and by the rules for co-operative behaviour that will be presented below. Since the cognitive states are updated after every move, the next move is not only determined by the previous one, as would be the case in a dialogue grammar, but also by the context of the move. Each play is a sequence – not necessarily finite – of linearly or hierarchically alternating moves.

The agents' communicative strategy is determined by the rules of the dialogue game, which is roughly played in line with the Gricean maxims of co-operation (Grice, 1975). Agents are forbidden to put forward any domain information they do not believe and are forbidden to ask anything they already believe or they believe that their partner does not believe. Questions should be answered if the information is available and, if not available, this should be indicated accordingly. This may take more than one turn, because the information to answer the question may be distributed among the participants. Beliefs or desires that are explicitly stated in the dialogue are called *manifested*.

To avoid unnecessary complexity, we will make two important simplifications. First, during the dialogue, the participants have no access to a domain of discourse, i.e. they are unable to observe or manipulate particular aspects of the domain. In other words, information only flows between the two dialogue partners like, for instance, in a telephone dialogue. Consequently, they only have *communicative* intentions. New explicit beliefs can thus only be modified in two ways: by communication with the partner and via a reasoning mechanism for the belief states of the agents. A second simplification is that the participants only hold positive information about the domain of discourse, i.e. negation is excluded. This implies that the two participants will never hold conflicting beliefs or desires, and, since alleged inconsistencies will never arise, the agents will never argue about a specific statement. The content of a statement about the domain of discourse is simply added to the belief state of the partner. This information may be incorrect with respect to a particular instance of the domain of discourse, but the incorrectness will never be discovered, since the agents have no access to the domain.

Additionally, both dialogue partners act according to the same dialogue rules, but have discrepancies in the content of their cognitive states.

3.1 The agents' cognitive state

Below, two types of domain information will be distinguished: simple propositions ('p', 'q', 'r',…) and compound propositions ('p→q',…), which connect a simple proposition (the antecedent) with a simple proposition (the consequence).

The agents' cognitive state consists of the following mental constructs:

- Private information of an agent about the domain of discourse (B_xp; '*x* believes that p').
- Information of which the other is ignorant ($B_x\neg B_y$p; '*x* believes that *y* does not believe that p').
- Desire about a particular state of the domain of discourse (D_xp; '*x* desires that p').
- A list of manifested intentions of the other (B_xIK_y<…,p>; '*x* believes that *y* intends to know that p').

Here IK_y<…,p> means that the agent *y* has an intention to know the truth value of several propositions, of which p is the last (in this case). The list may be empty, indicated by $IK_y<\varnothing>$, which means that *y* has no intentions.

We assume that the agents can reason about their beliefs by Modus Ponens (I1).

I1. B_xp & B_x(p→q)→ B_xq

The private belief states are monotonic, i.e. everything that can be inferred from previous states, can also be inferred from new belief states. Information about the desire and the intention state can be retracted after particular communicative acts. For example, if a question 'whether p' has been answered, the intention to answer this question is dropped. We are not concerned with the full details of the update mechanism, but assume that the cognitive states are updated in line with the principle I1 and the update rules presented below.

Below, we will use the function *link* that gives us the set of all antecedents that are connected to a particular consequence in a belief state. More precisely, *link* is defined in the following way:

- *link*(*x*,q) ≡ {p| B_x(p→q)}

For instance, if *x* believes that 'p→q' and believes that 'r→q', then *link*(*x*,q)={p,r}. If there is no compound proposition with q as its consequence in belief state *x*, the set is empty.

3.2 Communicative acts

Agents manifest their beliefs and intentions by means of communicative acts or moves, such as statements and questions. The content of a move consists of a simple proposition; the communicative function is tagged by one of the following markers ('?','!','*','♦'):

- Questions: [*x*,p]?
- Statements: [*x*,p]!
- Ignorance: [*x*,p]*
- Closure of the dialogue: *x*♦.

3.3 Dialogue rules

Speech acts (or moves) are fully determined by the cognitive state of the participant who performs the move and by the rules that are applicable to this state. A double arrow '⇒' links the preconditions of the move to the move itself. The left side of the arrow is of type proposition and represents the preconditions in terms of the cognitive state of an agent; the right side is of type action and represents the generated move.

Since the agents have no access to the domain of discourse, the initial move can only be a question:

G0. D_xq & ¬B_xq & $B_xIK_y<\varnothing>$ ⇒ [*x*,q]?

The first two preconditions indicate that the agent's state is out of balance[1]; the third precondition was included to avoid the agent from repeating the question every turn. Note that generation rule G0 is applicable to the initiator of the dialogue only; the rules presented below are applicable to both participants. Hence, below *x* and *y* range over both the initiator and follower.

Generation rule G1 expresses that if *x* believes that q is the last item on the manifested intention list (i.e. q has been asked for) of participant *y* and q is believed by *x*, then *x* will answer q:

G1. B_xIK_y<…,q> & B_xq ⇒ [*x*,q]!

[1] In short, a balanced state is a state where the desire of the agent is in agreement with his/her beliefs and, therefore, no further action is required.

If *x* does not know the answer, *x* may ask a counter-question. The counter-question can only be asked if the agent finds the antecedent of a linked proposition, and if he or she does not believe that the other is ignorant with respect to the linked proposition (G2):

$$\text{G2. } B_x IK_y\langle \ldots,q\rangle \;\&\; \neg B_x q \;\&\; p \in \mathit{link}(x,q) \;\&\; \neg B_x \neg B_y p \Rightarrow [x,p]^?$$

If *x* does not know the answer and cannot ask a counter-question, *x* will manifest his or her ignorance (G3).

$$\text{G3. } B_x IK_y\langle \ldots,q\rangle \;\&\; \neg B_x q \;\&\; \neg(p \in \mathit{link}(x,q) \;\&\; \neg B_x \neg B_y p) \Rightarrow [x,q]^*$$

Finally, in G4 a closing act is generated if the intention list is empty and if the situation is not imbalanced:

$$\text{G4. } B_x IK_y\langle \varnothing\rangle \;\&\; \neg(D_x q \;\&\; \neg B_x q) \Rightarrow x^{\blacklozenge}$$

To avoid an infinite sequence of closing acts, a meta-rule has been defined to close the dialogue:

> **Closing (CL)**
> Both dialogue partners stop generating communicative acts iff two successive closing acts are performed (i.e. the sequence $x^{\blacklozenge}$ & $y^{\blacklozenge}$).

3.4 The update of cognitive states

The update function yields a new cognitive state depending on the old state and the move just performed. To represent the consequences of a particular move, we introduce '>'. The left side is of type action and represents the performed move; the right side represents the postconditions and denotes how the cognitive states should be updated. If information has to be removed, the relevant attitudes are preceded by *Del*. So, for instance, $\mathit{Del}(B_x IK_y\langle q\rangle)$ means that q should be deleted from the intention list.

In update rule U1, it is expressed that if *x* utters a statement with content q, the following states are updated: a. *y* now believes that q, b. the content q will be removed from the intention list and, if relevant, c. *x* no longer believes that *y* is ignorant with respect to q:

$$\text{U1. } [x,q]^! > B_y q \;\&\; \mathit{Del}(B_x IK_y\langle q\rangle) \;\&\; \mathit{Del}(B_x \neg B_y q)$$

Rule U2 expresses that if *x* utters a question with content q, *y* subsequently believes that q is a communicative intention of *x* and, therefore, q will be added to the end of the list:

$$\text{U2. } [x,q]^? > B_y IK_x\langle q\rangle$$

Rule U3 expresses that if *x* indicates that he or she has no information about q, q will be added to *y*'s belief about *x*'s ignorance and q will be removed from the intention list and, if present, from the desire state:

$$\text{U3. } [x,q]^* > B_y \neg B_x q \;\&\; \mathit{Del}(B_x IK_y\langle q\rangle) \;\&\; \mathit{Del}(D_y q)$$

The last rule, U4, expresses that cognitive states do not change after a closing act:

$$\text{U4. } x^{\blacklozenge} > \otimes$$

4 A dialogue example

I turn now to an example where John and Mary play the co-operative dialogue game based on the previously introduced mental constructs, and the generation and update rules. First, we present an abstract version of the example, and next, we convert the example into a 'natural' language dialogue.

In Table 1, we have depicted the gameboard, i.e. the cognitive states of John and Mary, the communicative acts (MOVES) and, in addition, a reference to the applied update and generation rules. The information that represents the preconditions for the next move is indicated in bold italics; empty states are indicated by '∅'. In the example, we left out John's and Mary's desire state, since in this particular situation the desire states do not change during the course of the dialogue.

In the initial situation, John believes that $p \rightarrow q$ and believes that $r \rightarrow q$, John has no desires; Mary believes that $s \rightarrow p$ and believes that r, Mary has the desire that q. Therefore, Mary is the initiator and starts with the initial question whether q. John is unable to answer the question directly, but may find an answer if he has information about p or r.

Nr.	John			MOVES	Mary		
	B_J	$B_J I_M$	$B_J \neg B_M$		B_M	$B_M I_J$	$B_M \neg B_J$
Initial state G0	p→q r→q	∅	∅		s→p r	∅	∅
1				[M,q]?			
U2 G2	***p→q*** r→q	<***q***>	∅		s→p r	∅	∅
2				[J,p]?			
U2 G2	p→q r→q	<q>	∅		***s→p*** r	<***p***>	∅
3				[M,s]?			
U2 G3	p→q r→q	<q,***s***>	∅		s→p r	<p>	∅
4				[J,s]*			
U3 G3	p→q r→q	<q>	∅		***s→p*** r	<***p***>	***s***
5				[M,p]*			
U3 G2	***p→q*** ***r→q***	<***q***>	***p***		s→p r	∅	s
6				[J,r]?			
U2 G1	p→q r→q	<q>	p		s→p ***r***	<***r***>	s
7				[M,r]!			
U1 G1	p→q, r r→q, ***q***	<***q***>	p		s→p r	∅	s
8				[J,q]!			
U1 G4	p→q, r r→q, q	∅	p		s→p r,q	∅	s
9				M♦			
U4 G4	p→q, r r→q, q	∅	p		s→p r,q	∅	s
10				J♦			
U4 CL	p→q, r r→q, q	∅	p		s→p r,q	∅	s

Table 1: John and Mary try to solve the problem whether q is true. Initially, the information about q is distributed; in the final state, both John and Mary believe that q is true.

Hence, according to rule G2, John will generate the question whether p or the question whether r; in Table 1, John starts with the first solution (move 2). Now, the same rule is applicable to Mary's situation, therefore she asks for s, but subsequently, John has to inform Mary that he has no information about s (move 4). Dead end, the intention s will be retracted by rule U3. There is a way out, however: q is still on the intention list and p cannot be asked for again, since John now believes that Mary is ignorant about p. John will question r and, since Mary has direct information about r, the question can be answered by Mary (move 7). In turn, John can answer Mary's initial question and finally, since all manifested intentions are removed, the dialogue will be closed (move 9 and move 10).

To make the example a little more concrete, suppose a domain where the following propositions hold:

p: *'Peter smokes cigars'*

q: *'Peter is happy'*
r: *'Peter works in Utrecht'*
s: *'Peter is a manager'*

In correspondence with Table 1, John and Mary initially believe the following information:

B_{John}: *'If Peter smokes cigars, then he is happy'* ($p \rightarrow q$)
'If Peter works in Utrecht, then he is happy' ($r \rightarrow q$)

B_{Mary}: *'If Peter is a manager, then he smokes cigars'* ($s \rightarrow p$)
'Peter works in Utrecht' (r)

Mary wants Peter to be happy, but she has no direct information about Peter's emotional state. Below, the example is presented in 'natural' language.

Dialogue I
1. Mary: Is Peter happy?
2. John: Does Peter smoke cigars?
3. Mary: Is Peter a manager?
4. John: I do not know whether Peter is a manager.
5. Mary: I do not know whether Peter smokes cigars.
6. John: Does Peter work in Utrecht?
7. Mary: Peter works in Utrecht.
8. John: Peter is happy.
9. Mary: Thank you.
10. John: Thank you.

5 Discussion

Simple in its basic form the framework enables us to produce abstract conversations with some properties that agree strikingly with coherence properties found in, for instance, Conversation Analysis.

In Table 1 we notice the elementary structural phenomena discussed at the beginning of this paper. First of all, we can observe the linear management organisation of adjacency pairs, such as question-response (moves 3-4, and moves 6-7) and the closing of the dialogue (moves 9-10). Second, we observe the hierarchical organisation of insertion sequences, such as moves 3 and 4 between the question in 2 and its reply in 5, and moves 2-7 between 1 and 8. Depending on the initial states, the dialogue rules generate an arbitrary number of levels of subsequences and the final reply may be originated many turns away from the initial question. Syntactically, the structure of the dialogue can be schematised as follows (Q = Question; R = Response):

$$(Q_1((Q_2(Q_3{-}R_4)R_5)(Q_6{-}R_7))R_8)$$

Note that the R-utterances always cause a deletion of the intentional content of a question in Table 1.

Clearly, the dialogue is still unnatural and lacks many of the ingredients that we usually observe when we study the properties of natural language dialogue. The unnaturalness has many reasons. One is that the generation rules do not take into account the generation of extra management utterances and coherence markers; another reason is that the domain language is far too simple and that only simple propositions can be communicated. To obtain a more realistic dialogue, a 'decorated' version of Dialogue I was constructed. In the decorated dialogue, we included, for instance, pronominalisation and expressions that are not generated by the rules G0-G4, such as the opening of the dialogue, indirect questioning, thanking and particles on the process level ('Aha', 'Well', 'Uh'). The basic structure of both dialogues, however, is isomorphic to the structure presented in Table 1.

Dialogue II (decorated version of Dialogue I)
1. Mary: Hello John, I have a question. Can you tell me whether Peter is happy?
2. John: I don't know, does he smoke cigars?
3. Mary: Uh… that depends, is he a manager?
4. John: Sorry, I don't know.
5. Mary: Then I don't know whether he smokes cigars.
6. John: Aha, wait… does he work in Utrecht?
7. Mary: Yes, he does.
8. John: Well, in that case, don't worry, he is happy.
9. Mary: Great, thanks a lot, bye.
10. John: Okay, bye.

An extension along the line of Levinson's preference organisation (Levinson, 1983), in which linguistic markers are added in the dialogue if non-preferred sequences of moves are generated, looks promising. Although in a premature stage of development, another interesting candidate is a context-change approach where specific types of feedback are considered as side-effects of updates on the cognitive states. It should be stressed, however, that I consider these types of

management moves to be a second order effect, which are based on more fundamental principles such as the ones presented in this paper.

Although admittedly still incomplete, the framework thus provides an explanation of some important dialogue phenomena. The game-board accurately shows how, during a dialogue, the cognitive states of participants change as a result of the exchange of information. In order to generate these conversational units, neither a planning approach, nor a speech act grammar approach is needed (or wanted) to build coherent structures of conversation. Coherence relations can be described in a situated sense, based on the context of the dialogue in terms of the agents' cognitive state and the immediately preceding conversational unit. Since only a limited number of attitudes was included, the framework does not suffer from the same computational complexity as in most planning approaches where agents are not only able to reason about the discourse domain, but also about their own and their partner's beliefs and intentions.

6 Credits

Parts of this paper have been published in Pragmatics & Cognition 9:1 (2001), 37-68. ISSN 0929-0907.

References

Allen, J.F. & Perrault, C.R. 1980. Analyzing Intention in Utterances. *Artificial Intelligence*, 15: 143-178.

Amgoud, L., Maudet, N. & Parsons, S. 2000. Modelling Dialogues Using Argumentation. *Proceedings of the Fourth International Conference on Multi-Agent Systems* (ICMAS 2000), Boston (MA), July 10-12, 2000, 31-38.

Austin, J.L. 1962. *How to do Things with Words*. Oxford: Clarendon Press.

Bateman, J.A. & Rondhuis, K.J. 1997. Coherence Relations: Towards a General Specification. *Discourse Processes*, 24: 3-49.

Bunt, H.C. 1989. Information dialogues as communicative action in relation to partner modelling and information processing. In: M.M. Taylor, F. Néel & D.G. Bouwhuis (eds.) *The Structure of Multimodal Dialogue.* Amsterdam: North Holland, 47-73.

Clark, H.H. & Marshall, C.R. 1981. Definite Reference and Mutual Knowledge. In: A.K. Joshi, B.L. Webber and I.A. Sag (eds.) *Elements of Discourse Understanding*. Cambridge: Cambridge University Press, 10-63

Gernsbacher, M.A. & Givón, T. 1995. *Coherence in Spontaneous Text*. Amsterdam: John Benjamins Publishing Company.

Givón, T. 1995. Coherence in Text vs. Coherence in Mind. In: M.A. Gernsbacher & T. Givón (eds.) *Coherence in Spontaneous Text*. Amsterdam: John Benjamins Publishing Company. 59-115.

Grice, H.P. 1975. "Logic and Conversation". In: P. Cole & J. Morgan (Eds.): Speech Acts. Syntax and Semantics, Vol. 11. New York: Academic Press. pp. 41-58

Grosz, B.J. & Sidner, C.L. 1986. "Attention, Intentions, and the Structure of Discourse". Computational Linguistics, 12(3), 175-204.

Hamblin, C.L. 1971. Mathematical Models of Dialogue. *Theoria*, 37: 130-155.

Levinson, S. C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Longacre, Robert, E. 1996. The Grammar of Discourse. New York: Plenum Press.

Piwek, P. 1998. *Logic, Information & Conversation*. PhD Thesis. Eindhoven University of Technology.

Polanyi, L. 1988. "A Formal Model of the Structure of Discourse". Journal of Pragmatics, 12, 601-638.

Power, R. 1979. "The Organisation of Purposeful Dialogues". Linguistics, 17, 107-152.

Redeker, G. 1990. Ideational and Pragmatic Markers of Discourse Structure. *Journal of Pragmatics*, 14: 367-381.

Sanders, T.J.M. & Noordman, L.G.M. 2000. The Role of Coherence Relations and their Linguistic Markers in Text Processing. *Discourse Processes*, 29(1): 37-60.

Searle, J.R. 1992. Conversation. In: Searle, J.R. et al. *(On) Searle on Conversation*. Amsterdam: John Benjamins Publishing Company. 7-30.

Suchman, L.A. 1987. *Plans and Situated Actions: the Problem of Human-Machine Communication*. Cambridge: Cambridge University Press

Taylor, J.A., Carletta, J. & Mellish, C. 1996. Requirements for Belief Models in Co-operative Dialogue. *User Modelling and User-Adapted Interaction*, 6: 23-68.

Improving Text Fluency by Reordering of Constituents

Katja Filippova and **Michael Strube**
EML Research gGmbH
Schloss-Wolfsbrunnenweg 33
69118 Heidelberg, Germany
http://www.eml-research.de/nlp

Abstract

We present a method for improving local coherence in German with implications for automatically as well as for human-generated texts. We demonstrate that local coherence crucially depends on which constituent occupies the initial position in a sentence. We provide statistical evidence based on a corpus investigation and on results of an experiment with human judges to support our hypothesis. Additionally, we implement our findings in a generation module for determining the *Vorfeld* constituent automatically.

1 Introduction

Multi-document summarization extracts important sentences from different input documents and joins them together in one output document. Obviously, this procedure may not lead to well-written summaries as they may lack coherence. Even if the extracted sentences exhibit some coherence on the entity level, they cannot present the information in the right word order thus leading to difficult to read sentences.

In this paper we propose a method for improving local coherence of German texts by making transitions between sentences smoother. We show that the fluency of a transition from one sentence to the next one depends on which constituent occupies the initial position of the next one. This work is done within a project on automatic text-to-text biography generation which proceeds as follows: Given a number of documents about a certain person and a keyword query as input, first, the sentences which are relevant to the user are found; second, a coherent text is generated from them. The tasks performed during the generation phase, when selected sentences are being put together, concern the order of sentences (global coherence) as well as the order of constituents within a sentence and pronominalization (local coherence). In this paper we investigate the tasks constituent order and pronominalization.

Other applications which could benefit from our method are text summarization, machine translation, or any other application whose output consists of more than one sentence. Moreover, as we will demonstrate, simple rules can improve the fluency of a text produced by human writers.

Unlike some other approaches investigating the relation between information structure and word order, our scope is not limited to noun phrases only, but also includes adverbs and discourse connectives. Because of that we deliberately decided not to formalize our approach within such well-established frameworks as, for example, Centering (Grosz et al., 1995; Prince, 1999). The modifications needed for such formalization would require extending the notion of the (backward-, forward-looking) center not just to constituents other than NPs but also to propositions and would lead to a loss of conceptual simplicity of this framework.

The remainder of the paper is organized as follows: Having outlined related work on generation (Section 2) and on information structure (Section 3), we first motivate and present our approach (Section 4), then we introduce our data whose analysis provides statistical evidence for our hypothesis (Section 5). The results of an experiment with human judges which also confirm the claims concerning the functions of the VF and an application to generation are presented in Section 6 and Section 7 respectively.

2 Related Work

Recent papers on local coherence have suggested algorithms for ordering discourse units like sentences or clauses while phrase ordering within a sentence has not received as much attention. Barzilay et al. (2002) consider the task of sentence ordering within a multi-document summarization approach and experiment with majority and chronological ordering. Lapata (2003) infers constraints on sentence order from a corpus of domain specific texts and approaches the problem in a probabilistic manner. Karamanis et al. (2004) assume a set of clauses as the input and compute a metric for text structuring which utilizes the Centering perspective on coherence. Since all these studies concern English, the question of phrase or word ordering does not play an important role there. The German language, allowing for word order variations, introduces another challenge for generating locally coherent texts.

Kruijff et al. (2001) combine the Prague School and the systemic-functional frameworks and recognize the importance of the information structure for word order variation. They propose an approach to characterizing word order which can be equally well applied to different languages, no matter whether the word order is driven pragmatically or syntactically. In their study, they consider English, Czech and German and demonstrate that in each case the word order can be determined by so called *communicative dynamism* (Firbas, 1974) as well as by the language specific *systemic ordering* (Sgall et al., 1986). Generally, communicative dynamism prescribes that explicitly or implicitly given entities (termed context-bound) precede new information and systemic ordering describes the canonical order in a clause which in case of German corresponds to the following:

Actor < TemporalLocative < SpaceLocative < Means < Addressee < Patient < Source < Destination < Purpose

The authors apply their algorithm to English and Czech software instruction manuals and note that it can be applied to German as well.

3 Background on Information Structure

Due to divergencies in terminology, information structure is notoriously difficult to talk about (see Levinson (1983, p.x)). Therefore, given that the sentence topic is what our proposal relies on, it is a matter of necessity to provide an operative definition and clearly express similarities and differences to existing approaches before presenting our idea.

In general, there are two views on *topic*: as what the sentence is about, and as the measure of salience of an entity. The former has its origin in the work of Strawson (1964); an extreme example of the latter is Givon (1983) whose topic is very similar to the notion of the backward-looking center in the Centering model (Grosz et al., 1995). We adhere to the first view and define topic based on the pragmatic relation of aboutness only, thus excluding the discourse status of the referent from our definition. The topic is the referent the proposition is about, or more precisely, the referent the speaker assumes to be a center of current interest. Consequently, we do not subscribe to the view that the element about which the information is provided always occupies sentence initial position. On the contrary, like Reinhart (1981), we think that topiclike elements may and do appear on other positions as well. The role of the sentence initial element, on the other hand, is more similar to the role of 'real' topics as described by Chafe (1976) in that they 'are not so much "what the sentence is about" as "the frame within which the sentence holds"'. Splitting these two functions makes our approach different from Vallduví's (1990). For him, 'by starting a sentence with link speakers indicate to hearers that the focus must go to that address, and enter the information under its label.' (Vallduví, 1990, p.59). Although in many cases his link and our topic coincide, we find it unintuitive and improbable that dates or discourse connectives are the addresses where the new information is attached.

Apart from that, we distinguish between what has been introduced by Chafe (1987) as active, accessible and inactive referents. Topic and activeness correlate in that the most easily processed sentences are those whose topic referents are active in the discourse (Lambrecht, 1994, p.165).

Our approach is similar to the one of Kruijff et al. (2001) in that we also consider the relation between word order and information structure but differs from it in several respects. Firstly, Kruijff et al. (2001) concentrate on how to generate not just a grammatical but acceptable ordering whereas we focus on how to determine not just a grammatical and acceptable ordering but the one that makes the transition as smooth as pos-

sible. Secondly, we extend context-bound information to accessible and treat context-bound NPs, temporal expressions and discourse connectives in the same way. The fact that absolute temporal expressions are perfectly acceptable in the beginning of a sentence (Heidolph et al., 1981, ch.4) which can not be explained in terms of given and new information is noticed by Kruijff-Korbayová et al. (2002) where locations or temporal expressions are treated as the theme or point of departure of the clause. Extending given information to accessible makes it possible to treat these cases uniformly.

4 Our Hypothesis

Because of the fixed verb position, the German V-second clause is divided into two parts. The part preceding the finite verb, "prefield", or *Vorfeld (VF)* usually contains only one constituent, and the part between the finite verb or complementizer and the verbal elements at the end of the clause, "middle field", or *Mittelfeld (MF)* incorporates the rest. In (1) and all the following examples the VF is indicated by italics:

(1) *Marie Curie* wurde am 7. November
Marie Curie was on the 7th November
1867 in Warschau geboren.
1867 in Warsaw born.

'Marie Curie was born in Warsaw on the 7th of November 1867'

The problem of constituent ordering in German can be reformulated then: Which constituent is to be placed in the VF? What should be the order of constituents in the MF?

Concerning the VF, the following claims are made: the VF, being a cognitively prominent position (Gernsbacher & Hargreaves, 1988), has two major[1] functions: Whenever the topic of a sentence needs to be established, it is placed into the VF. Otherwise, if the topic has already been established and is still activated in the mind of the reader, it should be pronominalized and there is no need for it to occupy the VF (see Frey (2004) for recent research on the topic position in German). In this case the VF is the position responsible for a smooth transition from the previous sentence to the current one. The smoothness or fluency of transitions is ensured by placing a constituent in the VF which helps linking the introduced sentence to the representation readers have already built in their mind.

The best candidate for the VF is to be selected from the set of accessible elements. These are entities accessible due to the preceding context, e.g. repeated mentions or anaphoric elements, inferentially accessible constituents (bridging anaphora). Temporal expressions – absolute, *am 23. Mai 1900 (on the 24th of May)*, or relative, *Im gleichen Jahr (in the same year)* – belong to this group because of the relevance of the time scale for the biography genre. For newspaper texts locations (e.g. *Berlin, Sankt-Petersburg*) and other named entities (e.g. *SPD, Merkel, SAP*) are expected to be as readily accessible as temporal expressions here. Discourse connectives count as accessible constituents as well: They establish a relation between the proposition expressed in the current sentence and propositions expressed earlier in the discourse. *So (so), anschliessend (finally), dabei (in doing so)* are examples of such connectives but not *weil (because), obwohl (although)* which link two clauses within one sentence. Proadverbials, e.g. *damit (with that), darüber (about that)* are also included in this group.

The first impression might be that it is inconsistent to unify such diverse phenomena as discourse connectives and noun phrases. This impression may change if we distinguish between structural connectives and discourse adverbials and consider the latter as anaphora (Webber et al., 2003). From this point of view the fact that an adverbial connective, e.g. *sonst* (otherwise), and an inferrable NP, e.g. *die Familie* (the family) following a discourse where the parents are mentioned, are both treated as accessible elements, should not be surprising because both of them are instances of anaphora (bridging anaphora in the latter case).

To sum up, we identify the topic in the sentence, which is the address for new information, we also find *other* linking or framing elements and in case of the topic being activated place the best candidate from the linking list to the VF. We hypothesize that this strategy provides smoother transitions than reserving the VF for the topic. The rest of the paper provides evidence from different sources which confirm our hypothesis.

[1] Other elements, such as contrastive topics, are encountered in the VF as well but considerably less frequent.

5 Data

5.1 Preprocessing

The data we investigate is a collection of biography texts from the German version of Wikipedia[2]. The data is homogeneous in the sense that it contains all biographies under the Wikipedia category of physicists.

Fully automatic preprocessing in our system comprises the following stages: First, a list of people of a certain Wikipedia category is taken and for every person an article is extracted. The text is purged from Wiki tags and comments, the information on subtitles and paragraph structure is preserved. Second, sentence boundaries are identified with a Perl CPAN module[3] whose performance we improved by extending the list of abbreviations and modifying the output format. Next, the sentences are split into tokens. The TnT tagger (Brants, 2000) and the TreeTagger (Schmid, 1997) are used for tagging and lemmatizing. Finally, the texts are parsed with the CDG dependency parser (Foth & Menzel, 2006). Thus, the text is split on three levels: paragraphs, sentences and tokens, and morphological and syntactic information is provided.

A publicly available list of about 300 discourse connectives was downloaded from the Internet site of the Institute of the German Language[4] (Institut für Deutsche Sprache, Mannheim) and slightly extended. These are identified in the texts and annotated automatically as well. Named entities are classified according to their type using information from Wikipedia: *person, location, organization* or *undefined*. Given the peculiarity of our corpus, we are able to identify all mentions of the biographee in the text by simple string matching. We also annotate different types of referring expressions (*first, last, full name*) and resolve anaphora by linking personal pronouns to the biographee provided that they match in number and gender.

Temporal expressions (both relative and absolute) and VFs are identified automatically by a set of patterns. VFs, for example, are determined as the part of the sentence standing before the root verb.

[2]http://de.wikipedia.org

[3]http://search.cpan.org/~holsten/Lingua-DE-Sentence-0.07/Sentence.pm

[4]http://hypermedia.ids-mannheim.de/index.html

5.2 Corpus Analysis

We analyzed 370 texts with an average length of 17 sentences, 6521 sentences in total. 2857 of them mentioned the biographee (with the name or with a personal pronoun) and hence were of interest for us. Whenever such a sentence opens a new section in an article, we assume that the topic should be explicitly established, therefore a concrete reference to the person is needed and the referring expression should be placed in the VF, no matter what its syntactic function is. Whenever a sentence is preceded by one or several sentences which already are about the biographee, we assume the person to be activated in the mind of the reader. In such a case a pronominal reference should be used, and the preferred position for it is the MF. Examples (2) and (3) should make the point clearer:

(2) a Familie und frühe Jahre
Familiy and early years
'Family and early years'

b *Marie Curie* wurde am 7. November
Marie Curie was on 7th November
1867 als Maria Salomea Sklodowska in
1867 as Maria Salomea Sklodowska in
Warschau geboren.
Warsaw born.
'Marie Curie was born in Warsaw on the 7th of November 1867 as Maria Salomea Sklodowska.'

(3) a Zusammen mit ihrem Mann Pierre
Together with her husband Pierre
Curie und dem Physiker Antoine Henri
Curie and the physicist Antoine Henri
Becquerel erhielt sie 1903 den
Becquerel received she 1903 the
Nobelpreis für Physik.
Nobel prize in physics.
'Together with her husband Pierre Curie and the physicist Antoine Henri Becquerel, she received the Nobel prize in physics in 1903.'

b *Acht Jahre später* wurde ihr der
Eight years later was her the
Nobelpreis für Chemie verliehen.
Nobel prize in chemistry given.
'Eight years later, the Nobel prize in chemistry was given to her'

	pronoun	name	conn.	temp.expr.
VF	680	953	359	1358
MF	2177	602	1013	1355
Total	2857	1555	1372	2713

Table 1: Distribution of expressions according to their position

Following a title, (2b) opens the biography which is devoted to Marie Curie. The topic is established by placing the full name reference to the VF. Pronominalization and placing the constituent in the MF are deprecated. In (3) the situation is different: The biographee is already activated in the mind of the hearer, and in (3b) there is a better candidate for the VF – a temporal expression.

Considering sentences with a reference to the biographee, it was of interest to us to see which constituents usually occupy the VF. Table 1 shows the distribution of the expressions referring to the biographee (pronominal and non-pronominal), temporal expressions, and connectives with respect to their position in a sentence. The results clearly indicate that the VF is not a preferred position for pronouns, whereas non-pronominal reference may appear in the VF about one and a half times as often as in MF. Unfortunately, some connectives are ambiguous and can mark relations between clauses of one sentence as well as relations between sentences. In the future we plan to improve the annotation and rule out all instances of intrasentential connectives. The fact that temporal expressions appear in the VF as often as in the MF does not support our hypothesis so far. In order to find out which candidate is more preferrable, we performed an experiment with human judges.

6 Experiment

In order to verify our hypotheses on text fluency, we performed an experiment with human judges, all native speakers of German who were presented with 24 short text fragments from our corpus. Each fragment had two possible continuations which were identical in all aspects but for the word order. The order of the two alternative sentences as well as the order of the fragments was generated randomly. The judges were asked to choose from the two variants the one which continues the preceding text in the most fluent way or choose nothing in case of both continuations sound equally fluent.

(4) a Nach seiner Kriegsteilnahme am
After his War participation in the
Ersten Weltkrieg folgte er
First World War followed he
Berufungen nach Jena, Stuttgart,
invitations to Jena, Stuttgart,
Breslau und Zürich.
Wroclaw and Zürich.

'Having taken part in the First World War, he accepted invitations from Jena, Stuttgart, Wroclaw and Zürich'

b′ *Dort* belegte er den Lehrstuhl für
There hold he the chair for
Theoretische Physik, den vor ihm
theoretical physics, which before him
bereits Albert Einstein und Max von
already Albert Einstein and Max von
Laue inne hatten.
Laue had.

b′′ *Er* belegte dort den Lehrstuhl für
He hold there the chair for
Theoretische Physik, den vor ihm
theoretical physics, which before him
bereits Albert Einstein und Max von
already Albert Einstein and Max von
Laue inne hatten.
Laue had.

'He hold there the chair of theoretical physics, which was before him occupied by Albert Einstein and Max von Laue'

Sentences (4b′) and (4b′′) have the same propositional content and differ only in what stands in the VF: the proadverbial *there* or the personal pronoun *he*. If our hypothesis is right, then the judges would choose (4b′) more often than (4b′′).

The purpose of the experiment was twofold: to check, first, whether in cases where topic establishing is necessary (e.g. example (2)), the VF is the preferrable position for the topic. Second, whether an established topic occupying the VF makes the transition to the sentence smoother, or there are better candidates for this position (example (4)).

18 human judges (9 female and 9 male) took part in the experiment. The statistical significance of our results was computed using χ^2 test on the

	inferrable	temp.expr.	connective	proadverbial	total
pronoun	– + + +	∘∘ + +	– + + +	– + – – –	17
name	+		+		2

Table 2: Results of the experiment with human judges

$p = 0.01$ level or below. It turned out that the preference for a certain variant was significant if it was chosen by at least 15 judges.

6.1 Topic-establishing Sentences

We selected three section initial sentences which mention the biographee because such sentences open a new discourse topic (this is explicitly marked by using section titles) and therefore require non-pronominal reference to the person. Three pairs – a sentence and a propositionally equivalent variant of it – were presented to the judges. Example (2) is one of such fragments. In these three fragments the judges had a choice of what to place into the VF: an absolute temporal expression, an NP with a reference to a previously mentioned and therefore accessible person, and an inferentially accessible NP or a name reference to the biographee. In all three cases the biographee was preferred over other candidates for the VF position, and in two of the cases the difference was significant. This finding alone is in accordance with the well-known correlation between topics, subjects and sentence initial position and does not have a dramatic impact on coherence.

6.2 Sentences with the Established Topic

The second part of the experiment concerned sentences where the biographee is established as the topic due to the immediately preceding context (like (3a,b) and (4a,b)). From the 19 test pairs of this kind, seventeen contained a pronominal reference, and in two other pairs the biographee was referred to with the last name. For these examples, constituents of the following kinds were supposed to be better candidates for the VF: *inferrable constituents* (5 fragments), *temporal expressions* (4), *discourse connectives* (5), or *proadverbials* (5). Here we distinguish between connectives which have a distinct semantic meaning (e.g. temporal or additive), these are labeled as *discourse connectives*, from *proadverbials* (*dabei*, *darüber*) whose meaning is usually context-dependent.

Syntactic function was expected to play a minor role for the choice of the best constituent for the VF. This parameter was set in favour of the activated referent: in all sentences the syntactic role of the biographee is subject.

For *every* pair it turned out that the majority of judges preferred accessible constituents over activated subjects. In five cases, the judges preferred the modified version over the original sentence, i.e. the sentence from the Wikipedia article, because they found the modified fragment sound more fluent. A plus (+) in Table 2 stands for cases where the difference in preferences is significant on the $p = 0.01$ level, a circle (∘) for significance on the $p = 0.05$ level, a minus (–) for non-significant preference.

Interestingly, for both examples with a non-pronominal reference to the biographee the connective as well as the accessible constituent were preferred significantly more often. This brings us to the conclusion that for a fluent transition the established topic should not be placed into the VF no matter what its surface or syntactic realization is. The last two test sentence pairs let the reader choose between, first, a temporal expression and an accessible constituent; second, a temporal expression and a proadverbial. For the former case, no difference in preferences was found; for the latter, the proadverbial was picked significantly more frequent than the temporal expression.

Obviously, in order to rank candidates of different kinds more subtle experiments need to be performed: Form of the expression, semantics of connectives, and degree of accessibility should be taken into account. So far, it can only be stated that, concerning candidates for the VF, the established topic follows any of the listed above.

7 Implications for Generation

In this section we present an application of our findings to the automatic identification of the best candidate for the VF. This can be considered a first step towards the automatic generation of phrase and word order. We split our 370 articles corpus into training and testing sets and selected parsed sentences which mention a biographee. Thus we obtained 3080 and 616 sentences for training and

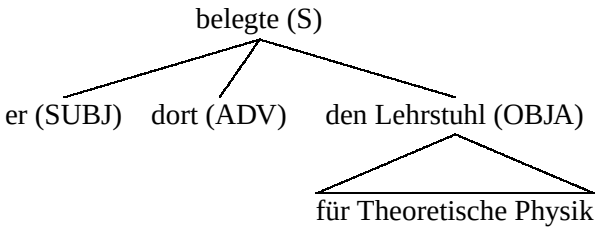


Figure 1: Essential part of example (4)

testing respectively. The number of candidate constituents for a sentence ranged from 1 to 8 being 4 on average. Consider the example (4b″) again: *Er belegte dort den Lehrstuhl für Theoretische Physik, den vor ihm bereits Albert Einstein und Max von Laue inne hatten.* For our purposes we may ignore the structure dependent on the OBJA *Lehrstuhl* and consider only the nodes dependent on the root verb (Figure 1). In this example there are three candidates which can occupy the VF because there are three constituents dependent on the main verb.

Using maximum entropy learning which has been successfully applied to a number of NLP tasks, including word order generation (Ratnaparkhi, 2000), we trained a binary classifier which for every constituent estimated the probability of it being in the VF. The three feature vectors for Figure 1 are presented in Table 3. The first seven features apply to any candidate, these are the word immediately dependent on the verb (DEP.WORD), the non-auxiliary root verb (VERB), the lexical head of the dependent constituent (LEX.HEAD), part of speech (POS), syntactic function (SYNT), maximal depth (DEP) and the length (LEN) of the constituent. If the constituent is a named entity, a temporal expression or a connective then this is expressed as TYPE. If it is a person, then it is marked whether it refers to the biographee (ROLE) and the type of the referring expression is given (REF.EXPR). For a temporal expression, REF.EXPR expresses whether it is an absolute or a relative one. The last line gives values of the temporal expression from example (2b) – *am 7. November 1867*. From all candidates for one sentence, the one with the highest probability was chosen as the best candidate. The results were evaluated against the original ordering. Note, that with this setting contextual information is totally absent, and inferrable constituents can not be identified.

From the 616 test instances the algorithm made a mistake in 211 cases, thus the accuracy is about

Wikipedia	MaxEnt	
pron	temp	17
pron	conn	8
name	temp	11
XP	pron	22

Table 4: Types of errors with their frequency

65%. Having analysed the first 100 errors, we summarize our observations in Table 4. In 17 cases the algorithm preferred a temporal expression over a pronoun which occupied the VF in the original Wikipedia article. This counts as a mistake although, as the experiment has demonstrated, human judges find text more coherent provided there is a temporal expression and not a pronoun in the VF. Likewise, the fact that 8 connectives were classified falsely does imply that the generated order would make the text less coherent than the original. Apart from that, name references may have been used in topic established sentences, which means that some of the 11 mistakes might not be errors, just as it is in the case of pronouns.

In 22 cases a pronominal reference to the biographee was chosen instead of a NP, PP or a sub-clause (labeled XP in the table) which were accessible due to the preceding context. By extending the list of features and taking the context into account we expect to improve the results significantly. Whereas temporal expressions, NEs and connectives can be identified relatively easily, identifying inferrable NPs is a much harder task. A straightforward way to measure inferrability is by means of string matching but, obviously, this method would work for the most trivial cases only. Measuring semantic relatedness (using GermaNet (Gurevych, 2005) or Wikipedia (Strube & Ponzetto, 2006)) could offer a more intelligent way of finding accessible referents.

8 Conclusions

Corpus investigation as well as experiments on constituent reordering confirmed our claims concerning the role of the VF: In most cases, it is either the topic establishing position, or the position for accessible constituents. In line with the hypothesis, human judges find transitions between sentences smoother when the VF is occupied by accessible elements, and not by topics, no matter what their discourse status is. The first results on automatic phrase ordering motivate fur-

DEP.WORD	VERB	LEX.HEAD	POS	SYNT	DEP	LEN	TYPE	ROLE	REF.EXPR.
[er]	belegte	er	pper	subj	d=0	l=1	pers	biogr	re=pron
[dort]	belegte	dort	adv	adv	d=0	l=1			
[lehrstuhl]	belegte	lehrstuhl	nn	obja	d=7	l=18			
[am]	geboren	november	card	pp	d=3	l=5	temp		re=abs

Table 3: Vectors for the three constituents in Figure 1 and the temporal expression from example (2b)

ther research in this direction. In the future we would like to automatically generate word order for whole sentences. The ultimate goal is to apply the method to generating coherent biographies.

Acknowledgments: This work has been funded by the Klaus Tschira Foundation, Heidelberg, Germany. The first author has been supported by a KTF grant. We would also like to thank our judges.

References

Barzilay, Regina, Noemie Elhadad & Kathleen R. McKeown (2002). Inferring strategies for sentence ordering. *Journal of Artificial Intelligence Research*, 17:35–55.

Brants, Thorsten (2000). TnT – A statistical Part-of-Speech tagger. In *Proceedings of the 6th Conference on Applied Natural Language Processing*, Seattle, Wash., 29 April – 4 May 2000, pp. 224–231.

Chafe, Wallace (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles Li (Ed.), *Subject and Topic*, pp. 25–55. New York: Academic Press.

Chafe, Wallace (1987). Cognitive constraints on information flow. In Russell S. Tomlin (Ed.), *Coherence and Grounding in Discourse*, pp. 21–52. Amsterdam, The Netherlands: John Benjamins.

Firbas, Jan (1974). Some aspects of the Czechoslovak approach to problems of functional sentence prespective. In F. Daneš (Ed.), *Papers on Functional Sentence Perspective*, pp. 11–37. Prague: Academia.

Foth, Kilian & Wolfgang Menzel (2006). Robust parsing: More with less. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 3–7 April 2006, pp. 25–32.

Frey, Werner (2004). A medial topic position for German. *Linguistische Berichte*, 198:153–190.

Gernsbacher, Morton A. & David J. Hargreaves (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27:699–717.

Givon, Talmy (1983). Topic continuity in spoken English. In T. Givon (Ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam, Philadelphia: John Benjamins.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Gurevych, Iryna (2005). Using the structure of a conceptual network in computing semantic relatedness. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing*, Jeju Island, South Korea, 11-13 October, 2005, pp. 767–778.

Heidolph, Karl Erich, Walter Flämig & Wolfgang Motsch (1981). *Grundzüge einer deutschen Grammatik*. Berlin: Akademie-Verlag.

Karamanis, Nikiforos, Massimo Poesio, Chris Mellish & Jon Oberlander (2004). Evaluating Centering-based metrics of coherence for text structuring using a reliably annotated corpus. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 21–26 July 2004, pp. 392–393.

Kruijff, Geert-Jan M., Ivana Kruijff-Korbayová, John Bateman & Elke Teich (2001). Linear Order as higher-level decision: Information Structure in strategic and tactical generation. In *8th European Workshop on Natural Language Generation*, Toulouse, France, July 6-7 2001, pp. 74–83.

Kruijff-Korbayová, Ivana, Geert-Jan Kruijff & John Bateman (2002). Generation of appropriate word order. In K. van Deemter & R. Kibble (Eds.), *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, pp. 193–222. Stanford: CSLI.

Lambrecht, Knud (1994). *Information Structure and Sentence Form*. Cambridge University Press.

Lapata, Maria (2003). Probabilistic text structuring: Experiments with sentence ordering. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 7–12 July 2003, pp. 545–552.

Levinson, Stephen C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Prince, Ellen F. (1999). How not to mark topics: 'Topicalization' in English and Yiddish. *Texas Linguistics Forum*.

Ratnaparkhi, Adwait (2000). Trainable methods for surface natural language generation. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, Wash., 29 April – 3 May, 2000, pp. 194–201.

Reinhart, Tanya (1981). Pragmatics and linguistics. An analysis of sentence topics. *Philosphica*, 27(1):53–93.

Schmid, Helmut (1997). Probabilistic part-of-speech tagging using decision trees. In Daniel Jones & Harold Somers (Eds.), *New Methods in Language Processing*, pp. 154–164. London, UK: UCL Press.

Sgall, Peter, Eva Hajičová & Jarmila Panevová (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: D. Reidel.

Strawson, Peter F. (1964). Identifying reference and truth-values. In D. Steinberg & L. Jacobovits (Eds.), *Semantics*, pp. 86–99. Cambridge: Cambridge University Press.

Strube, Michael & Simone Paolo Ponzetto (2006). WikiRelate! Computing semantic relatedness using Wikipedia. In *Proceedings of the 21st National Conference on Artificial Intelligence*, Boston, Mass., 16–20 July 2006. To appear.

Vallduví, Enric (1990). *The Informational Component*, (Ph.D. thesis). Philadelphia, Penn.: University of Pennsylvania, Department of Linguistics.

Webber, Bonnie, Matthew Stone, Aravind Joshi & Alistair Knott (2003). Anaphora and discourse structure. *Computational Linguistics*, 29(4):545–588.

Conceptual Coherence in the Generation of Referring Expressions *

Albert Gatt
Department of Computing Science
University of Aberdeen
agatt@csd.abdn.ac.uk

Kees van Deemter
Department of Computing Science
University of Aberdeen
kvdeemte@csd.abdn.ac.uk

Abstract

One of the challenges in the automatic generation of referring expressions is to identify a set of domain entities coherently, that is, from the same conceptual perspective. We describe and evaluate an algorithm that generates a conceptually coherent description of a target set. The design of the algorithm is motivated by the results of psycholinguistic experiments.

1 Introduction

Algorithms for the Generation of Referring Expressions (GRE) seek a set of properties that distinguish an intended referent from its distractors in a knowledge base. Much of the GRE literature has focused on developing efficient content determination strategies that output the best available description according to some interpretation of the Gricean maxims (Dale and Reiter, 1995), especially Brevity. Work on reference to sets has also proceeded within this general framework (van Deemter, 2002; Gardent, 2002; Horacek, 2004).

One problem that has not received much attention is that of *conceptual coherence* in the generation of plural references, i.e. the ascription of related properties to elements of a set, so that the resulting description constitutes a coherent cover for the plurality. As an example, consider a reference to $\{e_1, e_3\}$ in Table 1 using the Incremental Algorithm (IA) (Dale and Reiter, 1995). IA searches along an ordered list of attributes, selecting properties of the intended referents that remove some distractors. Assuming the ordering in the top row, IA would yield *the postgraduate and the chef*, which is fine in case **occupation** is the *relevant* attribute in the discourse, but otherwise is

	type	occupation	nationality
e_1	man	postgraduate	maltese
e_2	man	undergraduate	greek
e_3	man	chef	italian

Table 1: Example domain

arguably worse than an alternative like *the italian and the maltese*, because it is more difficult to see what a postgraduate and a chef have in common. Such examples lead us to hypothesise the following constraint:

> **Conceptual Coherence Constraint (CC):** As far as possible, describe objects using related properties.

Related issues have been raised in the formal semantics literature. Aloni (2002) argues that an appropriate answer to a question of the form *'Wh x?'* must conceptualise the different instantiations of *x* using a perspective which is relevant given the hearer's information state and the context. Kronfeld (1989) distinguishes a description's *functional relevance* – i.e. its success in distinguishing a referent – from its *conversational relevance*, which arises in part from implicatures. In our example, describing e_1 as *the postgraduate* carries the implicature that the entity's academic role is relevant. When two entities are described using contrasting properties, say *the student and the italian*, the contrast may be misleading for the listener.

Any attempt to port these observations to the GRE scenario must do so without sacrificing logical completeness. While a GRE algorithm should attempt to find the most coherent description available, it should not fail in the absence of a coherent set of properties. This paper aims to achieve a dual goal. First (§2), we will show that the CC can be explained and modelled in terms of lexical semantic forces within a description, a claim supported by the results of two experiments. Our

*This paper is a reprint, with minor revisions, of a paper that appears in the *Proceedings of the COLING-ACL 2006 Poster Session.*

focus on 'low-level', lexical, determinants of adequacy constitutes a departure from the standard Gricean view. Second, we describe an algorithm motivated by the experimental findings (§3) which seeks to find the most coherent description available in a domain according to CC.

2 Empirical evidence

We take as paradigmatic the case where a plural reference involves disjunction/union, that is, has the logical form $\lambda x\,(p(x) \vee q(x))$, realised as a description of the form *the* N_1 *and the* N_2. By hypothesis, the case where all referents can be described using identical properties (logically, a conjunction), is a limiting case of CC.

Previous work on plural anaphor processing has shown that pronoun resolution is easier when antecedents are ontologically similar (e.g. all humans) (Kaup et al., 2002; Koh and Clifton, 2002). Reference to a heterogeneous set increases processing difficulty.

Our experiments extended these findings to full definite NP reference. Throughout, we used a *distributional* definition of similarity, as defined by Lin (1998), which was found to be highly correlated to people's preferences for disjunctive descriptions (Gatt and van Deemter, 2005). The similarity of two arbitrary objects a and b is a function of the information gained by giving a joint description of a and b in terms of what they have in common, compared to describing a and b separately. The relevant data in the lexical domain is the grammatical environment in which words occur. This information is represented as a set of triples $\langle rel, w, w' \rangle$, where rel is a grammatical relation, w the word of interest and w' its co-argument in rel (e.g. $\langle$ *premodifies, dog, domestic* $\rangle$). Let $F(w)$ be a list of such triples. The information content of this set is defined as mutual information $I(F(w))$ (Church and Hanks, 1990). The similarity of two words w_1 and w_2, of the same grammatical category, is:

$$\sigma(w_1, w_2) = \frac{2 \times I(F(w_1) \cap F(w_2))}{I(F(w_1)) + I(F(w_2))} \quad (1)$$

For example, if *premodifies* is one of the relevant grammatical relations, then *dog* and *cat* might occur several times in a corpus with the same premodifiers (*tame*, *domestic*, etc). Thus, $\sigma(dog, cat)$ is large because in a corpus, they often occur in the same contexts and there is considerable information gain in a description of their common data.

Condition	a	b	c	distractor
HDS	spanner	chisel	plug	thimble
LDS	toothbrush	knife	ashtray	clock

Figure 1: Conditions in Experiment 1

Rather than using a hand-crafted ontology to infer similarity, this definition looks at real language use. It covers ontological similarity to the extent that ontologically similar objects are talked about in the same contexts, but also cuts across ontological distinctions (for example *newspaper* and *journalist* might turn out to be very similar).

We use the information contained in the SketchEngine database[1] (Kilgarriff, 2003), a largescale implementation of Lin's theory based on the BNC, which contains grammatical triples in the form of *Word Sketches* for each word, with each triple accompanied by a salience value indicating the likelihood of occurrence of the word with its argument in a grammatical relation. Each word also has a thesaurus entry, containing a ranked list of words of the same category, ordered by their similarity to the head word.

2.1 Experiment 1

In Experiment 1, participants were placed in a situation where they were buying objects from an online store. They saw scenarios containing four pictures of objects, three of which (the targets) were identically priced. Participants referred to them by completing a 2-sentence discourse:

S1 The *object1* and the *object 2* cost *amount*.

S2 The *object3* also costs *amount*.

If similarity is a constraint on referential coherence in plural references, then if two targets are similar (and dissimilar to the third), a plural reference to them in S1 should be more likely, with the third entity referred to in S2.

Materials, design and procedure All the pictures were artefacts selected from a set of drawings normed in a picture-naming task with British English speakers (Barry et al., 1997).

Each trial consisted of the four pictures arranged in an array on a screen. Of the three targets (a, b, c), c was always an object whose name in the norms was *dis*similar to that of a and b. The

[1] http://www.sketchengine.co.uk

semantic similarity of (nouns denoting) a and b was manipulated as a factor with two levels: **High Distributional Similarity (HDS)** meant that b occurred among the top 50 most similar items to a in its Sketchengine thesaurus entry. **Low DS (LDS))** meant that b did not occur in the top 500 entries for a. Examples are shown in Figure 2.1.

Visual Similarity (VS) of a and b was also controlled. Pairs of pictures were first normed with a group who rated them on a 10-point scale based on their visual properties. High-VS (HVS) pairs had a mean rating ≥ 6; Low-VS LVS) pairs had mean ratings ≤ 2. Two sets of materials were constructed, for a total of 2 $(DS) \times 2$ $(VS) \times 2 = 8$ trials.

29 self-reported native or fluent speakers of English completed the experiment over the web. To complete the sentences, participants clicked on the objects in the order they wished to refer to them. Nouns appeared in the next available space[2].

Results and discussion Responses were coded according to whether objects a and b were referred to in the plural subject of S1 ($a + b$ responses) or not ($a - b$ responses). If our hypothesis is correct, there should be a higher proportion of $a + b$ responses in the HDS condition. We did not expect an effect of VS. In what follows, we report by-subjects Friedman analyses (χ^2_1); by-items analyses (χ^2_2); and by-subjects sign tests (Z) on proportions of responses for pairwise comparisons.

Response frequencies across conditions differed reliably by subjects ($\chi^2_1 = 46.124, p < .001$). The frequency of $a + b$ responses in S1 was reliably higher than that of $a - b$ in the HDS condition ($\chi^2_2 = 41.371, p < .001$), but not the HVS condition ($\chi^2_2 = 1.755, ns$). Pairwise comparisons between HDS and LDS showed a significantly higher proportion of $a + b$ responses in the former ($Z = 4.48, p < .001$); the difference was barely significant across VS conditions ($Z = 1.9, p = .06$).

The results show that, given a clear choice of entities to refer to in a plurality, people are more likely to describe similar entities in a plural description. However, these results raise two further questions. First, given a choice of distinguishing properties for individuals making up a target set, will participants follow the predictions of the CC? (In other words, is distributional similarity relevant for content determination?) Second, does the similarity effect carry over to modifiers, such as adjectives, or is the CC exclusively a constraint on types?

[2]Earler replications involving typing yielded parallel results and high conformity between the words used and those predicted by the picture norms.

2.2 Experiment 2

Experiment 2 was a sentence continuation task, designed to closely approximate content determination in GRE. Participants saw a series of discourses, in which three entities (e_1, e_2, e_3) were introduced, each with two distinguishing properties. The final sentence in each discourse had a missing plural subject NP referring to two of these. The context made it clear which of the three entities had to be referred to. Our hypothesis was that participants would prefer to use semantically similar properties for the plural reference, *even if* dissimilar properties were also available.

Materials, design and procedure Materials consisted of 24 discourses, such as those in Figure 2.2. After an initial introductory sentence, the 3 entities were introduced in separate sentences. In all discourses, the pairs $\{e_1, e_2\}$ and $\{e_2, e_3\}$ could be described using either pairwise similar or dissimilar properties (similar pairs are coindexed in the figure). In half the discourses, the distinguishing properties of each entity were *nouns*; thus, although all three entities belonged to the same ontological category (e.g. all human), they had distinct types (e.g. *duke, prince, bachelor*). In the other half, entities were of the same type, that is the NPs introducing them had the same nominal head, but had distinguishing adjectival modifiers. For counterbalancing, two versions of each discourse were constructed, such that, if $\{e_1, e_2\}$ was the target set in Version 1, then $\{e_2, e_3\}$ was the target in Version 2. Twelve filler items requiring singular reference in the continuation were also included. The order in which the entities were introduced was randomised across participants, as was the order of trials. The experiment was completed by 18 native speakers of English, selected from the Aberdeen NLG Group database. They were randomly assigned to either Version 1 or 2.

Results and discussion Responses were coded 1 if the semantically similar properties were used (e.g. *the prince and the duke* in Fig. 2.2); 2 if the similar properties were used together with other properties (e.g. *the prince and the bachelor duke*); 3 if a superordinate term was used to replace the similar properties (e.g. *the noblemen*); 4 otherwise

	Three millionaires with a passion for antiques were spotted dining at a London restaurant.
e_1	One of the men, a Rumanian, is a dealer$_i$.
e_2	The second, a prince$_j$, is a collector$_i$.
e_3	The third, a duke$_j$, is a bachelor.
	The XXXX were both accompanied by servants, but the bachelor wasn't .

Figure 2: Example discourses

id	base type	occupation	specialisation	girth
e_1	woman	professor	physicist	plump
e_2	woman	lecturer	geologist	thin
e_3	man	lecturer	biologist	plump
e_4	man		chemist	thin

Table 2: An example knowledge base

(e.g. *The duke and the collector*).

Response types differed significantly in the nominal condition both by subjects ($\chi^2_1 = 45.89, p < .001$) and by items ($\chi^2_2 = 287.9, p < .001$). Differences were also reliable in the modifier condition ($\chi^2_1 = 36.3, p < .001$, $\chi^2_2 = 199.2, p < .001$). However, the trends across conditions were opposed, with more items in the 1 response category in the nominal condition (53.7%) and more in the 4 category in the modifier condition (47.2%). Recoding responses as binary ('similar' = 1,2,3; 'dissimilar' = 4) showed a significant difference in proportions for the nominal category ($\chi^2 = 4.78, p = .03$), but not the modifier category. Pairwise comparisons showed a significantly larger proportion of 1 ($Z = 2.7, p = .007$) and 2 responses ($Z = 2.54, p = .01$) in the nominal compared to the modifier condition.

The results suggest that in a referential task, participants are likely to conform to the CC, but that the CC operates mainly on nouns, and less so on (adjectival) modifiers. Nouns (or types, as we shall sometimes call them) have the function of categorising objects; thus similar types facilitate the mental representation of a plurality in a conceptually coherent way. According to the definition in (1), this is because similarity of two types implies a greater likelihood of their being used in the same predicate-argument structures. As a result, it is easier to map the elements of a plurality to a common role in a sentence. A related proposal has been made by Moxey and Sanford (1995), whose *Scenario Mapping Principle* holds that a plural reference is licensed to the extent that the elements of the plurality can be mapped to a common role in the discourse. This is influenced by how easy it is to conceive of such a role for the referents. Our results can be viewed as providing a handle on the notion of 'ease of conception of a common role'; in particular we propose that likelihood of occurrence in the same linguistic contexts directly reflects the extent to which two types can be mapped to a single plural role.

As regards modifiers, while it is probably premature to suggest that CC plays no role in modifier selection, it is likely that modifiers play a different role from nouns. Previous work has shown that restrictions on the plausibility of adjective-noun combinations exist (Lapata et al., 1999), and that using unlikely combinations (e.g. *the immaculate kitchen* rather than *the spotless kitchen*) impacts processing in online tasks (Murphy, 1984). Unlike types, which have a categorisation function, modifiers have the role of adding information about an element of a category. This would partially explain the experimental results: When elements of a plurality have identical types (as in the modifier version of our experiment), the CC is already satisfied, and selection of modifiers would presumably depend on respecting adjective-noun combination restrictions. Further research is required to verify this, although the algorithm presented below makes use of the Sketch Engine database to take modifier-noun combinations into account.

3 An algorithm for referring to sets

Our next task is to port the results to GRE. The main ingredient to achieve conceptual coherence will be the definition of semantic similarity. In what follows, all examples will be drawn from the domain in Table 3.

We make the following assumptions. There is a set U of domain entities, properties of which are specified in a KB as attribute-value pairs. We assume a distinction between *types*, that is, any property that can be realised as a noun; and *modifiers*, or non-types. Given a set of target referents $R \subseteq U$, the algorithm described below generates a description D in Disjunctive Normal Form (DNF), having the following properties:

1. Any disjunct in D contains a 'type' property, i.e. a property realisable as a head noun.
2. If D has two or more disjuncts, each a conjunction containing at least one type, then the disjoined types should be as similar as pos-

sible, given the information in the KB and the *completeness* requirement: that the algorithm find a distinguishing description whenever one exists.

We first make our interpretation of the CC more precise. Let T be the set of types in the KB, and let $\sigma(t, t')$ be the (symmetrical) similarity between any two types t and t'. These determine a semantic space $\mathbb{S} = \langle T, \sigma \rangle$. We define the notion of a perspective as follows.

Definition 1. Perspective
A perspective $\mathcal{P}$ is a convex subset of $\mathbb{S}$, i.e.:

$$\forall t, t', t'' \in T : \\ \{t, t'\} \subseteq \mathcal{P} \wedge \sigma(t, t'') \geq \sigma(t, t') \rightarrow t'' \in \mathcal{P}$$

The aims of the algorithm are to describe elements of R using types from the same perspective, failing which, it attempts to minimise the distance between the perspectives from which types are selected in the disjunctions of D. Distance between perspectives is defined below.

3.1 Finding perspectives

The system makes use of the SketchEngine database as its primary knowledge source. Since the definition of similarity applies to words, rather than properties, the first step is to generate all possible lexicalisations of the available attribute-value pairs in the domain. In this paper, we simplify by assuming a one-to-one mapping between properties and words.

Another requirement is to distinguish between type properties (the set T), and non-types (M)[3]. The Thesaurus is used to find pairwise similarity of types in order to group them into related clusters. Word Sketches are used to find, for each type, the modifiers in the KB that are appropriate to the type, on the basis of the associated salience values. For example, in Table 3, e_3 has *plump* as the value for **girth**, which combines more felicitously with *man*, than with *biologist*.

Types are clustered using the algorithm described in Gatt (2006). For each type t, the algorithm finds its nearest neighbour n_t in semantic space. Clusters are then found by recursively grouping elements with their nearest neighbours. If t, t' have a common nearest neighbour n, then $\{t, t', n\}$ is a cluster. Clearly, the resulting sets are

[3]This is determined using corpus-derived information. Note that T and M need not be disjoint, and entities can have more than one type property

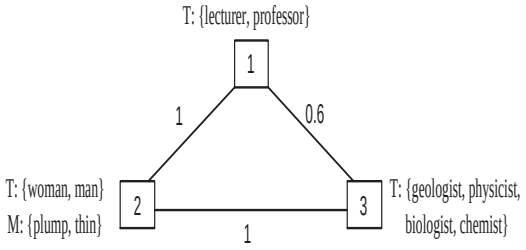


Figure 3: Perspective Graph

convex in the sense of Definition 1. Each modifier is assigned to a cluster by finding in its Word Sketch the type with which it co-occurs with the greatest salience value. Thus, a cluster is a pair $\langle \mathcal{P}, M' \rangle$ where $\mathcal{P}$ is a perspective, and $M' \subseteq M$. The distance $\delta(A, B)$ between two clusters A and B is defined straightforwardly in terms of the distance between their perspectives $\mathcal{P}_A$ and $\mathcal{P}_B$:

$$\delta(A, B) = \frac{1}{1 + \frac{\sum_{x \in \mathcal{P}_A, y \in \mathcal{P}_B} \sigma(x,y)}{|\mathcal{P}_A \times \mathcal{P}_B|}} \qquad (2)$$

Finally, a weighted, connected graph $\mathcal{G} = \langle V, E, \delta \rangle$ is created, where V is the set of clusters, and E is the set of edges with edge weights defined as the semantic distance between perspectives. Figure 3.1 shows the graph constructed for the domain in Table 3.

We now define the coherence of a description more precisely. Given a DNF description D, we shall say that a perspective $\mathcal{P}$ is *realised in* D if there is at least one type $t \in \mathcal{P}$ which is in D. Let $\mathbb{P}_D$ be the set of perspectives realised in D. Since $\mathcal{G}$ is connected, $\mathbb{P}_D$ determines a connected subgraph of $\mathcal{G}$. The *total weight* of D, $w(D)$ is the sum of weights of the edges in $\mathbb{P}_D$.

Definition 2. Maximal coherence
A description D is *maximally coherent* iff there is no description D' coextensive with D such that $w(D) > w(D')$.

(Note that several descriptions of the same referent may all be maximally coherent.)

3.2 Content determination

The core of the content determination procedure maintains the DNF description D as an associative array, such that for any $r \in R$, $D[r]$ is a conjunction of properties true of r. Given a cluster $\langle \mathcal{P}, M \rangle$, the procedure searches incrementally first through $\mathcal{P}$, and then M, selecting properties that

are true of at least one referent and exclude some distractors, as in the IA (Dale and Reiter, 1995).

By Definition 2, the task of the algorithm is to minimise the total weight $w(D)$. If $\mathbb{P}_D$ is the set of perspectives represented in D on termination, then *maximal coherence* would require $\mathbb{P}_D$ to be the subgraph of $\mathcal{G}$ with the lowest total cost from which a distinguishing description could be constructed. Under this interpretation, $\mathbb{P}_D$ corresponds to a Shortest Connection, or Steiner, Network. Finding such networks is known to be NP-Hard. Therefore, we adopt a weaker (greedy) interpretation. Under the new definition, if D is the only description for R, then it trivially satisfies maximal coherence. Otherwise, the algorithm aims to maximise *local coherence*.

Definition 3. Local coherence
A description D is *locally coherent* iff:

a. **either** D is maximally coherent **or**

b. there is no D' coextensive with D, obtained by replacing types from some perspective in $\mathbb{P}_D$ with types from another perspective such that $w(D) > w(D')$.

Our implementation of this idea begins the search for distinguishing properties by identifying the vertex of $\mathcal{G}$ which contains the greatest number of referents in its extension. This constitutes the root node of the search path. For each node of the graph it visits, the algorithm searches for properties that are true of some subset of R, and removes some distractors, maintaining a set N of the perspectives which are represented in D up to the current point. The crucial choice points arise when a new node (perspective) needs to be visited in the graph. At each such point, the next node n to be visited is the one which minimises the total weight of N, that is:

$$\min_{n \in V} \sum_{u \in N} w(u, n) \tag{3}$$

The results of this procedure closely approximate maximal coherence, because the algorithm starts with the vertex most likely to distinguish the referents, and then greedily proceeds to those nodes which minimise $w(D)$ given the current state, that is, taking all previously used nodes into account.

As an example of the output, we will take $R = \{e_1, e_3, e_4\}$ as the intended referents in Table 3. First, the algorithm determines the cluster with the greatest number of referents in its extension. In this case, there is a tie between clusters 2 and 3 in Figure 3.1, since all three entities have type properties in these clusters. In either case, the entities are distinguishable from a single cluster. If cluster 3 is selected as the root, the output is $\lambda x\,[physicist(x) \vee biologist(x) \vee chemist(x)]$. In case the algorithm selects cluster 2 as the root node the final output is the logical form $\lambda x\,[man(x) \vee (woman(x) \wedge plump(x))]$.

There is an alternative description that the algorithm does not consider. An algorithm that aimed for conciseness would generate $\lambda x\,[professor(x) \vee man(x)]$ (*the professor and the men*), which does not satisfy local coherence. These examples therefore highlight the possible tension between the avoidance of redundancy and achieving coherence. It is to an investigation of this tension that we now turn.

4 Evaluation

It has been known at least since Dale and Reiter (1995) that the best distinguishing description is not always the shortest one. Yet, brevity plays a part in all GRE algorithms, sometimes in a strict form (Dale, 1989), or by letting the algorithm *approximate* the shortest description (for example, in the Dale and Reiter's IA). This is also true of references to sets, the clearest example being Gardent's constraint based approach, which always finds the description with the smallest number of logical operators. Such proposals do not take coherence (in our sense of the word) into account. This raises obvious questions about the relative importance of brevity and coherence in reference to sets.

The evaluation took the form of an experiment to compare the output of our *Coherence Model* with the family of algorithms that have placed Brevity at the centre of content determination. Participants were asked to compare pairs of descriptions of one and the same target set, selecting the one they found most natural. Each description could either be optimally brief or not ($\pm b$) and also either optimally coherent or not ($\pm c$). Non-brief descriptions, took the form *the A, the B and the C*. Brief descriptions 'aggregated' two disjuncts into one (e.g. *the A and the D's* where D comprises the union of B and C). We expected to find that:

H1 $+c$ descriptions are preferred over $-c$.

H2 $(+c, -b)$ descriptions are preferred over ones that are $(-c, +b)$.

	Three old manuscripts were auctioned at Sotheby's.
e_1	One of them is a book, a biography of a composer.
e_2	The second, a sailor's journal, was published in the form of a pamphlet. It is a record of a voyage.
e_3	The third, another pamphlet, is an essay by Hume.
$(+c, -b)$	The biography, the journal and the essay were sold to a collector.
$(+c, +b)$	The book and the pamphlets were sold to a collector.
$(-c, +b)$	The biography and the pamphlets were sold to a collector.
$(-c, -b)$	The book, the record and the essay were sold to a collector.

Figure 4: Example domain in the evaluation

H3 $+b$ descriptions are preferred over $-b$.

Confirmation of H1 would be interpreted as evidence that, by taking coherence into account, our algorithm is on the right track. If H3 were confirmed, then earlier algorithms were (also) on the right track by taking brevity into account. Confirmation of H2 would be interpreted as meaning that, in references to sets, conceptual coherence is more important than brevity (defined as the number of disjuncts in a disjunctive reference to a set).

Materials, design and procedure Six discourses were constructed, each introducing three entities. Each set of three could be described using all 4 possible combinations of $\pm b \times \pm c$ (see Figure 4). Entities were human in two of the discourses, and artefacts of various kinds in the remainder. Properties of entities were introduced textually; the order of presentation was randomised. A forced-choice task was used. Each discourse was presented with 2 possible continuations consisting of a sentence with a plural subject NP, and participants were asked to indicate the one they found most natural. The 6 comparisons corresponded to 6 sub-conditions:

C1. **Coherence constant**
  a. $(+c, -b)$ vs. $(+c, +b)$
  b. $(-c, -b)$ vs. $(-c, +b)$

C2. **Brevity constant**
  a. $(+c, -b)$ vs. $(-c, -b)$
  b. $(+c, +b)$ vs. $(-c, +b)$

C3. **Tradeoff/control**
  a. $(+c, -b)$ vs. $(-c, +b)$
  b. $(-c, -b)$ vs. $(+c, +b)$

Participants saw each discourse in a single condition. They were randomly divided into six groups, so that each discourse was used for a different condition in each group. 39 native English speakers, all undergraduates at the University of Aberdeen, took part in the study.

	C1a	C1b	C2a	C2b	C3a	C3b
$+b$	51.3	43.6	–	–	30.8	76.9
$+c$	–	–	82.1	79.5	69.2	76.9

Table 3: Response proportions (%)

Results and discussion Results were coded according to whether a participant's choice was $\pm b$ and/or $\pm c$. Table 4 displays response proportions. Overall, the conditions had a significant impact on responses, both by subjects (Friedman $\chi^2 = 107.3, p < .001$) and by items ($\chi^2 = 30.2, p < .001$). When coherence was kept constant (C1a and C1b), the likelihood of a response being $+b$ was no different from $-b$ (C1a: $\chi^2 = .023, p = .8$; C1b: $\chi^2 = .64, p = .4$); the conditions C1a and C1b did not differ significantly ($\chi^2 = .46, p = .5$). By contrast, conditions where brevity was kept constant (C2a and C2b) resulted in very significantly higher proportions of $+c$ choices (C2a: $\chi^2 = 16.03, p < .001$; C2b: $\chi^2 = 13.56, p < .001$). No difference was observed between C2a and C2b ($\chi^2 = .08, p = .8$). In the tradeoff case (C3a), participants were much more likely to select a $+c$ description than a $+b$ one ($\chi^2 = 39.0, p < .001$); a majority opted for the $(+b, +c)$ description in the control case ($\chi^2 = 39.0, p < .001$).

The results strongly support H1 and H2, since participants' choices are impacted by Coherence. They do not indicate a preference for brief descriptions, a finding that echoes Jordan's (2000), to the effect that speakers often relinquish brevity in favour of observing task or discourse constraints. Since this experiment compared our algorithm against the current state of the art in references to sets, these results do not necessarily warrant the affirmation of the null hypothesis in the case of H3. We limited Brevity to number of disjuncts, omitting negation, and varying only between length 2 or 3. Longer or more complex descriptions might evince different tendencies. Nevertheless, the results show a strong impact of Coherence, compared to (a kind of) brevity, in strong support of the algorithm presented above, as a realisation of the Coherence Model.

5 Conclusions and future work

This paper started with an empirical investigation of conceptual coherence in reference, which led

to a definition of *local* coherence as the basis for a new greedy algorithm that tries to minimise the semantic distance between the perspectives represented in a description. The evaluation strongly supports our Coherence Model.

We are extending this work in two directions. First, we are investigating similarity effects *across noun phrases*, and their impact on text readability. Finding an impact of such factors would make this model a useful complement to current theories of discourse, which usually interpret coherence in terms of discourse/sentential structure.

Second, we intend to relinquish the assumption of a one-to-one correspondence between properties and words (cf. Siddharthan and Copestake (2004)), making use of the fact that words can be disambiguated by nearby words that are similar. To use a well-worn example: the 'financial institution' sense of *bank* might not make *the river and its bank* lexically incoherent as a description of a piece of scenery, since the word *river* might cause the hearer to focus on the aquatic reading of the word anyway.

6 Acknowledgements

Thanks to Ielka van der Sluis, Imtiaz Khan, Ehud Reiter, Chris Mellish, Graeme Ritchie and Judith Masthoff for useful comments. This work is part of the TUNA project (`http://www.csd.abdn.ac.uk/research/tuna`), supported by EPSRC grant no. GR/S13330/01

References

M. Aloni. 2002. Questions under cover. In D. Barker-Plummer, D. Beaver, J. van Benthem, and P. Scotto de Luzio, editors, *Words, Proofs, and Diagrams*. CSLI, Stanford, Ca.

C. Barry, C. M. Morrison, and A. W. Ellis. 1997. Naming the snodgrass and vanderwart pictures. *Quarterly Journal of Experimental Psychology*, 50A(3):560–585.

K. W. Church and P. Hanks. 1990. Word association norms, mutual information and lexicography. *Computational Linguistics*, 16(1):22–29.

R. Dale and E. Reiter. 1995. Computational interpretation of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(8):233–263.

Robert Dale. 1989. Cooking up referring expressions. In *Proc. 27th Annual Meeting of the Association for Computational Linguistics.*

C. Gardent. 2002. Generating minimal definite descriptions. In *Proc. 40th Annual Meeting of the Association for Computational Linguistics.*

A. Gatt and K. van Deemter. 2005. Semantic similarity and the generation of referring expressions: A first report. In *Proc. 6th International Workshop on Computational Semantics, IWCS-6.*

A. Gatt. 2006. Structuring knowledge for reference generation: A clustering algorithm. In *Proc. 11th Conference of the European Chapter of the Association for Computational Linguistics.*

H. Horacek. 2004. On referring to sets of objects naturally. In *Proc. 3rd International Conference on Natural Language Generation.*

P. W. Jordan. 2000. Can nominal expressions achieve multiple goals? In *Proc. of the 38th Annual Meeting of the Association for Computational Linguistics*.

B. Kaup, S. Kelter, and C. Habel. 2002. Representing referents of plural expressions and resolving plural anaphors. *Language and Cognitive Processes*, 17(4):405–450.

A. Kilgarriff. 2003. Thesauruses for natural language processing. In *Proc. NLP-KE, Beijing.*

S. Koh and C. Clifton. 2002. Resolution of the antecedent of a plural pronoun: Ontological categories and predicate symmetry. *Journal of Memory and Language*, 46:830–844.

A. Kronfeld. 1989. Conversationally relevant descriptions. In *Proc. 27th Annual Meeting of the Association for Computational Linguistics.*

M. Lapata, S. McDonald, and F. Keller. 1999. Determinants of adjective-noun plausibility. In *Proc. 9th Conference of the European Chapter of the Association for Computational Linguistics.*

D. Lin. 1998. An information-theoretic definition of similarity. In *Proc. International Conference on Machine Learning.*

L. Moxey and A. Sanford. 1995. Notes on plural reference and the scenario-mapping principle in comprehension. In C.Habel and G.Rickheit, editors, *Focus and cohesion in discourse.* de Gruyter, Berlin.

G.L. Murphy. 1984. Establishing and accessing referents in discourse. *Memory and Cognition*, 12:489–497.

A. Siddharthan and A. Copestake. 2004. Generating referring expressions in open domains. In *Proc. 42nd Annual Meeting of the Association for Computational Linguistics.*

K. van Deemter. 2002. Generating referring expressions: Boolean extensions of the incremental algorithm. *Computational Linguistics*, 28(1):37–52.

Entity versus Rhetorical Coherence for Information Ordering: Initial Experimentation

Nikiforos Karamanis
Natural Language and Information Processing Group
Computer Laboratory
University of Cambridge
Nikiforos.Karamanis@cl.cam.ac.uk

Abstract

This paper investigates whether the model of local rhetorical coherence suggested in Knott et al. (2001) can boost the performance of the Centering-based metrics of entity coherence employed by Karamanis et al. (2004) for the task of information ordering. Our results indicate that (a) the simplest metric continues to perform better than its competitors even when local rhetorical coherence is taken into account, and (b) this extra coherence constraint decreases its performance.

1 Introduction

As most literature in text linguistics argues, a felicitous text should be *coherent* which means that the content has to be organised in a way that makes the text easy to read and comprehend. The easiest way to demonstrate this claim is by arbitrarily reordering the sentences that an understandable text consists of. This process very often gives rise to documents that do not make sense although the information content is the same before and after the reordering. Hence, *information ordering* (Barzilay and Lee, 2004), i.e. deciding in which sequence to present a set of preselected information-bearing items (typically corresponding to clauses or sentences) is an important problem in automatic text production.

Entity coherence, which arises from the way NP referents relate subsequent clauses in the text, is an important aspect of textual felicity. *Centering Theory* (Grosz et al., 1995) has been an influential framework for modelling entity coherence in computational linguistics in the last two decades. Karamanis et al. (2004) were the first to evaluate Centering-based metrics of coherence for ordering clauses in a subset of the GNOME corpus (Poesio et al., 2004) which is reliably annotated with features related to Centering. Their test data consisted of descriptions of museum artefacts since Centering was expected to be particularly appropriate for information ordering in this genre.

Karamanis et al. assume a system, similar to the one discussed e.g. in Lapata (2003), which receives an unordered set of clauses as its input and uses a metric to output the highest scoring ordering of these clauses. They introduced a novel experimental methodology that treats the observed ordering of clauses in a text as the gold standard, which is scored by each metric. Then, the metric is penalised proportionally to the amount of alternative orderings of the same material that score equally to or better than the gold standard. This methodology is very similar to the way Barzilay and Lee (2004) and Barzilay and Lapata (2005) evaluate automatically their information ordering approach.

The main finding of Karamanis et al. was that the simplest metric (and most remote to Centering) sets a baseline that cannot be overtaken by other metrics which utilise additional Centering-specific notions. However, the baseline did not perform well enough to be used in practice for information ordering on its own.

This paper investigates whether the model of *local rhetorical coherence* in Knott et al. (2001) can boost the performance of the metrics of Karamanis et al. Our results indicate that (a) the baseline remains the best performing metric when compared to its competitors even when local rhetorical coherence is taken into account, and (b) supplementing the baseline with this extra coherence constraint decreases its performance.

Unit	CF list: {CP, next referent}	CB	Transition	CHEAPNESS CB_n=CP_{n-1}
(1a)	{de374, de375}	n.a.	n.a.	n.a.
(1b)	{de376, de374, ... }	de374	RETAIN	OK
(1c)	{de374, de379, ... }	de374	CONTINUE	*
(1d)	{de380, de381, ... }	–	NOCB	n.a.

Table 1: The CP (i.e. first member of the CF list), the next referent, the CB, NOCB or Centering transition (Table 2) and violations of CHEAPNESS (denoted with an asterisk) for each unit in example (1) from the GNOME-LAB corpus.

		COHERENCE: CB_n=CB_{n-1} or NOCB in CF_{n-1}	COHERENCE*: $CB_n \neq CB_{n-1}$
SALIENCE:	CB_n=CP_n	CONTINUE	SMOOTH-SHIFT
SALIENCE*:	$CB_n \neq CP_n$	RETAIN	ROUGH-SHIFT

Table 2: COHERENCE, SALIENCE and the table of Centering transitions.

The paper is structured as follows: First we present the annotation features of GNOME which are relevant to our study and discuss how local rhetorical coherence can be taken into account in our domain. After a brief presentation of the Centering-based metrics of coherence and our experimental methodology, we present the results of our study and discuss their implications. The paper is concluded with an outline of our related and future work.

2 Centering data structures in GNOME

Our experimental domain is GNOME-LAB, a subset of the GNOME corpus consisting of 20 museum labels as identified by Karamanis et al. The following example cites a characteristic text from this corpus:

(1) (a) [Item 144]$_S$ is a torc. (b) [Its present arrangement]$_S$, twisted into three rings, may be a modern alteration; (c) [it]$_S$ should probably be a single ring, worn around the neck. (d) [The terminals]$_S$ are in the form of goats' heads.

The text spans with indexes (a) to (d) correspond to annotated *finite units* in GNOME. Karamanis et al. used the computational tools of Poesio et al. to automatically derive from these units the basic data structures of Centering (known as the CF lists), in which referents of NPs such as de374 (that is, the referent of "Item 144") are ranked according to their prominence (see Table 1).

More specifically and following Brennan et al. (1987), the referent of the NP which bears the grammatical role of the subject (indicated with the subscript S in the example) is defined as the first member of the CF list (called the CP). Referents with the same grammatical role are ranked according to the linear order of the corresponding NPs in the text.

This way of computing the CF list is very commonly used in Centering and particularly appropriate for the information ordering approach assumed by Karamanis et al. who take the finite units to correspond to the database facts typically employed in concept-to-text generation systems such as MPIRO (Isard et al., 2003) and the referents of the NPs to accord to the arguments of those facts.[1]

The derived sequence of CF lists is then used to compute other important Centering concepts:

- The CB (Grosz et al., 1995), i.e. the referent that links the current CF list with the previous one such as de374 in (1b). The CB is defined by Centering's Constraint 3 as the highest ranked member of the current CF list which also appears in the previous CF list.

- NOCBs, that is, cases in which two subsequent CF lists do not have any referent

[1]Other information ordering approaches such as the one presented by Barzilay and Lapata (2005) operate on sentences, i.e. less granulated units, and are more appropriate for text-to-text production.

in common as in (1d).[2]

- Transitions (Brennan et al., 1987), exemplified in Table 2, and the preferences between them (known as Centering's Rule 2): CONTINUE is preferred to RETAIN, which is preferred to SMOOTH-SHIFT, which is preferred to ROUGH-SHIFT.

- The decomposition of transitions into the principles of COHERENCE and SALIENCE by Kibble and Power (2000), also exemplified in Table 2. We take account of the principle of CHEAPNESS as well (Strube and Hahn, 1999): see last column of Table 1.

3 Local rhetorical coherence

Not taking other coherence-inducing factors into account is quite common in most Centering-based studies, especially in text interpretation: see e.g. the collection of papers in Walker et al. (1998). Recently, Kibble (2001) argued that Centering needs to be supplemented with other models of coherence while Poesio et al. (2004) suggested that the model of local rhetorical coherence introduced by Knott et al. (2001) may be a good candidate to supplement Centering in our domain of interest.

The main claim of Knott et al. (2001) is that entity coherence in descriptive texts is supplemented by trees of *rhetorical relations* (Mann and Thompson, 1987) which apply **locally**, that is, among adjacent clauses, for instance:

(2) (a) Access to the cartonnier's lower half can only be gained by the doors at the sides, (b) because the table would have blocked the front.

In the representation of Karamanis et al., the units of this example give rise to two CF lists:

(3) CF list of (2a): {de12, de13, ... }
CF list of (2b): {de9, de18}

The local tree of rhetorical relations (RR-tree) in example (2) corresponds to an annotated *sentence* in GNOME, that is, a span of text ending with a full stop, a question mark or an exclamation point. Hence, the CF list of the RR-tree can be readily computed using *sentence* instead of *finite unit* (keeping the other Centering parameters such as the ranking of referents the same as in Karamanis et al.). The first two members of the CF list for the sentence that contains (2a) and (2b) are shown in example (4):

(4) Access to the cartonnier's lower half can only be gained by the doors at the sides, because the table would have blocked the front.
CF list of (4): {de12, de9, ... }

The CF list of (4) replaces the CF lists of (2a) and (2b) in the data structures we define.

Using a cue phrase such as "because", "but", "although", etc. as the signal for a local RR, we identified 19 local RR-trees in 12 texts from GNOME-LAB.[3] Those 12 texts form our subcorpus GNOME-RR. The remaining 8 texts are similar to example (1) in that they do not feature any rhetorical relation other than ELABORATION which is replaced by entity coherence in the model of Knott et al. (2001).

The sentence in (4) contains two NPs annotated as subjects: "Access to the cartonnier's lower half" (whose referent is de12) and "the table" (whose referent is de9). The CP of (4) is de12 because "Access to the cartonnier's lower half" precedes "the table" within the sentence. If (2b) preceded (2a) within the sentence, the CP of (4) would have been de9. Thus, we assume that the order of the units that a local rhetorical tree consists of is defined *before* the information ordering process that the Centering-based metrics are tested to be suitable for.

In all but one case, the finite units that are related with each other via a RR appear within the same sentence which consists only of these units as in example (2) and can be computed automatically. The CF list of the sole RR-tree consisting of two finite units each forming a single sentence was computed by hand, using the surface order of the sentences for the ranking of referents with the same grammatical role.

Despite the isomorphism between RR-trees and sentences in GNOME-RR, it would be a mistake to consider the relationship between sentences

[2] In order not to violate the assumption that referents corespond to arguments of database facts, Karamanis et al. ignored the annotated bridging relation (Clark, 1977) between the referent of "the terminals" de380 in (1d) and the referent of "it" de374 in (1c), by virtue of which de374 might be thought as being a member of the CF list of (1d).

[3] Although we acknowledge that cue phrases are not the only hint for a RR, it has been shown that they constitute a very reliable way of detecting one (Knott and Dale, 1994).

consisting of more than one finite unit and RR-trees as 1:1. We identified 15 sentences in GNOME-LAB consisting of more than one finite unit which are not related to each other via an explicit RR marked with a connective although they appear within the same sentence (units (1b) and (1c) are one such case). These units are represented as subsequent, rhetorically unrelated, CF lists.

Note that taking local RR-trees into account as just explained reduces the overall number of CF lists a text is analysed to (and the corresponding number of possible orderings). More specifically, the texts in GNOME-RR contain 1.58 fewer CF lists when compared to the average number of CF lists in GNOME-LAB (8.35).

4 Metrics of coherence

Karamanis (2003) discusses how Centering can be used to define many different metrics of coherence which might be useful for information ordering. In our experiments we made use of the four metrics employed in Karamanis et al. (2004):

- The baseline metric M.NOCB which prefers the ordering with the fewest NOCBs.
- M.CHEAP which selects the ordering with the fewest violations of CHEAPNESS.
- M.KP, introduced by Kibble and Power (2000), which sums up the NOCBs as well as the violations of CHEAPNESS, COHERENCE and SALIENCE, preferring the ordering with the lowest total cost.[4]
- M.BFP which employs the transition preferences of Brennan et al. (1987).

5 Experimental methodology

As already mentioned, our evaluation methodology is based on the premise that the gold standard ordering (GSO) of the clauses (or the corresponding CF lists) observed in a text is more coherent than any other ordering. If a metric takes a randomly produced ordering to be more coherent than the GSO, it has to be penalised.

[3]A more detailed and general study on the lack of isomorphism between document and rhetorical structure, motivated mainly by examples from GNOME's pharmaceutical section appears in Power et al. (2003).

[4]A more recent variant of this metric appears in Kibble and Power (2004).

GNOME-RR	M.NOCB			p
corpus	lower	greater	ties	
M.CHEAP	10	2	0	0.038
M.KP	11	1	0	0.006
M.BFP	7	5	0	0.774
N of texts	12			

Table 3: Comparing M.NOCB with M.CHEAP, M.KP and M.BFP in GNOME-RR.

Karamanis et al. (2004) introduce a measure called the *classification rate* which estimates this penalty as the weighted sum of the percentage of alternative orderings that score equally to or better than the GSO.[5] When comparing several metrics with each other, the one with the lowest classification rate is the most appropriate for ordering the CF lists that the GSO consists of.

In this study, we use the classification rate to measure the performance of the metrics and investigate the following questions: (a) Is the best performing metric in GNOME-RR different from the one in GNOME-LAB? (b) Does taking local RR-trees into account improve the performance of the metrics?

6 Results

6.1 Which is the best metric?

The experimental results of the comparisons of the metrics from section 4 are reported in Table 3. Following Karamanis et al. (2004), the tables compare the baseline metric M.NOCB with each of M.CHEAP, M.KP and M.BFP. The exact number of GSOs for which the classification rate of M.NOCB is lower (i.e. better) than its competitor for each comparison is reported in the second column of the Table.

For example, M.NOCB has a lower classification rate than M.CHEAP for 10 (out of 12) GSOs from GNOME-RR. M.CHEAP achieves a lower classification rate for just 2 GSOs, while there are no ties, i.e. cases in which the classification rate of the two metrics is the same. The p value returned by the two-tailed sign test for the difference in the number of GSOs, rounded to the third decimal place, is reported in

[5]The classification rate is computed according to the formula Better(M,GSO) + Equal(M,GSO)/2. Better(M,GSO) stands for the percentage of orderings that score better than the GSO according to a metric M, whilst Equal(M,GSO) is the percentage of orderings that score equal to the GSO.

GNOME-LAB	M.NOCB			p
corpus	lower	greater	ties	
M.CHEAP	18	2	0	0.000
M.KP	16	2	2	0.002
M.BFP	12	3	5	0.036
N of texts	20			

Table 4: Comparing M.NOCB with M.CHEAP, M.KP and M.BFP in GNOME-LAB.

	GNOME-RR			p
metric	lower	greater	ties	
M.NOCB	3	9	0	0.146
M.CHEAP	9	3	0	0.146
M.KP	10	2	2	0.038
M.BFP	5	7	5	0.774
N of texts	12			

Table 5: Changes in the classification rate of the metrics in GNOME-RR.

the fifth column of Table 3.[6]

Overall, the Table shows that M.NOCB does significantly better than M.CHEAP and M.KP in GNOME-RR. Since M.BFP fails to significantly overtake M.NOCB, the baseline can be considered the most promising solution in that case too by applying Occam's razor. This in turn indicates that simply avoiding NOCB transitions is more relevant to information ordering than the various combinations of the Centering notions that the other metrics make use of.

Table 4 shows the results of the evaluation of the metrics in GNOME-LAB from Karamanis et al. (2004) which are very similar to the ones just reported in that the baseline overtakes the three other metrics which employ additional Centering concepts. Hence, M.NOCB is the most suitable among the investigated metrics for information ordering in this domain irrespective of whether local RR-trees are taken into account for the computation of the CF list.

6.2 Does rhetorical coherence help?

We were also interested to see whether taking RR-trees into account improves the performance of the metrics. To do this we compared the classification rates of the 12 GSOs for each metric in GNOME-RR with the corresponding classification rates in GNOME-LAB (Table 5). The Table suggests that using local RR-trees for the computation of the CF list lowers (i.e. improves) the performance of M.CHEAP as well as M.KP (for which the difference is significant). However, since both these metrics are defeated overwhelmingly by M.NOCB (see previous section), this improvement seems to be of little use.

[6]The sign test was chosen by Karamanis et al. (2004) over its parametric alternatives to test significance because it does not carry specific assumptions about population distributions and variance and is more appropriate for small sample sizes.

Notably, M.NOCB continues to beat its opponents despite the fact that its own classification rates are increased (i.e. worsened) in 9 out of 12 GSOs. This observation is coupled by the value of the *average classification rate* of M.NOCB which is an estimate of how likely M.NOCB is to come up with the GSO if it is actually used to guide an algorithm which orders the CF lists in our corpora. The average classification rate in GNOME-RR is 23.24%, which means that on average M.NOCB takes approximately 1 out of 4 alternative orderings in GNOME-RR to be more coherent than the GSO. This compares poorly with the value of 19.95% in GNOME-LAB and suggests that RRs do not help M.NOCB become more efficient for information ordering in the investigated domain. In the following section we discuss the implications of our experimental results.

7 Discussion

First, we would like to point the attention of the reader to the fact that several effects *are* strong enough to reject the null hypothesis on the basis of statistical tests, despite of the small size of the employed samples. Hence, this simple preliminary study on combining certain aspects of entity and rhetorical coherence for information ordering enables several interesting observations to be made.

As already pointed out in Karamanis et al. (2004), the results suggest that if one is provided with the set of CF lists from a GSO in the domain of interest and has to choose which of the four candidate metrics to use to order them (aiming to arrive at the GSO as the output), the baseline M.NOCB is a better choice than M.KP, M.CHEAP and M.BFP. This is because there exist proportionally fewer alternative orderings that are taken to be more coherent than the GSO according

to M.NOCB in comparison to the coherence assessments made by the other metrics.

Avoiding NOCBs is hardly a Centering-specific requirement and is typically seen as just a prerequisite for computing other more Centering-related notions. Our work shows that NOCBs are much more useful than these notions as far as information ordering is concerned. Of course, concepts such as CHEAPNESS remain very important for other tasks such as anaphora resolution.

Our empirical results indicate that the predominance of M.NOCB holds irrespective of whether local RR-trees are taken into account for the computation of the CF lists. This renders M.NOCB as a very robust baseline against which other, perhaps even more informed, metrics may be compared.

Poesio et al. (2004) report that dispreferred transitions such as NOCBs are very frequent in GNOME, which leads them to the conclusion that Centering needs to be supplemented with another models of coherence such as the one suggested by Knott et al. Using hybrid models of entity and rhetorical coherence is also favoured by most text generation practitioners such as Kibble and Power.

The majority of transitions in both GNOME-LAB and GNOME-RR (57% and 53% respectively) are NOCBs. This accords with the findings of Poesio et al. and might cause one to think that entity coherence has indeed little to do with the investigated domain.

However, using the classification rate to estimate the effect of entity coherence in this domain sheds new light into the issue. The average classification rate of M.NOCB is approximately 20% in GNOME-LAB and 23% in GNOME-RR. This suggests that the GSO tends to be in greater agreement with the preference to avoid NOCBs that the overwhelming majority (i.e. 80% in GNOME-LAB and 77% in GNOME-RR) of alternative orderings. In this sense, it seems that the observed ordering in the corpus (that is, the GSO) does optimise with respect to the number of potential NOCBs to a great extent. This is not obvious if the effect of entity coherence is estimated simply on the basis of the transition frequencies as it has been done until now.

Since the number of possible orderings becomes smaller when local RR-trees are taken into account, the information ordering problem is somewhat simplified. One might also be tempted to think that since GNOME-RR contains 4% fewer NOCB transitions than GNOME-LAB, computing the CF list using local RR-trees should be preferred.

However, the 3% rise in the aforementioned classification rates provides evidence that there exist proportionally more orderings which are taken to be more coherent than the GSO in GNOME-RR than in GNOME-LAB. Thus, taking local RR-trees into account does not help M.NOCB improve its performance. This in turn indicates that a solution based on the model of Knott et al. is not particularly helpful, at least as far as information ordering in this domain is concerned.

Overall, our empirical results clarify which aspects of entity and local rhetorical coherence are more relevant to information ordering and puts other related work into perspective. Our experiments also provide researchers working in information ordering with a simple and easily extendable evaluation framework as well as a robust baseline to deploy for their own meaningful comparisons.

8 Related and future work

In related work, we applied the methodology discussed here to several additional domains: (a) 122 orderings of facts derived from the MPIRO generation system by Dimitromanolaki and Androutsopoulos (2003) and ordered by a domain expert and (b) 200 newspaper articles and 200 accident narratives collected by Barzilay and Lapata (2005). The results from these domains verify the ones reported here with the baseline overwhelmingly beating its competitors (Karamanis, 2003; Karamanis, 2006).

The enrichment of the deployed metrics with additional constraints of coherence remains the biggest challenge for the work reported in this paper. Initial results indicate that making use of features related to global focus in GNOME-LAB has the same effect as local RR-trees, i.e. they increase – instead of reduce – the classification rate of the metrics (Karamanis, 2003).

Given the abundance of possible Centering-based metrics and several different ways of instantiating Centering, one might be keen to investigate whether a different metric might overtake M.NOCB or whether using bridging

for the computation of the CF list affects its performance (also when local RR-trees are used).

Last but not least, the evaluation in this paper is based on purely corpus-based methods. These should ideally be supplemented with human judgments in the spirit of the work reported by Reiter and Sripada (2002) and Barzilay and Lapata (2005).

Acknowledgments

Many thanks to Massimo Poesio, Chris Mellish and Jon Oberlander for their invaluable guidance and advice, to James Soutter for significant programming assistance and to three anonymous reviewers for their comments. Support from the Greek State Scholarships Foundation (IKY) and the BBSRC-funded Flyslip grant (No 16291) is also acknowledged.

References

Regina Barzilay and Mirella Lapata. 2005. Modeling local coherence: An entity-based approach. In *Proceedings of ACL 2005*, pages 141–148.

Regina Barzilay and Lillian Lee. 2004. Catching the drift: Probabilistic content models with applications to generation and summarization. In *Proceedings of HLT-NAACL 2004*, pages 113–120.

Susan E. Brennan, Marilyn A. Friedman [Walker], and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of ACL 1987*, pages 155–162, Stanford, California.

Herbert. H. Clark. 1977. Bridging. In P. N. Johnson-Laird and P. C. Wason, editors, *Thinking: Readings in Cognitive Science*, pages 9–27. Cambridge University Press.

Aggeliki Dimitromanolaki and Ion Androutsopoulos. 2003. Learning to order facts for discourse planning in natural language generation. In *Proceedings of the 9th European Workshop on Natural Language Generation*, pages 23–30, Budapest, Hungary.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Amy Isard, Jon Oberlander, Ion Androutsopoulos, and Colin Matheson. 2003. Speaking the users' languages. *IEEE Intelligent Systems Magazine*, 18(1):40–45.

Nikiforos Karamanis, Massimo Poesio, Chris Mellish, and Jon Oberlander. 2004. Evaluating centering-based metrics of coherence using a reliably annotated corpus. In *Proceedings of ACL 2004*, pages 391–398, Barcelona, Spain.

Nikiforos Karamanis. 2003. *Entity Coherence for Descriptive Text Structuring*. Ph.D. thesis, Division of Informatics, University of Edinburgh.

Nikiforos Karamanis. 2006. Evaluating centering for information ordering in two new domains. In *Proceedings of NAACL 2006*.

Rodger Kibble and Richard Power. 2000. An integrated framework for text planning and pronominalisation. In *Proceedings of INLG 2000*, pages 77–84, Israel.

Rodger Kibble and Richard Power. 2004. Optimizing referential coherence in text generation. *Computational Linguistics*, 30(4):401–416.

Rodger Kibble. 2001. A reformulation of rule 2 of centering theory. *Computational Linguistics*, 27(4):579–587.

Alistair Knott and Robert Dale. 1994. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1):35–62.

Alistair Knott, Jon Oberlander, Mick O'Donnell, and Chris Mellish. 2001. Beyond elaboration: The interaction of relations and focus in coherent text. In T. Sanders, J. Schilperoord, and W. Spooren, editors, *Text Representation: Linguistic and Psycholinguistic Aspects*, chapter 7, pages 181–196. John Benjamins, Amsterdam.

Mirella Lapata. 2003. Probabilistic text structuring: Experiments with sentence ordering. In *Proceedings of ACL 2003*, pages 545–552, Saporo, Japan, July.

William C. Mann and Sandra A. Thompson. 1987. Rhetorical structure theory: A theory of text organisation. Technical Report RR-87-190, University of Southern California / Information Sciences Institute.

Massimo Poesio, Rosemary Stevenson, Barbara Di Eugenio, and Janet Hitzeman. 2004. Centering: a parametric theory and its instantiations. *Computational Linguistics*, 30(3):309–363.

Richard Power, Donia Scott, and Nadjet Bouayad-Agha. 2003. Document structure. *Computational Linguistics*, 29(2):221–260.

Ehud Reiter and Somayajulu Sripada. 2002. Should corpora texts be gold standards for NLG? In *Proceedings of INLG 2002*, pages 97–104, Harriman, NY, USA, July.

Michael Strube and Udo Hahn. 1999. Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors. 1998. *Centering Theory in Discourse*. Clarendon Press, Oxford.

Generating coherence relations via internal argumentation

Rodger Kibble
Dept of Computing, Goldsmiths College
University of London, UK
r.kibble@gold.ac.uk

Abstract

A key requirement for the automatic generation of argumentative or explanatory text is to present the constituent propositions in an order that readers will find coherent and natural, to increase the likelihood that they will understand and accept the author's claims. Natural language generation systems have standardly employed a repertoire of coherence relations such as those defined by Mann and Thompson's Rhetorical Structure Theory. This paper models the generation of persuasive monologue as the outcome of an "inner dialogue", where the author attempts to anticipate potential challenges or clarification requests. It is argued that certain RST relations such as Motivate, Evidence and Concession can be seen to emerge from various pre-empting strategies.

1 Introduction

A key requirement for the automatic generation of argumentative or explanatory text is to present the constituent propositions in an order that readers will find coherent and natural, to increase the likelihood that they will understand and accept the author's claims. Ideally, any objections or clarification requests that an audience might raise will already have been countered by elements of the author's argument. In fact this paper models the generation of persuasive monologue as the outcome of an "inner dialogue", where the author attempts to anticipate potential challenges or clarification requests. It will be argued that certain coherence relations can be seen to emerge from various strategies for pre-empting or "obviating" challenges or clarification requests.

This paper assumes a model of dialogue as updating participants' information states (IS), where an IS consists of a **commitment store** or a record of each interlocutor's propositional and practical commitments (cf (Hamblin, 1970; Brandom, 1994; Walton and Krabbe, 1995)) rather than "mental states" such as belief and intention (cf (Cohen and Levesque, 1990)). This approach is motivated at greater length and contrasted with other commitment-based approaches such as (Matheson et al., 2000) in (Kibble, 2004; Kibble, 2006b), where more details are given of the proposed structure and contents of commitment stores. The key assumptions for the purposes of this paper are:

1. Each agent in a dialogue keeps a score of social commitments for all participants, including itself. Commitments can be classified into *practical* (commitments to act, corresponding to *intentions* in mentalistic accounts) and propositional or *doxastic* (commitments to justify an assertion, corresponding to *beliefs*).

2. Agents play one of three dynamically assigned roles at any given point in a dialogue: Speaker (**Sp**), Addressee (**Ad**), or Hearer (**He**) who is not directly addressed.

3. For an agent α to assert ϕ is to acknowledge commitment to ϕ; other agents may also attribute consequential commitments to α.

4. Additionally, a dialogue act constitutes an attempt to commit Addressee(s) to a proposition or a course of action, as detailed in the following section.

5. Addressee's options include accepting the proffered commitment, challenging it or requesting clarification.

This paper will focus on modelling persuasive monologue, or extended dialogue turns, as emerging from a process of internal argumentation, with the virtual agents Planner (**Pl**) in place of **Sp** and Critic **(Cr)** substituted for **Ad**. I will aim to show how a variety of Mann and Thompson's RST relations such as Motivate, Justify, Evidence, Concession and Elaboration can be seen to emerge from different text planning strategies (Mann and Thompson, 1987; Taboada and Mann, 2006) . It might be argued that this is an essentially trivial exercise in shifting information from a pre-defined set of coherence relations to a pre-defined set of dialogue acts and moves. However, there are independent motivations for developing models for

dialogue and argumentation, and the argument in this paper is that a (possibly partial) account of coherence relations in monologue emerges as a side-effect of these models. The paper will conclude by addressing some apparent differences between dialogue and monologue as discussed by (Reed, 1998) and (Green and Carberry, 1999).

2 Argumentation and discourse relations

The full framework will include specifications for the proto-speech acts listed below. Note that I use upper-case Greek letters such as Φ to represent speech acts themselves and lower-case letters such as ϕ for the propositional content of the speech acts.

assert(Sp, ϕ, Ad, He) undertake commitment to justify a propositional claim; attempt to bestow same commitment on **Ad**.

instruct(Sp, ϕ, Ad, He) attempt to bestow a practical commitment on Addressee.

endorse(Sp, ϕ, Ad, He) Speaker adopts a commitment specified by Addressee

challenge(Sp, Φ, Ψ, Ad, He): require agent to justify or retract a commitment offer Φ, with Ψ as an optional counter-commitment. Note that the challenge may be directed at the propositional content ϕ, or at the appropriateness of the speech act itself.

respond(Sp, challenge(Ad, Φ, Ψ, Sp, He), Ξ, Ad, He)

respond to a challenge with a dialogue act Ξ which may be:

- asserting ξ as evidence for ϕ, or as justification for uttering Φ;
- retracting commitment to ϕ, the propositional content of Φ;
- withdrawing a claim to justification for the speech act Φ;
- challenging Ψ;
- requesting clarification of Ψ;
- ϵ - the null act. How this is interpreted will depend on the particular conventions currently in force: it may be understood at different times as implicit endorsement, implicit denial or non-committal.

retract(Sp, ϕ, Ad, He) withdraw a commitment to ϕ.

query(Sp, Φ, Ad, He) request clarification of Φ

respond(Sp, query(Ad, Φ, Sp, He), Ψ, Ad, He)

respond to request for clarification of Φ by uttering the speech act Ψ.

2.1 Examples of dialogue and monologue

The following examples consist of a short dialogue followed by two variants of a monologue expressing roughly the same content and exemplifying particular rhetorical structures.

Example (a)

A: You should take an umbrella.
B: Why?
A: It's going to rain.
B: It doesn't look like rain to me. It's sunny
A: Michael Fish predicted it.
B: Who's he?
A: He's a weather forecaster on the BBC.
B: OK.

In terms of the speech acts defined above, this exchange can be represented (somewhat simplified) as follows:

A: instruct(A, *take-umbrella(B)*, B, _);

B: challenge(B, *take-umbrella(B)*, _, A, _);

A: respond(A, challenge(B, *take-umbrella(B)*, _, A, _), assert(A, *rain-later*, B, _), B, _)

B: challenge(B, *rain-later*, *sunny-now*, A, _);

A: respond(A, challenge(B, *rain-later*, *sunny-now*, A, _), assert(A,*predict(fish,rain)*,B, _), B, _)

B: query(B, *predict(fish,rain)*, *A*, _)

A: respond(A, query(B, *predict(fish,rain)*, *A*, _), assert(A, *BBC(fish)*, B, _), B, _)

B: endorse(B, {*BBC(fish)*; *predict(fish,rain)*; *rain-later*; *take-umbrella(B)*}, A, _)

Example (b)

A: You should take an umbrella. It's going to rain. I heard it on the BBC.

A possible RST analysis of this example is:

Motivate

Nucleus *take-umbrella(B)*

Satellite: Evidence

Nucleus *rain-later*
Satellite *BBC-forecast(rain)*

Example (b′)

A: You should take an umbrella. It's going to rain, even though it looks sunny right now. I heard it on Michael Fish's slot. He's a weather forecaster at the BBC.

Proposed RST analysis:

Motivate

Nucleus *take-umbrella(B)*

Satellite

Evidence
Nucleus
Concession
Nucleus *rain-later*
Satellite *sunny-now*
Satellite
Background
Nucleus *predict(fish,rain)*
Satellite *BBC(fish)*

Example (c)

A: The BBC forecast was for wet weather. It will rain later. You should take an umbrella.

Proposed RST analysis: same rhetorical structure as (b) but realised in a satellite-first sequence:

Motivate

Satellite: Evidence

Satellite *BBC-forecast(rain)*
Nucleus *rain-later*

Nucleus *take-umbrella(B)*

2.2 Speaker strategies

In the above scenario, suppose A has the goal that B undertake a practical commitment to carry an umbrella. Examples (a - c) illustrate three different strategies:

(i) Issue a bare instruction; offer justification only if challenged.

(ii) Issue an instruction, followed by an assertion that **pre-empts** a potential challenge, and recursively pre-empt challenges to assertions.

(iii) **Obviate** the challenge by uttering the justification **before** the instruction, and recursively obviate potential challenges to assertions.

(The terms **pre-empt** and **obviate** are used with these particular meanings in this paper, which may not be inherent in their ordinary usage.) Note that examples (a) and (b′) exhibit the same sequence of propositions, which is consistent with the assumption that (b′) results from a process of internal argumentation with a virtual agent that raises **Ad**'s potential objections. The following section will sketch a formulation of strategies (i - iii) in terms of the Text Planning task of natural language generation.

3 Dialectical text planning

I will assume some familiarity with terms such as "text planning" and "sentence planning". These are among the distinct tasks identified in Reiter's "consensus architecture" for Natural Language Generation (Reiter, 1994); see also (Bateman and Zock, 2003):

Text Planning/Content Determination - deciding the content of a message, and organising the component propositions into a text structure (typically a tree). I will make a distinction between the **discourse plan** where propositions in the initial message are linked by coherence relations, and the **text plan** where constituents may be re-ordered or pruned from the plan.

Sentence Planning - aggregating propositions into clausal units and choosing lexical items corresponding to concepts in the knowledge base; this is the level at which the order of arguments and choice of referring expressions will be determined.

Linguistic realisation - surface details such as agreement, orthography etc.

3.1 Discourse planning

Text planning is modelled in what follows as the outcome of an inner dialogue between two virtual agents, the Planner (**Pl**) and the Critic (**Cr**). The Critic is a user model representing either a known interlocutor or a "typical" reader or hearer. A's options (i -iii) in Section 2.2 above can be seen to correspond to three different strategies which I will call *one-shot*, *incremental* and *global*. These strategies are presented in rather simplified pseudo-code below, in particular I only consider the assert action and selected responses to it.

One-shot planning

Speaker produces one utterance per dialogue turn which may be:

- a bare assertion ϕ;
- response to a challenge or clarification request from Addressee;

- challenge to Address's most recent or salient assertion, or request for clarification;
- ϵ

The message is passed directly to the text planner without being checked by the Critic. This strategy is appropriate when no user model is available.

Incremental Planning

Speaker generates the "nuclear" utterance and then calculates whether a challenge is likely, and recursively generates a response to the challenge if possible. This is the strategy of **pre-empting** challenges referred to in section 2.2. The response is immediately committed to the right frontier of Speaker's text plan.

procedure inc-tp(Φ)
  where Φ is some speech act with propositional content ϕ;
send Φ to text planner;
assert(Pl, ϕ, Cr, _);
if challenge(Cr, ϕ, ψ, Pl, _)
then do inc-tp(respond(Pl, challenge(Cr, ϕ, ψ, Pl, _), Ξ, Cr, _);
else quit.

This strategy is appropriate when a suitable user model is available but resource limits or time-criticality make it desirable to interleave discourse planning, text planning and sentence generation.

Goal-directed Planning

The sequence is globally planned in order to rebut potential challenges by generating responses to them ahead of the nuclear proposition. This is the strategy I have dubbed **obviating** challenges in section 2.2.

procedure gd-tp(Φ)
  where Φ is some speech act with propositional content ϕ;
initialise stack = [ ];
call gd-tp-stack(Φ);
do until stack = [ ]:
  pop Ψ from stack;
  add Ψ to text plan;
end gd-tp()

procedure gd-tp-stack(Φ)
stack = [Φ | stack];
assert(Pl, ϕ, Cr, _);
if challenge(Cr, ϕ, ψ, Pl, _)
then do gd-tp-stack(respond(Pl, challenge(Cr, ϕ, ψ, Pl, _), Ξ, Cr, _);
else quit gd-tp-stack
end gd-tp-stack()

This strategy is appropriate for applications where resources allow for the full discourse plan to be generated in advance of text planning so that constituents may subsequently be reordered or pruned to produce a possibly more "natural" and readable text.

3.2 Text planning and plan pruning

If we consider the examples in section 2.1: (b), (b′) are typical products of incremental planning and (c) of goal-directed planning. The former will result in **nucleus-first** structures, while the default ordering resulting from the latter will realise satellites **before** nuclei. Two refinements are discussed in this section: **plan pruning** and **re-ordering** of the text plan.

The differences between (b) and (b′) demonstrate that the text planner has a choice over whether to realise only the Planner's contributions or those of the Critic as well. The latter option, retaining the proposition *sunny-now*, results in instances of RST's Concession relation. This is a special case of **plan pruning** as described by (Green and Carberry, 1999), where a constituent may be removed if it is inessential to the speaker's purpose: for instance it may be inferrable from other material in the plan. Green and Carberry motivate this with the aid of the following example (their (13a-e)), illustrating how a question-answering system might decide how much unrequested information to include in an indirect answer to a yes-no question.

Example (d)

(i) Q: Can you tell me my account balance?
(ii) R: [No.]
(iii) [I cannot access your account records on our computer system.]
(iv) The line to our computer system is down.
(v) You can use the ATM machine in the corner to check your account.

Items (ii - iii), shown in square brackets, can be suppressed since (iii) is inferrable from (iv) and in turn implies (ii). This assumes that the user is aware, or can accommodate the fact that their account balance is kept on the computer system. This example is compared with an "imaginary dialogue" where each statement responds to a specific question from the user.

As stated above, the planning strategies outlined in section 3 produce texts that are uniformly either satellite-first or nucleus-first by default. There is a need to generalise the strategies so that the planner

can dynamically switch from one to the other, in order to produce texts such as:

Example (e)

It's going to rain. I heard it on the BBC. You should take an umbrella.

RST analysis:

Motivate
Satellite: Evidence
Nucleus *rain-later*
Satellite *BBC-forecast(rain)*
Nucleus *take-umbrella(B)*

By distinguishing between the **discourse plan** and **text plan** we allow for re-ordering of constituents at the level of the text plan, within the partial ordering defined by the discourse plan. For instance, a different ordering of propositions might improve the referential coherence of a text according to Centering Theory (Kibble and Power, 2004).

3.3 Summary

In contrast to approaches to text generation that carry out top-down planning using pre-defined coherence relations I have argued that certain RST relations can be seen to emerge sequences of internalised dialogue moves that aim to pre-empt or obviate potential challenges or clarification requests, as follows:

instruct-challenge-respond underlies Motivation or Justify depending on the content of the challenge and response;

assert-challenge-respond underlies Evidence if the propositional content is challenged, or Justify if the appropriateness of the assert act itself is at issue.

<**any-speech-act**>**-challenge-respond** underlies Concession if the content of the challenge is realised in the text.

<**any-speech-act**>**-query-respond** underlies Background.

It remains to be seen if further RST relations can be modelled using the "dialectical" method.

4 Discussion and future work

4.1 Objections to "implicit dialogue"

Reed (Reed, 1998) argues against identifying a persuasive monologue with an implicit dialogue and emphasises the importance of distinguishing the *process* of creating a monologue from the *product*, the monologue itself. Now, it is not argued here that a monologue is nothing more than a trace of the dialogical process of constructing an argument. The "goal-directed" strategy allows for a phase of pruning and re-ordering the text plan (not described in detail here) although the default is for propositions to be realised in the sequence in which they are added to the discourse plan.

Reed puts forward an important argument: that a crucial difference is the fact that unlike a dialogue, a "pure" monologue must not contain a *retraction* in the sense of asserting a proposition and its negation. This has implications for the discussion of text planning strategies in section 3 above, since there is the possibility of a contradiction occurring in a sequence of responses to recursive challenges. On the one hand, goal-directed planning could be extended with a backtracking facility and consistency checking such that indefensible claims or even the nuclear proposition itself could be withdrawn before proceeding to sentence generation, if a challenge generated by the Critic shows up a contradiction in the existing plan. However, the essence of incremental planning is intended to be that each proposition is committed to the text plan, to be passed on to the sentence planner, *before* considering potential challenges. The algorithm as adumbrated above certainly allows the possibility that contradictory propositions will be added to the plan, as a consequence of limitations on speakers' memory and reasoning capabilities.

The proscription of overt retraction would certainly be a reasonable design feature for a computer system generating argumentative text. However, this paper is also concerned with modelling the ways in which human speakers might construct an argument, and so this comes down to an *empirical* question as to whether speakers delivering an extempore monologue will ever realise part-way through that there are insuperable objections to their initial claim (or a subordinate claim), and end up withdrawing it. For instance, the medium of communication might be an electronic "chat" forum such that all keystrokes are instantly and irrevocably transmitted to other logged-on users. It is not obvious that this possibility should be ruled out in principle, or even that it can be ruled out in a resource-limited system following "incremental planning" as defined here.

4.2 Complexity of speaker and hearer strategies

To generalise the remarks in section 2.2 above: Speakers have a choice between uttering a bare speech act Φ (e.g. assertion, instruction); uttering Φ followed by a supplementary speech act Ψ that pre-empts potential objections; and preceding Φ with a supplementary Ψ that "obviates" an anticipated challenge - with the last two options potentially applying recursively.

Conversely the Addressee appears to have the following options (cf (Kibble, 2001)):

(i) Endorse A's instruction, assertion etc and adopt the offered commitment ϕ;

(ii) Challenge Φ and demand justification or clarification;

(iii) Defer the challenge in case a justification follows.

(iv) Reject Φ out of hand: e.g. if it is contradictory or grossly offensive.

This can be partially formalised as the update function $\langle \sigma, Stack \rangle[.]$ where σ is Addressee's self-attributed commitment store:

1. Initial state: $\langle \emptyset, \top \rangle$
2. Updating with an empty stack:
   $\langle \sigma, \top \rangle[\Phi] =$
   a. $\langle \sigma \cup \{\phi\}, \top \rangle$ if the commitment ϕ is endorsed; else
   b. $\langle \sigma, (\Phi, \top) \rangle$ if the commitment is neither endorsed nor rejected; else
   c. $\langle \sigma, \top \rangle$

   This clause involves only PUSH operations: Addressee will either update σ with ϕ (a), defer Φ by placing it on the stack (b), or reject it, leaving the commitment store unchanged (c).
3. Updating with a non-empty stack:
   $\langle \sigma, (\Psi, S) \rangle[\Phi] =$
   a. $\langle \sigma \cup \{\phi\}, S) \rangle[\Psi]$ if the commitment ϕ is endorsed; else
   b. $\langle \sigma, (\Psi, (\Phi, S)) \rangle$ if the commitment is neither endorsed nor rejected; else
   c. $\langle \sigma, (\Psi, S) \rangle$

   This inclues the POP clause: after updating with a new speech act ϕ, Addressee will reevaluate any speech act Ψ still on the stack.

Note that (2/3b) cover Addressee options (ii - iii) above: the update effect on this section of the commitment store is the same whether Addressee has challenged Speaker or is silently waiting for a justification.
Let

Φ = instruct(A, *take-umbrella*, B, _);
Ψ = assert(A, *rain-later*, B, _);
Ξ = assert(A, *BBC-forecast(rain)*, B, _)
ϕ = *take-umbrella(B)*
ψ = *rain-later*
ξ = *BBC-forecast(rain)*

Assume B is disposed to reason as follows:

- A is a reliable reporter of the BBC weather forecast;
- If the BBC forecasts rain, it will rain
- If it is going to rain, one should take an umbrella

Examples (b) and (c)(repeated below) will be processed as follows.

Example (b)

A: You should take an umbrella. It's going to rain. I heard it on the BBC.

$\langle \emptyset, \top \rangle[\Phi] = \langle \emptyset, (\Phi, \top) \rangle$
$\langle \emptyset, (\Phi, \top) \rangle[\Psi] = \langle \emptyset, (\Psi, (\Phi, \top)) \rangle$
$\langle \emptyset, (\Psi, (\Phi, \top)) \rangle[\Xi] =$
$\langle \{\xi\}, (\Phi, \top) \rangle[\Psi] =$
$\langle \{\xi, \psi\}, \top \rangle[\Phi] =$
$\langle \{\xi, \psi, \phi\}, \top \rangle$

In this example, ϕ relies on the support of ψ which in turn depends on ξ so the initial speech acts are deferred by being pushed onto the stack until the contents of the commitment store allow for them to be endorsed.

Example (c)

A: The BBC forecast was for wet weather. It will rain later. You should take an umbrella.

$\langle \emptyset, \top \rangle[\Xi] = \langle \{\xi\}, \top \rangle$
$\langle \{\xi\}, \top \rangle[\psi] = \langle \{\xi, \psi\}, \top \rangle$
$\langle \{\xi, \psi\}, \top \rangle[\phi] = \langle \{\xi, \psi, \phi\}, \top \rangle$

Each successive proposition is supported by the contents of the existing commitment store, or in the case of ξ is acceptable on its own, so no stacking takes place.

This analysis suggests that the various strategies impose directly opposed processing loads on Speaker and Addressee. Recall that (b) was presented as a typical product of **incremental planning** while (c) has the default ordering of **goal-directed planning**. Goal-directed planning requires more computational resources on the part of the Speaker but evidently results in (satellite-initial or mixed) texts that are easier for Hearers to process. The question arises whether speakers optimise their utterances for the audience or follow a path of least effort. This is a topic of debate amongst researchers in psycholinguistics, as evidenced by the claims put forward by (Pickering and Garrod, 2004) and the various responses collected together in the same journal issue.

4.3 Future work

The following issues will be addressed in future research:

Coherence, user modelling and reasoning. It is assumed that for a text to be *coherent* as perceived by the intended audience means that there is an increased likelihood that they will endorse the proffered (practical or doxastic) commitments *and* that this will require less cognitive effort on the audience's part, by comparison with less coherent texts. The success of a dialectical, user-model oriented text planning regime will clearly depend crucially on the reliability of the user models and the validity of the reasoning processes by which the planner calculates potential challenges and suitable responses. Some important topics are:

- modelling *specific* users to whom a message is directed, versus *typical* readers of a text which is not directed at any particular individual;
- modelling information states of the virtual agents **Pl** and **Cr**, in view of arguments that speakers and hearers have asymmetric context models in dialogue (Ginzburg, 1997).

Preempting clarification requests. This paper has modelled the Background relation as resulting from preemption of a clarification request (CR).) Studies including (Ginzburg and Cooper, 2004) have shown that CRs can be directed at various levels of linguistic representation or content. In the following example (constructed for this paper), the elliptical query *Maclean?* could have any of the responses shown:

Example (f)

(i)A: Maclean's defected to the USSR.
(ii) B: Maclean?
(iii) A: Yes, Maclean of all people.
(iv) A: Donald Maclean, head of the American desk at the FO.
(v) A: That's M - a - c - l - e - a - n.

This raises architectural issues since it has been assumed in this paper that preemptions are generated at the discourse planning stage, where details of linguistic realisation such as how to spell a proper name may not be available. Future work will address the question of whether and how clarifications at distinct levels of representation can be integrated into the dialectical planning model.

Acknowledgments

A shorter version of this paper will be presented at the ECAI 2006 workshop Computational Models of Natural Argumentation as (Kibble, 2006a). I'm grateful to the ESSLLI and CMNA reviewers for their helpful comments.

References

John Bateman and Michael Zock. 2003. Natural language generation. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, pages 284 – 304. Oxford University Press, Oxford.

Robert Brandom. 1994. *Making it Explicit.* Harvard University Press, Cambridge, Massachusetts and London.

Philip Cohen and Hector Levesque. 1990. Persistence, intention and commitment. In Philip Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in Communication*, pages 33 – 69. MIT Press, Cambridge, Massachusetts and London.

Jonathan Ginzburg and Robin Cooper. 2004. Clarification ellipsis and the nature of contextual updates in dialogue. *Linguistics and Philosophy*, 27(3):297 – 365.

Jonathan Ginzburg. 1997. On some semantic consequences of turn taking. In *Proceedings of the 11th Amsterdam Colloquium.*

Nancy Green and Sandra Carberry. 1999. A computational model for taking initiative in the generation of indirect answers. *User Modeling and User-Adapted Interaction*, 9(1/2):93–132. Reprinted in Computational Models of Mixed-Initiative Interaction, Susan Haller, Alfred Kobsa, and Susan McRoy, eds., Dordrecht, the Netherlands, 277-316.

Charles Hamblin. 1970. *Fallacies.* Methuen, London.

Rodger Kibble and Richard Power. 2004. Optimizing referential coherence in text generation. *Computational Linguistics*, 30 (4):401–416.

Rodger Kibble. 2001. Inducing rhetorical structure via nested update semantics. In *Proceedings of the Fourth International Workshop on Computational Semantics*, University of Tilburg, The Netherlands.

Rodger Kibble. 2004. Elements of a social semantics for argumentative dialogue. In *Proceedings of the Fourth Workshop on Computational Modelling of Natural Argumentation*, Valencia, Spain.

Rodger Kibble. 2006a. Dialectical text planning. In Floriana Grasso, Rodger Kibble, and Chris Reed, editors, *Proceedings of 6th Workshop on Computational Models of Natural Argumentation*, Riva del Garda, Italy.

Rodger Kibble. 2006b. Reasoning about propositional commitments in dialogue. To appear in *Research on Language and Computation.*

William C. Mann and Sandra A. Thompson. 1987. Rhetorical structure theory: A theory of text organization. Technical report, Marina del Rey, CA: Information Sciences Institute.

Colin Matheson, Massimo Poesio, and David Traum. 2000. Modelling grounding and discourse obligations using update rules. In *Proceedings of NAACL 2000*.

Martin Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–225.

Chris Reed. 1998. Is it a monologue, a dialogue or a turn in a dialogue? In *Proceedings of the 4th International Conference on Argumentation (ISSA98)*, Amsterdam. Foris.

Ehud Reiter. 1994. Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible? In *Proceedings of 7th International Natural Language Generation Workshop*, pages 163–170.

Maite Taboada and William Mann. 2006. Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies*, 8(3). To appear.

Douglas Walton and Eric Krabbe. 1995. *Commitment in dialogue*. State University of New York Press, Albany.

Proceedings of the ESSLLI Workshop on Coherence in Generation and Dialogue

SDRT and Multi-modal Situated Communication

Andy Lücking **Hannes Rieser** **Marc Staudacher**
Bielefeld University, SFB 360, B3

Abstract

Classical SDRT (Asher and Lascarides, 2003) discussed essential features of dialogue like adjacency pairs or corrections and up-dating. Recent work in SDRT (Asher, 2002; Asher, 2005) aims at the description of natural dialogue. We use this work to model situated communication, *i.e.* dialogue, in which sub-sentential utterances and gestures (pointing and grasping) are used as conventional modes of communication. We show that in addition to cognitive modelling in SDRT, capturing mental states and speech-act related goals, special postulates are needed to extract meaning out of contexts. Gestural meaning anchors Discourse Referents in contextually given domains. Both sorts of meaning are fused with the meaning of fragments to get at fully developed dialogue moves. This task accomplished, the standard SDRT machinery, tagged SDRSs, rhetorical relations, the up-date mechanism, and the Maximize Discourse Coherence constraint generate coherent structures. In sum, meanings from different verbal and non-verbal sources are assembled using extended SDRT to form coherent wholes.

1 Credits

We are grateful to Nicholas Asher for having taught us SDRT in the years 2003-2005 and for letting us work with unpublished SDRT material, especially Asher (2005). Our work on SDRT was supported by the CRC "Situated Artificial Communicators," project "Deixis in Construction Dialogue" (DEIKON) at Bielefeld University, funded by the German Research Foundation (DFG). Thanks to three anonymous reviewers whose remarks were helpful for improving our paper.

2 Situated Communication

Recently, the interest in retrieval and representation of *non-sentential speech* has been growing, as the collection Elugardo and Stainton (2005) shows. The debate on how to account properly for the phenomena is still ongoing. However, it emerges that it puts further constraints on how mainstream linguistics *should* be done. For if non-sentential speech is an essential part of language, notions such as grammaticality and coherence have to be applicable to it. In this paper, we are concerned more specifically with issues of the semantics/pragmatics-interface of non-sentential speech. We understand this kind of language use as being part of *situated communication* and propose a formal model for it. Thus, we start by characterising situated communication. Consider the two examples (1) and (2).

(1) World economic growth slowed noticeably in 2005 from the strong expansion in 2004.

(2) In a two-person dialogue between I and C in a room with some bolts on a table:
   a. I: This bolt in the rear there (while I is pointing)
   b. C: This one? (while C is grasping some bolt)
   c. I: Yes

In opposition to (1), the kind of language use as in (2) is what we call *situated communication*.[1] Language use of this kind can be recognized by a couple of characteristics. First, utterances are typically sub-sentences and not "full-fledged sentences" in a grammatical sense. On a standard account, only sentences (and not parts) express

[1] In contrast, the use of some fragments such as question-answer- or request-answer-pairs is determined by rules of grammar. We are interested in cases which are extremely context dependent and need inference for their resolution. These are cases calling for "resolution-via-inference" in the terms of the Schlangen and Lascarides (2003) approach.

propositions. Still, sub-sentences can be used to express propositions. For example, (2-a) says of a particular bolt on the table that it is the one to be grasped. So, after all, utterances of sub-sentences can express propositional content.

Secondly, such utterances are typically accompanied by linguistically relevant non-verbal behaviour such as pointing gestures or graspings. Deixis is typical for this kind of language use. In (2-b), for example, it is asked of a certain bolt on the table whether it is the one I meant in (2-a). To establish the reference to that bolt, C's grasping seems to be essential.

Thirdly, such utterances as in (2) can be used to perform speech acts. It can be meaningfully asked what the illocutionary role of such an utterance is (*e.g.* (2-b) is a `Check-back`) and which proposition is thereby expressed. However, it cannot be a property of the expression's content that makes it express a certain speech act or proposition. For example consider an utterance of 'scissors' in a sewing shop, in the rock-paper-scissors-game, or on a shopping list. In each context, the utterance is used to express something different. While the first two can be taken to express a proposition, the inscription on the list cannot. It might merely be some mnemonic device to perform the shopping. Moreover, the different uses of 'scissors' seem to correlate with conventions governing its use. So, a special stock of conventions seems to regulate its interpretation. Being conventions each of them is mutually believed (in some dispositional sense). Together they allow agents the use of sub-sentential utterances and gestures to successfully communicate as (2) shows.

From these three properties of situated communication we derive the minimal requirements for a formal model of situated communication. As a framework we are going to use SDRT. The minimal requirements are: The theory has to explain which sentential content a non-sentential utterance expresses and which speech act is performed by uttering it. The explanation has to make use of a special stock of conventions. Moreover, discourse coherence should be explained.

For purposes of illustration we use discourse (3) as our main example:

(3) a. *I: Die rote* $\searrow_a$ *Holzscheibe*
I: The red $\searrow_a$ wooden disc

b. *C:* $\downdownarrows_a$ *Diese?*
C: $\downdownarrows_a$ This one?

c. *I: Ja*
I: Yes

Some comment about (3) is in order. Dialogues like (3) are called *Object Identification Games* and have been examined in project B3 of the Collaborative Research Centre "Situated Artificial Communicators" (SFB 360)[2].

In (3-a) the symbol '$\searrow_a$' indicates, when the stroke of a pointing gesture occurs. The symbol is written after the word whose occurrence is immediately preceding in time. The index indicates the object a the pointing refers to. Likewise in (3-b), the symbol '$\downdownarrows_a$' indicates the grasping of the object a. Two video-stills showing the pointing and the grasping in (3) are provided in Fig. 1(b) and 1(c), respectively.

(3) is a gloss for a corpus entry which has been built from the experimental data. Each corpus entry is a description of a dialogue which occurred in the experimental setting. The corpus annotation format features both verbal and non-verbal elements in such a way that the role of pointing gestures can be studied theoretically. Fig. 1(a) shows a graphical representation of a corpus entry.

In Object Identification Games, two persons, the *instructor* (I) and the *constructor* (C), are involved in a coordination task. It is a two-player-game of spotting an object in a given situation. The instructor has the role of the "description-giver". The constructor has the role of the "object-identifier". The players interact by performing moves in the game. The game starts with the instructor's choosing a certain object out of the parts of a toy air plane spread on a table. She instructs the constructor to identify the object she has chosen by referring to it. The constructor then has to resolve the instructor's reference act and to give feedback. Thereby, reference has to be negotiated and established using a special kind of dialogue game. The game ends, if the constructor has identified the correct object on the table and the instructor has accepted it.

3 The standard exposition of SDRT

As a dynamic discourse representation theory modelling the semantics/pragmatics-interface, SDRT is an apt framework for modelling situated communication. For our purposes it is important to note that "standard" SDRT as presented in the

[2] `http://www.sfb360.uni-bielefeld.de/`

(a) The annotation in TASX. (b) The pointing in (2a). (c) The grasping in (2b).

Figure 1: Annotation of a natural dialogue in project B3 of SFB 360.

2003 book (Asher and Lascarides, 2003) requires its input to be of a type corresponding to sentences in the grammar.

To understand this point we illustrate SDRT's general architecture (Fig. 2) and its implicit notion of discourse construction using the sample dialogue (3).

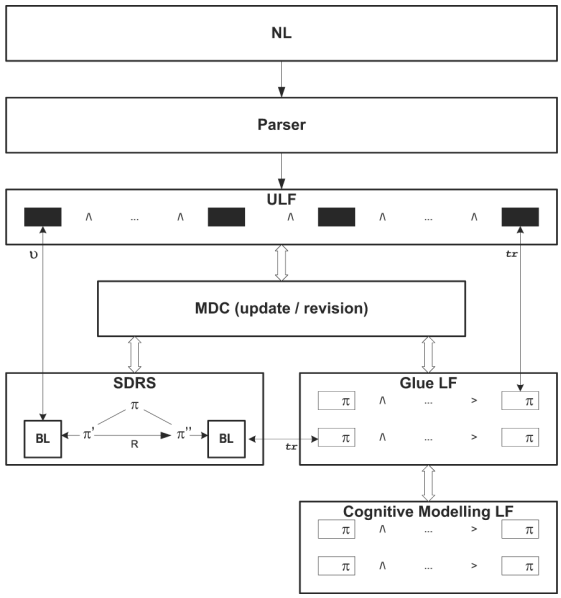


Figure 2: The general SDRT architecture.

Since SDRT provides no grammar, the *NL-input* is assumed to be available as *underspecified logical forms (ULFs)* constructed by a *parser*. The underspecification reflects the fact that, in general, the grammar does not determine a unique logical form but a set of possible forms corresponding to the interpretation licensed by the grammar alone. ULFs describe base logical forms, *i.e. SDRSs*.

So, SDRT's processing begins by assuming some context C (a potentially empty description of SDRSs) and by assuming the ULF of the dialogue's first utterance (3-a) being part of it. In the next step, C is updated with the ULF of (3-b) yielding a new context C'. SDRT's *update mechanism* assumes that the new utterance is related to an available attachment point by means of an underspecified *discourse relation* $R(a, b)$. In this case, the most coherent one is (3-a). However, discourse relations relate only content having satisfaction conditions.

At this point SDRT fails with sub-sentential utterances. The interpretations of (3-b) licensed by the grammar alone are not contents having satisfaction conditions. Intuitively, the new context C' describes SDRSs in which a and b are related through some admissible resolution of R, *e.g.* $Q-Elab$. So, what $Q-Elab$ should relate is of the wrong semantic type. Something having satisfaction conditions is required, however in the case of (3-b) an NP-denotation is present.

To make the illustration of the general architecture complete, let us assume that (3-b) had a sentential content. Then its ULF (inter alia) would be translated to the *Glue logic* and to the *Cognitive Modelling Logic* in order to resolve underspecification by pragmatic reasoning. The resolutions are translated back to the logic of ULF and added to the description. The update mechanism restricts the resolutions to those that are consistent, *i.e.* describe well-formed SDRSs.

4 SDRT's Recent Developments

One of the main achievements of standard SDRT was the application of tools originally developed for the description of monologue to dialogue including corrections, a major step, if we consider rhetorical relations and dialogue up-date and down-date problems. Three papers, Asher (2002) on Deixis, Binding and Presupposition, Asher (2005), Bielefeld Lectures on SDRT, and Schlangen and Lascarides (2003)'s work on the Interpretation of Non-Sentential Utterances in Dialogue, show that difficult NL dialogue data can be handled using SDRT's full theoretical power plus some additional assumptions. Asher (2002) deals with the following issues, relevant for our example: treatment of presuppositions, analysis of defi-

nite descriptions, especially their deictic uses, anchoring of definites in the non-linguistic context, the notion of internal and external anchors, the relation between anchoring and speech act related goals (SARGs), the cognitive effects of anchoring, the generation of MB *wrt* an object anchored. These concepts are briefly and somewhat fragmentarily introduced below.

As to presuppositions, Asher argues that the Heim-van der Sandt-Geurts account of them is incomplete and yields wrong predictions, the reason being that presupposition accommodation in the case of deictically used NPs is not always adequate. Definite descriptions introduce an underspecified relation, called *bridging relation*, between the referent and some other contextually given object, set to identity by default. Deictically used definites have to be anchored to some object in the non-linguistic context. As a consequence, anchoring involves a *de re* attitude towards the object, some sort of *knowing how* needed to solve the conversational goals (SARGs) of the speaker. SDRT uses, in opposition to specifying anchoring contextually as undertaken in Kaplan's Context Theory or Situation Semantics, DRT's external and internal anchors (Kamp, 1990). Anchoring requires linking an agent A's epistemic attitude to conversational goals. If an anchoring relation between the presupposition of a definite ψ and some element in the discourse context exists for the agent A, he is supposed to have a computable means of getting to the referent of ψ from the present non-linguistic context of utterance under some given purpose ϕ; to capture this, a notion of *path* is defined. If the anchoring function of a deictically used definite is accepted by the participants in dialogue, they are assumed to mutually believe that the definite picks out the same object for them. Hence, anchoring amounts to coordination or alignment.

Of similar importance as the discussion of definites, presupposition, binding and anchoring is the handling of fragments in dialogue, since, normally, natural dialogue does not come with utterances which can be mapped onto well-formed sentences in the theory of grammar sense. The idea in (Asher, 2005) is that fragments can be resolved iff the context in which the communication is situated provides us with two things: First, it must be mutual knowledge that a fragment with some meaning has been produced by an agent and secondly, it must be mutually believed that the fragment as produced expresses some more comprehensive content ϕ wrapped around the information reconstructed as a presupposition. In our example, the more comprehensive content ϕ is given by 'Grasp the red wooden disc!' and 'This one?', respectively. The status of these assumptions in the theoretical set up of SDRT is not yet clear, presumably, they belong to Cognitive Modelling, since mental states are involved.

Another approach to fragments is elaborated in (Schlangen and Lascarides, 2003). The idea is to assimilate sub-sentential utterances to sentences since such utterances express sentential content. From this point of view such utterances have "holes" which need to be filled in in order to express the intended content. Schlangen and Lascarides understand hole-fillers as the resolution of semantically underspecified content (and as such these are not syntactic ellipses). *I.e.* the linguistic form of such utterances is of the category "sentence fragment" which in turn consists of the usual linguistic items such as an NP. The logical form is assumed to have a semantically underspecified relation linking its variables such that each resolution expresses sentential contents, among them the intended one. Schlangen and Lascarides' main thesis is that the resolution of such utterances can be modelled as a by-product of establishing coherence in discourse.

Schlangen and Lascarides acknowledge that their approach has difficulties with sub-sentential utterances which need a "resolution-via-inference", *i.e.* a resolution that cannot use the immediate linguistic context containing a "copy" of the material needed (as in the case of short answers to wh-questions). We, following (Asher, 2002; Asher, 2005), propose a new direction for accounting for this class of utterances. Our thesis is that competent speakers have linguistic knowledge in form of situated conventions allowing the speakers to properly use and understand such utterances. Moreover, our original data shows that the role of gestures and graspings is central to correctly resolve newly introduced definites. Without a notion of external anchoring resolution cannot be explained correctly. As a by-product of the introduction of the notion of external anchoring resolution-via-inference becomes more tractable.

5 Coherence from the 2005 SDRT Perspective: A Giant Step for SDRT

SDRT's notion of coherence up to (Asher, 2005) rested on several mechanisms, the use of rhetorical relations and their semantics, especially the division into coordinating and subordinating relations, the use of SDRSs as context change potentials in the Kamp-Heim-tradition, the extended definition of up-date capturing revision in dialogue and, finally, the filter mechanism "Maximize Discourse Coherence" (MDC). All these notions were ultimately founded upon the notion of *complete* meaning, of whatever type and however explicated. These meanings in turn were conceived of as coming solely from verbal expressions using a construction algorithm in DRT fashion.

This picture fundamentally changes with 2005's SDRT: First of all, the information provided by the fragments of the description giver 'the red wooden disc' and the object identifier 'this one?', respectively, are not complete. The intuition is that the fragments combine with meanings from the context to give us complete meanings. Roughly, we want 'Grasp the red wooden disc!' on the description giver's and 'Do you mean this one?' on the object identifier's side. Once we arrive at complete meanings, the normal SDRT machinery can be put to work again. However, in order to get there, we have to use special postulates, which under specific conditions let agents in cooperative dialogue extend the fragments to directives and clarification questions, respectively. The missing information for the directive comes from the context at the beginning of the object identification game, in which the director of the experiment assigns the roles of description giver and object identifier, saying for example, 'you, A, tell the other one to grasp one of the objects in the domain' and 'you, B, identify the object described, pointed at etc. and indicate whether you have identified it'. These roles are preserved throughout the contexts developed, at least as a fall back option. In terms of SDRT: The director of the experiment fixes the type of the speech-act-related goals (SARGs) of the participating agents. Secondly, the dialogue is multimodal as the example shows, the object introduced by the description is anchored to the context by the demonstration. Similarly, the pure demonstrative used in the clarification question is anchored to the context by the object identifier's grasping. Definiteness information is treated as presuppositional, entertaining the idea that presuppositions are locally bound. On the whole, detailed context information plays a much greater role in the 2005 SDRT version as compared to the standard one, due to the fact that the meaning of the fragments has to be filled up using context information.

6 Tying Things Together

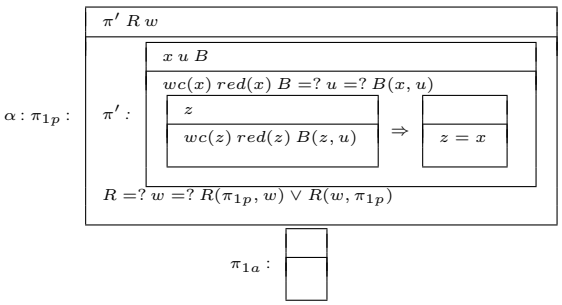
We now apply the theory to our main example (3) using a DRT-style notation. The application of the theory shows how demonstrations, discourse relations, a special stock of conventions and MDC interact in order to arrive at the intended interpretations of situated communication. We assume that in the context of Object Identification Games a special stock of conventions of the following form holds:

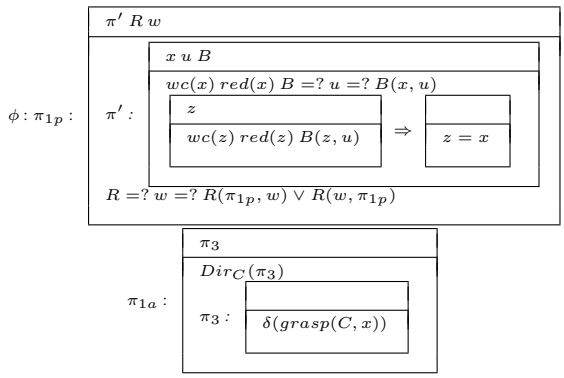
$$\mathcal{K}_{I,C}(\alpha(\pi_1) \wedge Ag(\pi_1) = I \wedge (\mathcal{MB}_{I,C}(\alpha(\pi_1) \wedge Ag(\pi_1) = I) \rightarrow Say_I(p_\phi))) \rightarrow \alpha(\pi_1) \text{ resolves to } \phi$$

Such conventions express linguistic knowledge which competent communicators in Object Identification Games are assumed to have. They say that, if certain preconditions are met, an utterance α is used to say that ϕ. The formula can be read as follows: If both communicators I and C know ('$\mathcal{K}_{I,C}$') that if I utters α and if it is mutually believed ('$\mathcal{MB}_{I,C}$') that if I utters α she says that ϕ, then α resolves to ϕ.

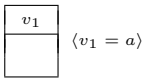
In our example dialogue, the sub-sentential utterance 'The red wooden disc' in (3-a) meets the preconditions by assumption. The relevant convention says that if I utters it, then it resolves to the directive addressed to C that she should grasp the object pointed at.

Recall that SDRT distinguishes *wrt* definites between *presupposed and asserted information*. Consequently, the utterance of (3-a) gives us the presupposed information π_{1p} in α and the asserted information π_{1a} in α. π_{1a} in α should be read as 'There is an SDRS but I don't know which one'. ϕ, in turn, expresses what the utterance α resolves to if the preconditions are met.

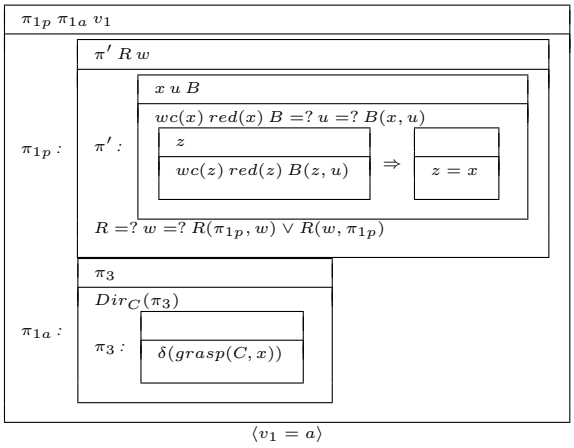




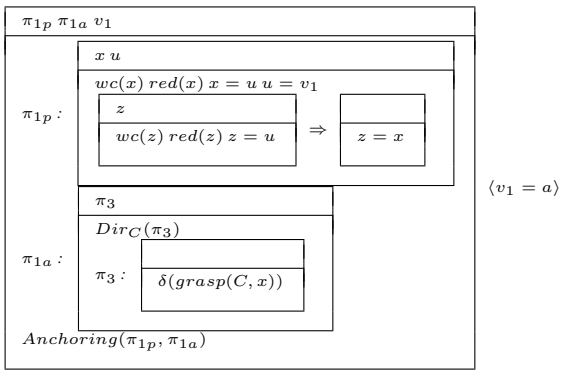
So, ϕ is what we get from the application of the linguistic information to the special convention. We assume a speech act theory style imperative semantics. Consequently, Dir_C is to be read as 'C is commanded that ...' and $\delta(grasp(C,x))$ in π_3 is the action commanded, namely that agent C grasp x. In a next step, the gestural information is represented. The pointing in (3-a) provides very little content. It merely relates some discourse referent to some external object:



Combining the linguistic and gestural information, the result of an apt multi-modal integration strategy is:

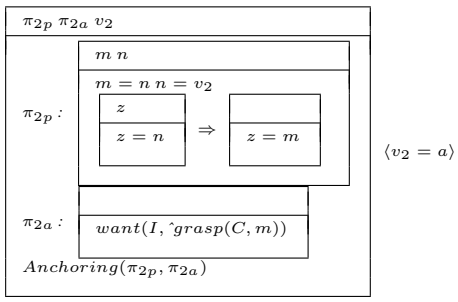


Now, underspecification can be resolved by using a tacit best-update-strategy. Thereby, we resolve the B-relation to identity ($\lambda x.\lambda y.x = y$), u to the externally anchored v_1, w to π_{1a} and R to $Anchoring$. Thus we get:

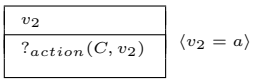


So, in the first turn I introduces a discourse referent v_1 which is externally anchored to the wooden disc a. The directive in π_{1a} presupposes that there is some object which can be grasped by C. The presupposition is satisfied through best-update's resolution of R to $Anchoring$ in such a way that π_{1p} anchors π_{1a}.

The next turn is analysed similarly. There is, likewise, a special convention regulating the interpretation of (3-b) which says that, if certain preconditions are met, when C utters the deictic 'Diese?' she thereby says that she wants to satisfy the directive. Combining the presupposed and the asserted content as before we get:



So, part of what the grasping does is that it externally anchors v_2. However, it seems that graspings have richer but underspecified content since they can be used to perform many things. We reflect this by assigning a highly underspecified content of type "action" to it:



In our dialogue, the grasping presumably carries out the action demanded by I. So, this suggests that the grasping in (3-b) is used to satisfy I's request in π_{1a} and part of its SARG. Using best update, this amounts to saying that $?_{action}(C, v_2)$ resolves to $grasp(C, v_2)$ and that $Sat-Request(\pi_2 g, \pi_1 a)$ holds. Thus the grasping elaborates on π_{2a} yielding

$Q-Elab(\pi_{2g}, \pi_{2a})$. Usual reasoning additionally gives us $Q-Elab(\pi_{1a}, \pi_{2a})$ and explains why 'This one?' in (3-b) is uttered. While the grasping satisfies the directive (see $Sat-Request$), it might not be *mutually believed* that it is satisfied. So, if $Q-Elab(\pi_{1a}, \pi_{2a})$ holds, it also mutually believed that it does (using SDRT's axiom schemata `Sincerity`, `Competence` and `Mutual Belief`). Moreover, by SARG-transitivity, the SARG of π_{1a} is (part of) the SARG of π_{2a}. Thus by satisfying π_{2a}'s SARG the SARG of π_{1a} is satisfied. So, finally, we get the resulting SDRS depicted in Fig. 3.

7 Related Research

Dealing with natural multi-modal dialogue in our paper, we touch on several research areas. Leaving out special SDRT literature here, the focus is on grammar-in-dialogue, description of fragments, and problems of integrating information from other channels.

The issue of syntax-in-dialogue was treated by Schegloff (1979) from the perspective of discourse analysis, mainly focussing on hesitations, restarts, turn construction, and repairs. Clark et al. (1990) generalised the ethno-methodological approach and studied cooperation in syntax production, formulating principles of cooperative contributions for NPs-in-dialogue. A corpus investigation from the perspective of syntax cooperation is provided in Skuplik (1999). Fine tuned coordination on all grammatical levels, named 'alignment', forms the backbone of Pickering and Garrod (2004)'s theory, completions and fragments being their favourite examples for establishing implicit common ground. Based on Skuplik (1999) and hooking up to SDRT, change of speaker roles, completions and inference in task-oriented dialogue were studied in Poncin and Rieser (2000) using Von Wright's Practical Syllogism and Asher and Morreau's Default Inference. A reconstruction of completions and similar phenomena within PTT is undertaken in Poesio and Rieser (2006). Recently, even if restricted to sentences/ propositions, the interest in retrieval and representation of fragmentary information has been growing, as the collection of articles in Elugardo and Stainton (2005) and their introduction to the volume shows. Above all, representation of ellipsis and fragmentary information has been investigated in the paradigm of Dynamic Syntax (Cann et al., 2005; Purver et al., 2005; Purver and Kempson, 2004) for some time, using advanced theory of grammar.

Since SDRT does not come with a worked out construction algorithm, it does not have a multimodal interface. Its contribution to multi-modality issues lies therefore in applying the separation of presuppositional versus assertional information and especially in the notion of anchoring. Principles of interface construction and compositionality matters concerning speech and gesture integration are discussed in Lücking et al. (2006), see also (Rieser, 2004; Rieser, 2005), where one can see which problems have to be overcome. Once the mapping from verbal expressions to SDRSs is organised, SDRT could, in principle, be part of an MM interface.

8 Ideas for Linking SDRT Logical Description Grammars (LDGs)

In (Asher and Lascarides, 2003, p. 122) it is assumed that some syntax-semantics-interface maps verbal input into ULF (underspecified logical forms), which, judging from the set-up of SDRT (p. 431), forms its bottom layer. Underspecified logical forms have models in the logic of information content, represented as SDRSs of some sort. In the simple case, where we have no underspecification, we get only one model. In order to get the mapping from language to ULF going, we can start from Muskens' concept of Logical Description Grammars (Muskens, 2001). LDGs use a version of Lexicalized Tree Adjoining Grammar (LTAG) which can capture underspecification in a similar way as the Constraint Language for Lambda Structures (Egg et al., 2001) does, for example concerning PP-attachment, quantifier ambiguity and polysemy. The semantic structures which we can use to tag LTAG-trees can be either type-logical formulae, as in (Muskens, 2001) or DRSs in the style of compositional DRT as in (Muskens, 1996). These we can take as substitutes for single SDRSs. Underspecification could arise due to syntactic structure or semantic ambiguity, *i.e.* we could get several SDRSs for one LTAG-formula. Once we reach this level, we seem to be done, since ULFs can be translated into Glue Logic, the place where the axioms substantiating admissible rhetorical relations are introduced.

We haven't yet tested this assumption in detail, we hope to report about it in the workshop. Ob-

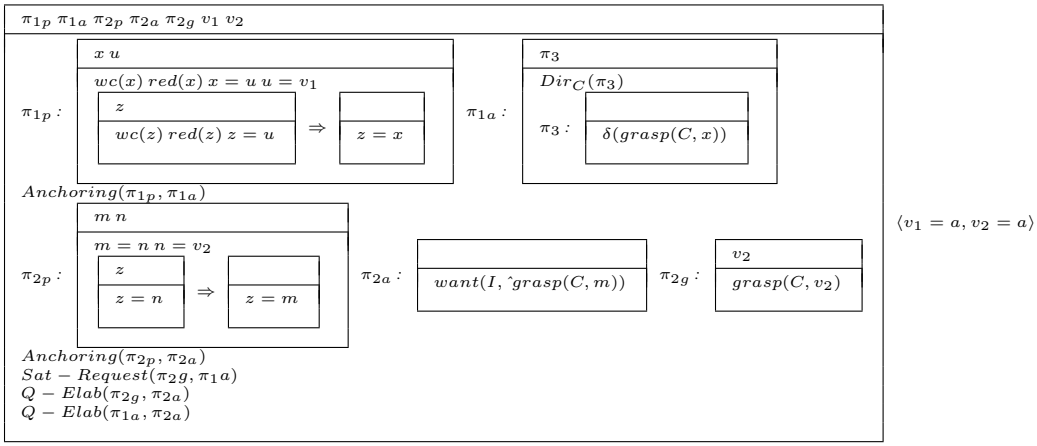


Figure 3: The resulting SDRS.

serve that with respect to our example we have to face additional problems due to the fragments encountered. As a consequence, we would have to use additional axioms in our mapping process.

References

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge.

Nicholas Asher. 2002. Deixis, Binding and Presupposition. forthcoming in: Festschrift for Hans Kamp.

Nicholas Asher. 2005. Bielefeld Lectures on SDRT.

Ronnie Cann, Ruth Kempson, and Lutz Martin. 2005. *The dynamics of language: an introduction*. Syntax and Semantics; 35. Elsevier, Amsterdam [a.o.].

Herbert H. Clark and Deanna Wilkes-Gibbs. 1990. Referring as a collaborative process. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intention in communication*, pages 463–493. MIT Press, Cambridge, MA.

Markus Egg, Alexander Koller, and Joachim Niehren. 2001. The constraint language for lambda structures. *Journal of Logic, Language, and Information*, 10(4):457–485.

Reinaldo Elugardo and Robert J. Stainton, editors. 2005. *Ellipsis and nonsentential speech*, volume 81 of *Studies in Linguistics and Philosophy*. Springer, Dordrecht [a.o.].

Hans Kamp. 1990. Prolegomena to a structural theory of belief and other attitudes. In C. Anthony Anderson and Joseph Owens, editors, *Propositional attitudes: the role of content in logic, language, and mind*. CSLI, Stanford.

Andy Lücking, Hannes Rieser, and Marc Staudacher. 2006. Multi-modal integration. Submitted to Brandial'06.

Reinhard Muskens. 1996. Montague semantics and discourse representation. *Linguistics and Philosophy*, 19:143–186.

Reinhard Muskens. 2001. Talking about trees and truth-conditions. *Journal of Logic, Language and Information*, 10(4):417–455.

Martin J. Pickering and Simon Garrod. 2004. Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.

Massimo Poesio and Hannes Rieser. 2006. Prolegomena to a theory of completions, continuations, and coordination in dialogue. Submitted.

Kristina Poncin and Hannes Rieser. 2000. Multi-speaker utterances and coordination in task-oriented dialogue. Revised and extended version of Gothenburg paper (Rieser and Skuplik (2000)). Report 2000/02, SFB 360, Univ. of Bielefeld, to appear in JoP 751. Technical Report 2000/06, SFB 360, Bielefeld University.

Matthew Purver and Ruth Kempson. 2004. Incremental parsing, or incremental grammar? In *Proceedings of the ACL Workshop on Incremental Parsing: Bringing Engineering and Cognition Together*, pages 74–81. Barcelona.

Matthew Purver, Ronnie Cann, and Ruth Kempson. 2005. Grammars as parsers: meeting the dialogue challenge. To appear in: Research on Language and Computation.

Hannes Rieser. 2004. Pointing in dialogue. In Jonathan Ginzburg and Enric Vallduví, editors, *Catalog '04. Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue*, pages 93–101. Barcelona.

Hannes Rieser. 2005. Pointing and grasping in concert. In Manfred Stede, Christian Chiarcos, Michael Grabski, and Luuk Lagerwerf, editors, *Salience in discourse: multidisciplinary approaches to discourse*, pages 129–139. Nodus Publikationen, Münster.

Emmanuel A. Schegloff. 1979. The relevance of repair to syntax-for-conversation. In Talmy Givón, editor, *Syntax and semantics: Discourse and syntax*, volume 12, pages 261–286. Academic Press, New York.

David Schlangen and Alex Lascarides. 2003. The interpretation of non-sentential utterances in dialogue. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*. Sapporo, Japan.

Kristina Skuplik. 1999. Satzkooperationen. Definition und empirische Untersuchung. Technical Report 1999/03, SFB 360, Bielefeld University.

Hyper-Document Structure: Representing Cognitive Coherence in Non-Linear Documents

Clara Mancini
Centre for Research in Computing
The Open University
Milton Keynes, MK7 6AA, UK
C.Mancini@open.ac.uk

Donia Scott
Centre for Research in Computing
The Open University
Milton Keynes, MK7 6AA, UK
D.Scott@open.ac.uk

Abstract

Hypertext introduces new possibilities for discourse organisation, as time and space can be exploited as new dimensions of document structuring. However, the technical characteristics of the non-linear medium pose a serious challenge to the representation of discourse coherence. We propose that, in the hypertext environment, graphics and animation can constitute a visual language to represent discourse coherence and support hyper-document construction.

1 Introduction

There is a long and well-established literature on textual devices that signal the coherence structure of a discourse to the reader, within both theoretical (e.g., van Dijk, 1977; Halliday and Hasan, 1976; Grimes, 1975; Brown and Yule, 1983) and computational linguistics (e.g., Hobbs, 1985; Mann and Thompson, 1988; Schiffrin, 1987; Knott and Mellish, 1996). Most of the work so far addresses the traditional conceptualisation of text as a two dimensional array on a physical page, traversed in a set pattern (e.g., left to right, top to bottom in the Western tradition).

However, hypertext is very different from traditional text: it is electronic, in that it can only be read on a computer screen, and it is non-linear, in that there are several paths available through the document. The reader moves from node to node by mouse-clicking on links. A node can be the equivalent of a traditional text page or can contain just a few sentences. A link can be a word in the text or a graphical element in the node. As nodes contain multiple links, the author can only partially control the order in which the reader will access them. In other words, with hypertext, a new conceptualisation of text has emerged as a three-dimensional array on a computer screen, which can be traversed in any number of ways.

The well-understood coherence markers of the traditional notion of text do not work for this new medium, therefore a new set of devices, not only textual but graphical, is needed together with formation rules to govern their usage, supported by sound theoretical frameworks. Here we explore new possibilities for constructing coherence in non-linear documents. Precisely because in non-linear documents discourse is organised as a network of self-standing units rather than as a hierarchy of interdependent segments, our analysis of discourse coherence departs from the tradition whereby text is described as a hierarchical structure (e.g., Mann & Thompson, 1988). Instead, we take a cognitive approach according to which coherence is a characteristic of the mental representation that the reader constructs during the process of text interpretation (e.g., Johnson-Laird, 1983).

2 Coherence representation in linear text

Understanding a text depends on the reader's ability to construct a coherent representation of its content. To do so the reader needs to be able to identify the conceptual relations holding between the set of discourse elements (sentences, paragraphs or entire text sections). In linear text this identification is facilitated by a number of cohesive elements. Over the years, the study of text coherence has concentrated on two types: those which function at the level of discourse structure, and those which function at the level of document structure. Much work has focussed on discourse structure. Whether data driven (Halliday and Hasan, 1976; Martin, 1992; Knott and Dale, 1994) or theory driven (Hobbs, 1985; Kamp and Ryle, 1993; Mann and Thompson, 1988; Sanders et al., 1993), this work has mainly

studied the use of discourse markers and referring expressions. Other work has highlighted the role played by graphical features such as punctuation and layout in text organisation, distinguishing text structure from syntactic structure (Nunberg, 1990). Text structure is defined by *concrete features*, such as punctuation and other graphical marks (parentheses, dashes, etc.), as well as by *abstract features*, such as layout and – we suggest – graphical formatting (titles, emphasis, etc.).

Elsewhere (Power, Scott, & Bouayad-Agha, 2003) we propose that both layout and formatting features deserve a separate descriptive level, which we term *abstract document structure*. This is an intrinsic part of text structure (Piwek, Power, Scott, & Van Deemter, 2005), but its constituents work differently from the way in which both discourse markers and concrete textual features work, because whereas discourse markers and punctuation are textual, devices like layout and formatting are visual.

3 Abstract discourse structure: visual vs. textual

Written text is a symbolic code in which the association between signifier and signified is arbitrary. This characteristic allows written text to explicitly express abstract concepts, which means that discourse markers can connect two text segments by explicitly expressing the relation holding between them. For instance, in the sentence "I was late for the meeting *because* I had missed the bus", the relation of causality holding between the segments is made explicit by the connective "because".

Its symbolic nature also implies that text can deploy along a single line, which can be articulated using punctuation, dashes, parentheses and the like (concrete textual features). These are purely graphical symbols, which signal different types of textual articulation and inflection, and whose use is also regulated by strict conventions.

Substantially different from both discourse markers and concrete textual features, abstract features transform the line of text into a visual configuration capable of conveying discourse structure on the space of the page. In visual configurations the association between a sign and its meaning is characterised by a degree of isomorphism, which makes this association partially motivated. For instance, in the sentence "I had a busy morning: I had a work meeting, I went for shopping, I picked up the children", the text segments in the list play an equivalent role within the sentence (Pander Maat, 1999). This rhetorical equivalence could be expressed as a bulleted list, in which the segments are given the same visual rendering: each segment starts on a new line with a bullet. Likewise, the title of the sections in a text will be visually more prominent than the title of the subsections in order to render the hierarchy of the text structure, just as emphasis is visually expressed through a format that stands out.

Unlike textual representations, visual representations tend to be regulated by conventions that are less strict and more dependent on the context of use. For instance, a list of clauses could be indented or not, bulleted, numbered or scored; whatever the chosen configuration, it is important that all listed clauses are rendered in the same way and occupy the same horizontal position. Even though they respond to flexible conventions, however, visual features can express discourse connections so effectively that the use of cue phrases or punctuation becomes redundant. So, in a bulleted list the use of connectives, commas and full stop is superfluous, as the conventions at work in the visual configuration of the list override the conventions that regulate the use of discourse connectives and punctuation.

4 Coherence in non-linear text

The devices described above constitute cohesive elements that can be used to express discourse coherence in linear text, either on paper or in electronic documents that maintain linearity. However, discourse markers such as relational and referential connectives can only be effectively used when discourse units are arranged in a predefined sequence, so that they are accessed in a univocal order. But because hypertext is a network of interconnected nodes, the order in which discourse parts will be accessed can only be partly controlled. Order can be established locally (a node can be linked to another node), but it is hardly possible to establish it globally through extended structures (unless one resorts to constrained paths, which would defeat the purpose of using a non-linear medium).

So, relational and referential connectives cannot be used to signal the discourse relation between nodes, because each node is accessible in more than one way. Consequently, hypertext nodes tend to be written as self-standing units of text. A hypertext node typically will not use pro-

nouns or referential phrases to refer to the content of another node, instead any information contained in the latter that would need to be referred to in the former has to be repeated. In fact, text sentences or paragraphs that are strongly related (for instance, by causality) will normally be kept within the same node: since they constitute strongly inter-dependent discourse parts, the writer is reluctant to put them in different nodes, because the reader might miss one or the other. However, it's less problematic to separate into different nodes discourse parts that are less strongly related (for instance, by elaboration or background) and therefore less inter-dependent can more easily be put into different nodes, their connection being expressed paratactically via a link (Mancini & Buckingham Shum, 2004). Finally, the same limitations that apply to discourse connectives also apply to punctuation and the like, which usually only work within nodes and do not facilitate the transition between link words and their target nodes.

If the non-linearity of hypertext does not lend itself to the use of discourse markers and concrete features, however, things are different for abstract document features, because they are visual and work in space. Because of its technical characteristics, hypertext is a spatial medium, and indeed numerous proposals that tackle the issue of non-linearity seek to compensate for the lack of control on discourse order by exploiting the spatial nature of hypertext. Some have proposed spatial metaphors as a way of describing discourse structure (Landow, 1991; Bolter, 1991; Kolb, 1997); others propose the use of maps, schemas, outlines (Carter, 2000) or navigational patterns (Bernstein, 1998) to return to the author's hands as much control as possible on the way in which discourse takes shape before the reader's eyes and coheres in their mind. But it is also a temporal medium, in which spatial structures have a temporal dimension and realisation (Luesebrink, 1998). So, both space and time can be exploited in hypertext to express discourse coherence and, we contend, in hypertext the notion of abstract document structure consists of both spatial and temporal configurations working in a three-dimensional space.

5 Hyper-document structure: managing space and time

If coherence is a cognitive phenomenon, then it is possible to express coherence relations not only through discourse markers, but also through visual patterns. And if this can be done by using spatial abstract features in linear documents, then it can also be done by using spatial and temporal abstract features in non-linear documents. In particular, we propose that graphics and animation could be used to express discourse coherence in hypertext (see Mancini & Buckingham Shum, 2004).

At present, most hypertexts (especially on the web) make no use of graphical features to signal rhetorical relations between nodes, and nodes often consist of long text pages with a few links targeting other pages, from where the source page can no longer be seen. However, we think that the non-linear medium could be used in a far more expressive and articulated way, if graphic features were exploited as discourse markers to support coherence. Our work precisely aims at identifying visual devices that can play the role of discourse markers in the non-linear, three dimensional space of hypertext.

One of these devices could consist of creating much smaller hypertext nodes and using the screen as a visual field across which they can distribute, as links are clicked and new nodes appear, composing meaningful patterns. The appearance and distribution of the nodes should signify the rhetorical role that their content plays within the discourse. To achieve that, rhetorical relations could be used as document structuring principles during discourse construction to define hypertext links. These could then be dynamically rendered during navigation through the consistent and concurrent use of the medium's spatial and temporal graphic features.

In this respect, having established a parallel between textual and visual processing (Riley & Parker, 1998), Gestalt theory has proposed useful principles of document design (Campbell, 1995). Furthermore, a number of representational rules for visually expressing discourse relations between hypertext nodes could be derived from the semiology of graphics, according to which graphic features can be employed to express conceptual relationships of *similarity*, *difference*, *order* and *proportion* exploiting the properties of the visual image, in a bi-dimensional static space (Bertin, 1967) as well as in a three-dimensional dynamic space (Koch, 2001). Using these rules, we have designed and begun testing a series of prototype visual patterns expressing coherence relations in non-linear discourse (Mancini, 2005).

6 Visualising rhetorical patterns

Based on cognitive parameterisations of coherence relations (Sanders et al., 1993; Pander Maat, 1999; Louwerse, 2001), we selected a set of relations for experimental rendering and evaluation. The set included: CAUSALITY, CONDITIONALITY, SIMILARITY, CONTRAST, CONJUNCTION, DISJUNCTION, ELABORATION and BACKGROUND. For the criteria of selection and for the discussion of all the renderings, see (Mancini, 2005). Here we report on three examples: CONJUNCTION, DISJUNCTION and CAUSALITY.

Relations	Basic Operation	Polarity
CONJUNCTION	additive	positive
DISJUNCTION	additive	negative
CAUSALITY	causal	positive

Table 1. Parametrical description of Conjunction, Disjunction and Causality (Sanders et al., 1993).

The graphical renderings of the relations were designed based on their parametrical description. In our descriptions of reference the bipolar parameters defining CONJUNCTION, DISJUNCTION and CAUSALITY were: *basic operation*, according to which a relation can be *causal* or *additive*, and *polarity*, according to which a relation can be *positive* or *negative*. The values of each cognitive parameter defining the relations (Table 1) were rendered through graphical features. As a result, each relation was visually defined by the sum of the graphical features rendering the cognitive values that define it. The graphical representation of CONJUNCTION was defined by the features rendering the values *additive* and *positive*. The representation of DISJUNCTION was defined by the features rendering the values *additive* and *negative*. The representation of CAUSALITY was defined by the features rendering the values *causal* and *positive*. The renderings of the values are described in Table 2.

To reify the relation renderings, examples of argumentative passages were taken from a history of science text, whose conceptual complexity and literary style were very accessible. Out of all the material provided by the book, a particular subject (theories about the orbiting of planets in the solar system) was selected, so that all the relations would be reified in the text within the same conceptual context. Short passages of text were then isolated, each passage consisting of a pair or a triple of sentences. The sentences of each pair or group held with each other one of the eight selected relations, all signaled by appropriate connectives. Finally, each pair or triple of related sentences was represented on screen respectively within a pair or triple of related text windows, and those windows were attributed certain graphical properties expressing the relation holding between the content of one sentence and the content of the other. On screen, all connective were removed from the text within the windows, and the connective function between the text spans was entirely delegated to the windows' graphical properties.

Parameter	Value	Rendering of each parametrical value
Basic Operation	additive	Windows aligned along horizontal axis. Same value throughout or at initial stage. Second window appearing next to the first or overlapped on one side.
	causal	Windows aligned along vertical axis. Gradual value intensification from one stage to the other. Second and third windows in turns slide down from behind respectively behind first and second.
Polarity	positive	Value intensification or stability, from appearance of one window to appearance of the other.
	negative	Value of the window appearing first in the visual field changes to contrast the value of the object appearing second.

Table 2. Description of the features used to design the parametrical values defining: conjunction, disjunction and causality.

In order to be as differentiated as possible, each representation had to be kept as minimalist as possible, making use of no more formal elements than strictly necessary. A small number of graphical variables (Koch, 2001) were used following specific rules of graphics. For a detailed discussion of the design process for all the relational renderings see (Mancini, 2005). Below is the description of three examples: CONJUNCTION, DISJUNCTION and CAUSALITY.

CONJUNCTION – In this relation two entities or phenomena coexist in the same place at the same time, but the reasons of their copresence is unspecified. However, in the context of their occurrence they play an equivalent role. In our example, the relation was reified by the text spans:

A. *Part of Newton's astronomical theory derives from Galileo's kinematic laws of falling bodies and projectiles, and from the completion of his principle of inertia.*

B. *Part of Newton's astronomical theory derives from Kepler's descriptive laws of planetary*

motion, and from the completion of his conception of gravitation.

The two respective text windows were given the same value and their vertical sides were given the same length; they appear on the screen next to each other, one at a time, the window containing the first text span appearing on the left and the window containing the second text span appearing on the right after 2 seconds. Firstly, the concept of *addition* was rendered by the windows appearing next to each other, with the order of appearance following the direction of reading that we are familiar with (in the Western culture). Secondly, the concept of *equivalence* was rendered by the value of the windows' areas, and reinforced by the fact that their sides were of identical length, and they appeared next to each other and not, say, one under the other. The way in which the windows positioned themselves was the simplest possible one, to render the fact that the two entities are related as complementary components of a whole (Figure 1).

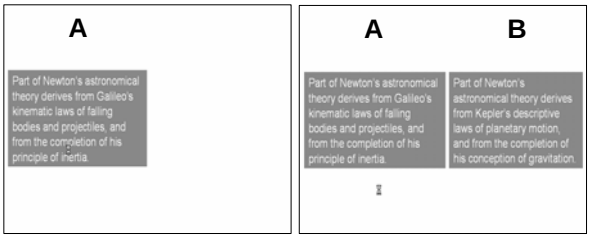


Figure 1. Two screen shots from the animated graphic rendering of conjunction (the letters above the text boxes are for illustration purposes only).

DISJUNCTION – In this relation two entities or phenomena do not coexist in a space-temporal interval, but are *alternative* to one another. The text spans selected to reify the disjunctive relation were:

A. *In Galileo's times, one could have embraced the heliocentric theory incurring the consequence of being considered a heretic by the Catholic Church.*

B. *In Galileo's times, one could have rejected the heliocentric theory and still have the chance of being considered a good Catholic.*

The text windows were given the same appearance as those used to represent the additive relation, with the difference that as the second window appeared on the right 2 seconds after, the window on the left had the value of its background changed to a very light grey, which made it difficult to read the text. The concept of reciprocal exclusion of the two situations, was rendered through the fact that, as the second span of text appeared, the first one would become unreadable (Figure 2).

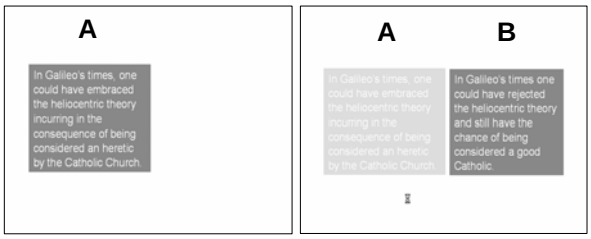


Figure 2. Two screen shots from the animated graphic rendering of disjunction (the letters above the text boxes are for illustration purposes only).

CAUSALITY - This is the strongest cognitive relation. It implies conjunction (the connected elements are part of the same context), sequence (one element necessarily follows the other) and the first element directly produces the second. The text spans, three this time, selected to reify causality were:

A. *Galileo ignored Kepler's demonstration of the elliptical orbits of planets and continued to believe that planetary revolutions were a "natural" motion requiring no external mover.*

B. *Galileo failed to see that the actual geometry of the heavens contradicted any spherical model.*

C. *Galileo missed the problem of how planets were retained in their elliptical orbits.*

The three windows respectively containing the three text spans were arranged one under the other, the second sliding down from behind the first as soon as the first had appeared, and the third sliding down from behind the second as soon as it had reached its position. They all shared the same width, while the height of each was determined by the quantity of text contained in each window. The value of the windows' background became increasingly darker from the first to the third, and the ratio of increment was the same from the first to the second and from the second to the third, that is, they were equidistant, as far as the value was concerned. In this configuration, the order of the events was rendered by the arrangement of the text windows, while the fact that the second and the third windows appeared by sliding down from the previous one rendered the fact that the second and the third events followed, and were brought about, respectively by the first and the second event. At the same time, the darkening of the background rendered the idea of progression in the forging of a logical chain. Finally, the cohesion between the three events was reinforced by the fact that the three windows had the same width (Figure 3).

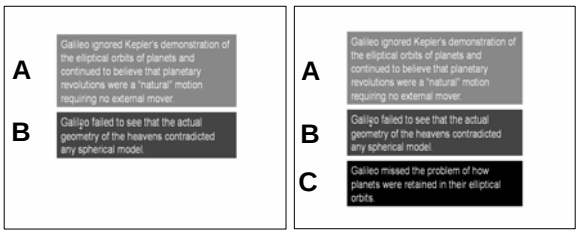


Figure 3. Two screen shots from the animated graphic rendering of causality (the letters beside the text boxes are for illustration purposes only).

The whole set of relations was rendered with the purpose of testing the renderings and their impact on users. In particular we wanted to find out whether the concurrent and consistent use of visual features according to certain perceptual principles and design criteria would determine the expressiveness of the configurations designed to represent the selected sub-set of discourse relations and whether people would discriminate the relational expressiveness of different visual configurations.

7 Testing the graphical renderings

As a first form of verification, we designed and conducted an empirical study with a group of **24** participants. We asked them to choose from three different representations the one that in their judgement best expressed each relational concept. For each relation, three different representations were presented to the participants: the one that had been designed to represent that particular relation, plus two alternative representations originally designed to express different relations.

One at the time, the participants were given the original text that had been used to reify each relation, as well as an abstract definition of the relation in question, then were shown the three animations associated with it, from which they had to choose what they thought to be its most expressive representation. They were asked to go through a second round, in which they were allowed to modify, one way or the other, the choices made in the first round.

R	Caus.	Cond.	Conj.	Disj.	Sim.	Cont.	Back.	Elab.
1st	19	10	18	12	16	20	21	20
2nd	**22**	**13**	**21**	**12**	**18**	**20**	**21**	**21**
χ^2	37	4.750	32.25	3.25	19.75	28	32.25	27.25
p	<0.001	N/S	<0.001	N/S	<0.001	<0.001	<0.001	<0.001

Table 3. Results of the experiment conducted with 24 participants, showing the renderings designed to respectively express each relation. 1st and 2nd = votes obtained by each rendering respectively in the first and in the second round.

For each given relation, the great majority of participants converged on the same option, which in fact corresponded to the animated pattern that had been specifically designed to render that particular relation. For 6 of the relations - CAUSALITY, CONJUNCTION, SIMILARITY, CONTRAST, ELABORATION, BACKGROUND - the results were statistically significant (see Table 3).

In brief, albeit not conclusive, the results of this first study suggest that people did recognize a particular expressiveness in the options that had been designed to render the subset of discourse coherence relations. In other words, there is positive evidence that the concurrent and consistent use of graphical elements, according to certain perceptual principles and design criteria, can support the visual expression of relational concepts.

The fact that for two of the relations - CONDITIONALITY and DISJUNCTION - the renderings did not obtain the same consensus obtained by the others could be explained with the fact that both conditionality and disjunction are characterized by a greater degree of cognitive complexity. From a cognitive point of view, CAUSALITY, CONJUNCTION, SIMILARITY, CONTRAST, ELABORATION and BACKGROUND hold within a space-temporal continuity, or along one possible line of events. However, conditionality and disjunction hold across two possible lines of events. That is, they implicate the cognitive projection into an alternative space-temporal dimension (or narrative axis), before the conditioned or disjuncted situations can be presented. Such an abstraction is easy to express in natural language, but it is not as easy to express in visual languages.

8 Future work

Still in its infancy, this work is at this stage more concerned with identifying the right questions than with presenting the right answers. However, our aim is to identify ways of presenting hypertext discourse which employ graphical features in a systematic and principled way, extending the notion of abstract document structure, so that it applies to hypertext as well as linear text, by making articulate use of the space-temporal dimensions of the electronic medium.

We have not implemented a system yet, but that is our goal, and the experimental results obtained so far are encouraging. As a next step we will be carrying out further tests on the visual renderings of rhetorical relations. For example,

we intend to test the same relational renderings with a larger number of participants from different backgrounds, carrying out a qualitative analysis of their responses. We have also started to construct hypertext mock-ups using our set of coherence relations to define the links between nodes and rendering the connections through their corresponding visual patterns. These are to be tested with users: as they navigate and visual patterns take shape on the screen, they will be asked to identify the relations holding between nodes, which will be indicated solely by the graphical clues. Further tests will also be designed.

Our long term goal is the application of this work to a larger effort in natural language generation, whereby the same semantic content is rendered differently for different readerships. In particular, we are generating paraphrases that vary not just along the traditional dimensions (discourse, syntax, lexicalisation) but also in terms of graphical presentation (e.g., as textual reports in different styles - including linear vs. non-linear - or as slides for a presentation).

9 Conclusion

If a reader is to understand a text, their mental representation of its content has to (at least to some degree) reflect the coherence structure intended by the writer. In linear documents, a number of textual devices signalling the coherence structure of discourse facilitate this process of reconstruction. However, these devices only work within a linear structure and they are no longer helpful in the interpretation of non-linear documents. When it comes to non-linear media, such as hypertext, a different set of signalling devices is required, which are visual rather than textual. These visual elements constitute the abstract document structure in traditional text, where they work within the bi-dimensional space of the page. However, in hypertext they have to work in a three-dimensional space as well as in time, which pushes the boundaries of the notion of abstract document structure. We have only begun to study this new concept of document structure and the principles that regulate the use of its features are yet to be established, but both theoretical and empirical work suggest that this is the route to follow, if we are to fully exploit the potential of electronic text.

Acknowledgments

We wish to acknowledge the reviewers of this paper for the useful comments.

Reference

Bernstein, M. (1998). Patterns of Hypertext. In *Proceedings of ACM Hypertext'98:* Pittsburgh, PA, ACM Press, New York, pp.21-29

Bertin, J. (1967). *Semiologie Graphique.* English translation (1983). *Semiology of Graphics: Diagrams, Networks, Maps.* University of Wiscounsin Press, Madison

Bolter, J.D. (1991). *Writing Space: The Computer, Hypertext, and the History of Writing*, Eastgate Systems, Cambridge MA

Brown, G. and Yule, G. (1983). Discourse Analysis. Cambridge University Press, New York

Campbell, K.S. (1995). *Coherence, Continuity, and Cohesion. Theoretical Foundations for Document Design*. Lawrence Erlbaum Associates Publishers, Hillsdale (NJ), Hove (UK)

Carter, L.M. (2000). Arguments in Hypertext: A Rhetorical Approach. In *Proceedings of ACM Hypertext '00*, ACM Press, New York, pp.87-91

Dijk, van T.A. (1977). Explorations in the Semantics and Pragmatics of Discourse. Longman, London - NY

Grimes, J.E. (1975). The Thread of Discourse. Mouton Publishers, Berlin, New York, Amsterdam

Halliday, M.A.K., Hasan, R. (1976). *Cohesion in English*. Longman

Hobbs, J.R. (1985). *On the Coherence and Structure of Discourse*. Technical Report CSLI-85-37, Stanford

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Kamp, H. and Ryle, U. (1993). *From Discourse to Logic.* Dordrecht: Kluwer.

Knott, A., Dale, R. (1994). Using Linguistic Phenomena to Motivate a Set of Coherence Relations. In *Discourse Processes*, 18(1), pp.35-62

Knott, A., Mellish, C. (1996). A feature-based account of the relations signalled by sentence and clause connectives. In *Language and Speech*, 39(2/3), pp.142-183

Koch, W.G. (2000/2001). Jaques Bertin's Theory of Graphics and its Development and Influence on Multimedia Cartography. In *Information Design Journal*, 10(1), pp.37-43

Landow, G.P. (1991). The Rhetoric of Hypermedia: Some Rules for Authors. In Delany, P., Landow, G.P. (Eds.) *Hypermedia and Literary Studies*, MIT Press, Cambridge, MA, pp.81-104

Kolb D. (1997). Scholarly Hypertext: Self-Represented Complexity. In *Proceedings of ACM Hypertext '97*, ACM Press, New York, pp.29-37

Louwerse, M. (2001). An Analytic and Cognitive Parametrization of Coherence Relations. In *Cognitive Linguistics*, 12 (3), pp. 291-315

Luesebrink, M. (1998). The Moment in Hypertext. In *Proceedings of ACM Hypertext '98*, ACM Press, pp.106-112

Mancini, C. (2005). Cinematic hypertext. Investigating a new paradigm. Amsterdam: IOS Press.

Mancini, C., Buckingham Shum, S. (2004). Towards Cinematic Hypertext. In *Proceeding of ACM Hypertext '04*, Santa Cruz, CA, USA, Aug 9-13, ACM Press, New York, pp.115-124

Mann, W.C., Thompson, S.A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organisation. In *Text*, 8 (3), pp.243-281

Martin, J.R. (1992). *English Text. System and Structure*. John Benjamins Publishing Co., Amsterdam

Nelson, T. (1981). *Literary Machines.* Swardmore

Nunberg, G. (1990). The Linguistics of Ponctuation. CSLI, Stanford, USA

Pander Maat, H. (1999). The Differential Linguistic Realisation of Comparative and Additive Coherence Relations. *Cognitive Linguistics, 10*, 147-184

Piwek, P., Power, R., Scott, D., van Deemter, K. (2005) Generating multimedia presentations: from plain text to screenplay Intelligent Multimodal Information Presentation, Text Speech and Language Processing, vol. 27 pp. 203-226 O. Stock and M. Zancanaro (ed.) Kluwer: Dordrecht

Power, R., Scott, D., Bouayad-Agha, N. (2003). Document Structure. *Computational Linguistics*, vol. 29 issue 4 pp. 211-260

Riley, K., Parker, F. (1998). Parallels between visual and textual processing. *IEEE Transactions on Professional. Communication, 41*, 175-185.

Sanders, T.J.M., Spooren, W.P.M., Noordman, L.G.M. (1993). Coherence Relations in a Cognitive Theory of Discourse Representation. In *Cognitive Linguistics*, 4(2), pp.93-133

Schiffrin, D. (1987). Discourse Markers. Cambridge University Press, New York

Tolva, T. (1996). Ut Pictura Hyperpoesis: Spatial Form, Visuality, and the Digital Word. In *Proceedings of ACM Hypertext '96*, ACM Press, New York, pp.66-73

Meaning and Dialogue Coherence: A Proof-theoretic Investigation

Paul Piwek
Centre for Research in Computing
The Open University
Milton Keynes, UK
p.piwek@open.ac.uk

Abstract

This paper present a novel proof-theoretic account of dialogue coherence. It focuses on cooperative information-oriented dialogues and describes how the structure of such dialogues can be accounted for in terms of a multi-agent hybrid inference system that mixes natural deduction with information transfer and observation. We show how the structure of dialogue arises out of the interplay between the inferential roles of logical connectives (i.e., sentence semantics), a rule for transferring information between agents, and rules for information flow between agents and their environment. Our order of explanation is opposite in direction to that adopted in the game-theoretic semantics tradition, where sentence semantics (or a notion of valid inferences) is derived from (winning) dialogue strategies. The approaches may, however, be reconcilable, since we focus on cooperative dialogues, whereas the latter concentrates on adversarial dialogue.

1 Introduction

Models of coherence come in many different shapes, from proposals based on scripts, grammars, and social rule following to models of topic continuity. A now slightly dated collection that provides an overview of the multitude of approaches to *dialogue coherence* is Craig and Tracy (1983). More recently, Mann (2002) surveys a number of extant analyses of dialogue coherence.

The aim of this paper is to work out in detail a notion of coherence for one particular type of cooperative dialogues, rather than to criticize or dismiss other approaches. In our view, coherence is a complex phenomenon that is likely to require analyses from more than one single perspective.

We provide an explication of dialogue coherence in terms of the meaning of the expressions that are used in a dialogue against the background of the participants' discursive dispositions. Thus, coherence is modelled as a property of dialogues whose meaning-bearing parts fit together in a certain way in context. Our analysis of dialogue coherence will only provide the foundations for certain aspects of dialogue coherence in general. At the end of this section, we specify the precise scope of the current proposal.

To construct an explication of dialogue coherence along these lines, we adopt the following strategy. Firstly, we describe a theory of meaning that provides the foundation for the current endeavour. This theory complies with the Wittgensteinian slogan that "meaning is use".[1] This pragmatist slogan is fleshed out by identifying the meaning of an expression with its role in reasoning. This role is given by the circumstances of appropriate *application* of the expression and the appropriate *consequences* of such an application. The meaning of logical vocabulary will be assigned a privileged status in this undertaking and receive a formalization in terms of a variant of Gentzen's (1934) calculus of Natural Deduction. Secondly, this standard Natural Deduction calculus for solitary reasoners is extended to a calculus for *multiple situated* reasoners. Thirdly, this extended Natural Deduction calculus is used to model dialogue coherence. Whereas Gentzen's

[1] Note that in an important respect our investigation is not Wittgensteinian; we do not share the later Wittgenstein's skepticism about the possibility of rigorous theories of language use.

calculus allows us to characterize valid inferences, the extended calculus demarcates a certain type of coherent dialogue. We provide examples of dialogues that are generated by the calculus and use these to bolster the initial plausibility of the claim that coherence according to the extended calculus mirrors coherence of natural language dialogue. This is achieved by drawing attention to a number of structural properties of naturally occurring dialogues that are also found in dialogues that are generated with the extended calculus.

Walton and Krabbe (1995) point out that dialogue comes in many varieties. Each variety has its own distinctive purpose and participant aims and, as a result, concomitant notion of coherence. We do not intend to define coherence regardless of dialogue type, but rather restrict our attention to a specific type of dialogue, which we call the *cooperative information-oriented dialogue*. The *main purpose* of is this type of dialogue is the exchange of information. The *participants' aim* is to cooperate with each others' requests for information; in particular, no persuasion, negotiation or coercion is required. Cooperative information-oriented dialogues have been a central topic of study in computational linguistics and natural language processing, e.g., witness the large number projects on dialogue systems for providing travel information.

Finally, a remark on the theoretical orientation of this paper. It is not concerned with directly accounting for instances of naturally occurring dialogues. Rather, our aim is to provide an abstract model of cooperative information-oriented dialogue that captures and accounts for certain abstract patterns that have been observed in naturally occuring dialogues by conversation analysts (Sudnow, 1972). We discuss these patterns, specifically adjacency pairs and insertion sequences, in more detail further on in this paper.

2 Meaning as Inferential Role

The theory of meaning that we employ follows broadly the meaning-theoretic deliberations of Brandom (1994). The formalization is along the lines described in Sundholm (1986),[2] though there are also important differences (see section 7). Meaning is characterized in terms of inferential role, rather than truth-conditions. For the purpose of this paper, no specific theory of truth is put forward; we get by without that notion. This does, however, not exclude the possibility of a reconstruction of truth in terms of the framework described in this paper. *Cf.* Brandom (1994).

We start with a framework involving a single agent, henceforth α. Our system captures the practical ability of this agent to reason with expressions of a language $\mathcal{L}$. This language consists exclusively of atomic formulae $At \subset \mathcal{L}$, *and* formulae that are constructed from formulae in $\mathcal{L}$ using the connectives for implication '$\rightarrow$' and conjunction '&': if $A, B \in \mathcal{L}$, then $(A \rightarrow B) \in \mathcal{L}$ and $(A\&B) \in \mathcal{L}$.[3]

Inferences are formalized in terms of *judgements* of the form $[\alpha]\ \ H \vdash A$. These should be read as agent α (henceforth, references to agents are omitted when it is clear from the context which agent the judgement belongs to) affirms/derives A, given the temporary assumptions H (i.e., assumptions that are only accessible for the duration of the inference). In addition to the collection of temporary assumptions (H), an agent, such as α, also relies on a set of persistent assumptions (Γ_α). In our system, Γ_α functions like a global variable in a programming language whose value is accessible any time during an inference. The value of Γ_α can be updated through declarations, as is common for global variables. For instance, the following declaration adds the proposition letter a to Γ_α's current value: $\Gamma_\alpha \ := \ \Gamma_\alpha \cup \{a\}$. Note that we use capitals (e.g., A and B) as meta-variables over proposition letters and lower case (e.g., a and b) for the actual proposition letters.

An assumption $A \in (H \cup \Gamma_\alpha)$ is thought of in terms of the disposition of α to affirm A. This disposition is made explicit by the following deduction rule:

(1) (member) $$\frac{A \ \in \ \Gamma \cup H}{H \ \vdash \ A}$$

This rule says that formula A can be inferred/derived/deduced from H and the implicit persistent assumptions Γ, *if* A is a member of the union of Γ and H. The set $\Gamma \cup H$ plays an inferential role in this system. This inferential role takes the place of a classical explication in terms of representation/truth-conditions.

[2] Sundholm bases much of his formalization on proposals by the philosophers/logicians Michael Dummett and Dag Prawitz.

[3] Henceforth, brackets will be omitted when there is no danger of ambiguity.

The inferential role of logical vocabulary is given a special place in the current scheme: a logical connective allows an agent to formulate explicitly a pattern of inference that it already follows. For instance, an agent who is disposed to deriving 'The tiles get wet' from 'it rains', can make this practical activity explicit by affirming 'If it rains, the tiles get wet'.

The meaning of the logical connective is given by the circumstances of appropriate application of that connective and the appropriate consequences of such an application. For the conditional '$\rightarrow$' the appropriate circumstances of application are given by the following rule for introducing a conditional:

(2) (arrow intro) $$\frac{H \cup \{A\} \vdash B}{H \vdash A \rightarrow B}$$

Thus, we can derive $A \rightarrow B$, if we can derive B from our assumptions extended with A. The appropriate consequences of using '$\rightarrow$' are given by the following rule for eliminating '$\rightarrow$':

(3) (arrow elim) $$\frac{H \vdash A \rightarrow B \quad H \vdash A}{H \vdash B}$$

This rule is chosen so that the arrow intro and elim rules together introduce only inferences regarding the logical connective '$\rightarrow$'. In Dummett's terms, the rules are in harmony with antecedent inferential practices. This requirement is essential, because of the explicative role of logical vocabulary: it should serve to make explicit existing inferential practices; it should not license novel inferences involving the pre-existing vocabulary, since that would destroy its explicative role with respect to that pre-existing vocabulary.

The rules for conjunction introduction and elimination are the following:

(4) (conj. intro) $$\frac{H \vdash A \quad H \vdash B}{H \vdash A\&B}$$

(5) (conj. elim) $$\frac{H \vdash A\&B}{H \vdash A} \qquad \frac{H \vdash A\&B}{H \vdash B}$$

3 System $\mathcal{S}_1$: Situated Inferential Practice and Dialogue

The system presented so far is limited to solitary reasoners that are isolated both from other reasoners and the world around them. In this section, we present an extension which removes the former limitation. We will refer to the system described in the current section as $\mathcal{S}_1$.

3.1 The Transfer Rule

We introduce a set of agents $\mathcal{A}$. We use α, β, γ, ... as meta-variables over members of $\mathcal{A}$. We can now a add rule for *transferring* proof goals between agents:

(6) (tr) $$\frac{[\beta]\ H \vdash A}{[\alpha]\ H \vdash A} \quad \Gamma_\alpha := \Gamma_\alpha \cup \{\bigwedge H \rightarrow A\} \text{ and } \langle \alpha, \beta \rangle \in \mathcal{C}$$

This transfer rule (tr) tells us that if agent β can derive A under the assumptions in H, then agent α can also derive A under the assumptions in H, provided that the two side conditions (given on the right-hand side) are satisfied. The first condition says that the context Γ_α of assumptions entertained by α, should be extended with $\bigwedge H \rightarrow A$. Here, $\bigwedge H$ stands for the conjunction $A_1 \& A_2 \& \ldots$ of the formulae $A_1, A_2, \ldots$ that are members of H. If H is empty, $\bigwedge H \rightarrow A = A$. The second condition ($\langle \alpha, \beta \rangle \in \mathcal{C}$) says that there should be a transfer channel between α and β (where $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$). We use $\mathcal{C}$ in combination with this side condition to model situations in which not every agent can exchange information with every other agent in $\mathcal{A}$. However, unless stated otherwise, we will henceforth assume that $\mathcal{C} = \mathcal{A} \times \mathcal{A}$, i.e., information can be transferred between any pair of agents.

3.2 Example of a Proof Tree

Take a situation involving the agents α, β and γ in which $\Gamma_\alpha = \emptyset$, $\Gamma_\beta = \{a\}$, and $\Gamma_\gamma = \{b\}$. Let us assume that α wants to build a proof for $a\&b$. Since neither a nor b is part of Γ_α, α will need to access information held by β and γ. The following proof tree illustrates how exactly:

(7)

$$\dfrac{\dfrac{\dfrac{[\beta]\ a \in \Gamma_\beta \cup \emptyset}{[\beta]\ \emptyset \vdash a}\ (mem.)}{[\alpha]\ \emptyset \vdash a}\ (1)(tr) \qquad \dfrac{\dfrac{\dfrac{[\gamma]\ b \in \Gamma_\gamma \cup \emptyset}{[\gamma]\ \emptyset \vdash b}\ (mem.)}{[\alpha]\ \emptyset \vdash b}\ (2)(tr)}{[\alpha]\ \emptyset \vdash a\&b}\ (conj.intro)$$

SIDE CONDITIONS: (1) $\Gamma_\alpha := \Gamma_\alpha \cup \{a\}$; (2) $\Gamma_\alpha := \Gamma_\alpha \cup \{b\}$.

Note that we omitted the side condition regarding the transfer channel. We conveniently assumed that all agents can exchange information with all other agents.

Execution of the two side conditions results in $\Gamma_\alpha = \{a, b\}$. As a result of the construction of this

proof, we have arrived at a Γ_α in which a proof for $a\&b$ can be constructed directly, without recourse to the transfer. In other words, $a\&b$ has become part of α's information.

3.3 From Proof Trees to Dialogue Structure

The last stage consists of the transformation of the proof tree to a dialogue (structure). Here we provide an outline of the algorithm, which has been fully implemented.[4] We proceed in two steps. Firstly, we map the hierarchical tree to a linear structure where each tree node is represented by an item in the linear structure. (e.g., items 4 and 9 below each represent a single node; the nodes in question are the terminal nodes of tree 7), and possibly a second item indicating that the part of the tree dominated by the node has been closed (e.g., the pairs $\langle 1, 12\rangle$ and $\langle 2, 6\rangle$ represent single tree nodes; the former corresponds to the root node of tree 7). For the tree in 7, we obtain the following linear representation:

(8)
1. α : goal-know-if($a\&b$)
2. α : (transfer) goal-know-if(a)
3. β : goal-know-if(a)
4. β : in-assumptions(a)
5. β : confirmed(a)
6. α : confirmed(a)
7. α : (transfer) goal-know-if(b)
8. γ : goal-know-if(b)
9. γ : in-assumptions(b)
10. γ : confirmed(b)
11. α : confirmed(b)
12. α : confirmed($a\&b$)

For brevity's sake, we have omitted reference to the empty set of temporary hypotheses. Strictly, speaking we should, for instance, have written goal-know-if($a\&b$,given-that,$\emptyset$) instead of goal-know-if($a\&b$).

The sequence in 8 is not yet a straightforward dialogue. It contains various locutions which can be thought of as internal monologues of the interlocutors with themselves, but not actual dialogue locutions. For example 6. α : confirmed(a), is superfluous after 5. β : confirmed(a). a can be taken to have been confirmed by α implicitly, simply by α proceeding with the dialogue.

For the mapping from an extensive dialogue representation, such as 8, to a more economic dialogue structure we use the following rules:

- α_i : goal-know-if(A) $\mapsto$ α_i : I am wondering whether A.
- α_i : (transfer) goal-know-if(A), α_j : I am wondering whether $A \mapsto \alpha_i$: Tell me α_j, A?
- α_i : confirmed(A), α_j : confirmed(A) $\mapsto$ α_i : confirmed(A).
- α_i : in-assumptions(A), α_i : confirmed(A) $\mapsto$ α_i : A.
- α_i : confirmed(A) $\mapsto$ α_i : That confirms A.

When these mapping rules are applied to 8, we obtain:

(9)
1. α : I am wondering whether $a\&b$.
2. α : Tell me β, a?
3. β : a.
4. α : Tell me γ, b?
5. γ : b.
6. α : That confirms $a\&b$.

Note that this dialogue structure exhibits two well-known conversation analytical configurations (Sudnow, 1972): the adjacency pairs (2,3), (4,5) and (1,6) and the insertion sequence (2,3,4,5).

4 Generative Systems as Abstract Models of Dialogue

Before we proceed with presenting some extensions to the system $\mathcal{S}_1$, let us take a step back and make explicit what such systems have in common. Each of them functions as an abstract model of cooperative information-oriented dialogue, and has the following components:

1. A hybrid inference system $\mathcal{I}$ consisting of:
   (a) A language $\mathcal{L}$ (e.g., the language of propositional logic or a fragment thereof);
   (b) A set of agents $\mathcal{A}$, each with a set of assumptions Γ_{α_i};
   (c) A communication channel $\mathcal{C}$ that specifies which agents can communicate with what other agents (i.e., $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$);
   (d) A set of hybrid inference rules $\mathcal{R}$ for the language and the agents. The rules are hybrid because they can encompass natural deduction, observation and communication. The rules enable us to build proof trees (or proof search trees, as we will see in a moment).

[4] See `mcs.open.ac.uk/pp2464/resources`

2. A specification of the set of potential dialogues $\mathcal{D}_P$ between the agents, given the language $\mathcal{L}$.

3. A mapping m from proof trees, generated with $\mathcal{I}$, to coherent dialogues $\mathcal{D}$.

In short, a generative system $\mathcal{S}$ is a tuple of the form $\langle \mathcal{I}, \mathcal{D}_P, m \rangle$. The purpose of such a system is the characterization of coherent dialogues (members of $\mathcal{D}$). This is achieved by using $\mathcal{I} = \langle \mathcal{L}, \mathcal{A}, \mathcal{C}, \mathcal{R} \rangle$ to generate proof trees, that is trees representing valid inferences (or searches for valid inferences). These proof trees are then mapped by m to members of $\mathcal{D}_P$ (the set of potential dialogues). A member of $\mathcal{D}_P$ that can be generated from a proof tree using m is a member of the set of proper, i.e., coherent, dialogues $\mathcal{D}$ (with $\mathcal{D} \subset \mathcal{D}_P$). The mapping m basically turns a hierarchical proof tree into a linear dialogue representation (omitting most proof steps that do not involve communication between agents).

We investigate systems for the generation of abstract representations of coherent dialogues. The adequacy of such systems can be thought of in terms of their correctness and completeness:

- CORRECTNESS: Each member of $\mathcal{D}$ generated by $\mathcal{S}$ should represent the structure of a coherent dialogue. Here, coherence is understood roughly speaking in terms of our pre-theoretical understanding of dialogue coherence.
- COMPLETENESS: 1. GLOBAL: Each structure of a coherent dialogue (again, we refer to our pre-theoretical insight into dialogue coherence) should be generated by $\mathcal{S}$, that is, it should be a member of $\mathcal{D}$. 2. LOCAL: Each structure of a coherent dialogue that is a member of $\mathcal{D}_P$ should be generated by $\mathcal{S}$, that is, it should be a member of $\mathcal{D}$.

These notions of correctness and completeness are to guide the construction of the generative systems. For each system, we will attempt to satisfy correctness. As we construct further systems, the main aim will be to add features that make the new system better approximate either local or global completeness.

5 System $\mathcal{S}_2$: From Proof Trees to Proof Search Trees

System $\mathcal{S}_1$ has one major drawback: it only allows for dialogues generated from completed proof trees. What is lost, is the *search* for a proof which many cooperative information-oriented dialogues revolve around. In short, System $\mathcal{S}_1$ is not globally complete: people in conversation will often explore unfruitful paths, and have to use locutions such as: 'I could not resolve the question whether A?' or 'I don't know whether A'. A first step toward remedying this situation is the addition of such locutions to $\mathcal{D}_P$. Let us be precise, and spell out the set of potential dialogues $\mathcal{D}_P$ using BNF notation:

(10) $\langle D_P \rangle$ `:==` $\langle Loc \rangle$`,`$\langle D_P \rangle$ | ϵ
$\langle Loc \rangle$ `:==` $\langle Agent \rangle$`: I am wondering whether` $\langle Prop \rangle$ | $\langle Agent \rangle$`: Tell me` $\langle Agent \rangle$`,` $\langle Prop \rangle$`?` | $\langle Agent \rangle$`: I don't know whether` $\langle Prop \rangle$`.` | $\langle Agent \rangle$`:` $\langle Prop \rangle$`.` | $\langle Agent \rangle$`: That confirms` $\langle Prop \rangle$`.`
$\langle Agent \rangle$ `:==` $\alpha \mid \beta \mid \ldots$
$\langle Prop \rangle$ `:==` $a \mid b \mid \ldots \mid \langle Prop \rangle \& \langle Prop \rangle \mid \langle Prop \rangle \rightarrow \langle Prop \rangle$

Now, the problem is one of local completeness: there are now members of D_P which are intuitively coherent dialogues (involving the locution '$\langle Agent \rangle$: I don't know whether $\langle Prop \rangle$'), but which cannot be generated in $\mathcal{S}_1$. To address this problem, we first need to define the mapping m for proof *search* trees, rather than proof trees.

Let us examine an example of a proof search tree.

(11)

$$\dfrac{\dfrac{\dfrac{}{(\star_1)\ [\beta]\ \emptyset \vdash a} \quad \dfrac{[\gamma]\ a \in \Gamma_\gamma \cup \emptyset}{(\star_2)\ [\gamma]\ \emptyset \vdash a}}{[\alpha]\ \emptyset \vdash a} \quad \dfrac{\dfrac{[\gamma]\ a \in \Gamma_\gamma \cup \emptyset}{[\gamma]\ \emptyset \vdash a}}{[\alpha]\ \emptyset \vdash b}}{[\alpha]\ \emptyset \vdash a \& b}$$

This tree is very similar to the proof tree 7. We have omitted rule labels and conditions to fit the tree on this page. That 11 is a proof *search* tree rather than a proof tree, is indicated by the use of $\star$, which marks alternative search branches. Here we have an unsuccessful branch $\star_1$ and a successful one, i.e., $\star_2$ (henceforth we assume that successful search branches are always to the right of unsuccessful ones). This tree would, for example, fit the situation where we initially set out with $\Gamma_\alpha = \Gamma_\beta = \emptyset$ and $\Gamma_\gamma = \{a, b\}$.

As before, we map the tree to a dialogue in two steps. The result of applying the first mapping is:

(12)
1. α : goal-know-if($a\&b$)
2. α : (transfer) goal-know-if(a)
3. β : goal-know-if(a)
4. β : not-resolved(a)
5. α : not-resolved(a)
6. α : (transfer) goal-know-if(a)
7. γ : goal-know-if(a)
8. γ : in-assumptions(a)
9. γ : confirmed(a)
10. α : confirmed(a)
11. α : (transfer) goal-know-if(b)
12. γ : goal-know-if(b)
13. γ : in-assumptions(b)
14. γ : confirmed(b)
15. α : confirmed(b)
16. α : confirmed($a\&b$)

The second half of the mapping requires the mapping rules of $\mathcal{S}_1$ and two additional rules:

- α_i : not-resolved(A), α_j : not-resolved(A) $\mapsto$ α_i : not-resolved(A)
- α_i : not-resolved(A) $\mapsto$ α_i : I don't know whether A.

Application of the extended set of mapping rules to 12 results in:

(13)
1. α : I am wondering whether $a\&b$.
2. α : Tell me β, a?
3. β : I don't know whether a.
4. α : Tell me γ, a?
5. γ : a.
6. α : Tell me γ, b?
7. γ : b.
8. α : That confirms $a\&b$.

6 System $\mathcal{S}_3$: Adding Observation

In $\mathcal{S}_1$ and $\mathcal{S}_2$, we went beyond common inference systems, by moving from the model of a solitary reasoner to a community of reasoners (agents) who can exchange information with each other. Communication is, however, not the only way reasoners can acquire new information. In particular, observation of the environment is a further means of information acquisition that the traditional model of logical inference does not deal with. In this respect, our systems $\mathcal{S}_1$ and $\mathcal{S}_2$ are also incomplete.

It is beyond the scope of this paper to address all the intricacies of interspersing reasoning with observation. What we do offer is an outline of how observation can be integrated with multi-agent inference. We explore a minimal extension of $\mathcal{S}_2$ with the following rule:

(14) (obs.) $$\frac{A \in \mathcal{O}_\alpha \quad obs(\alpha, A)}{[\alpha]H \vdash A} \quad \Gamma_\alpha := \Gamma_\alpha \cup \{A\}$$

This rules states that if the proposition A is an observable proposition for agent α (written as $A \in \mathcal{O}_\alpha$) and α actually observes that A (i.e., $obs(\alpha, A)$), then α can derive A. There is one side condition which requires that Γ_α is extended with A; that is: $\Gamma_\alpha := \Gamma_\alpha \cup \{A\}$.

To give an idea of how the extended calculus can be applied, we model a conversation between π and δ. π has the flu and rings δ (information desk of π's surgery) to find out whether she needs to see a doctor (sd). For that purpose, she needs to find out whether she has a temperature (ht). We have $\Gamma_\pi = \emptyset$ and $\Gamma_\delta = \{ht \rightarrow sd\}$. π's goal is to find out whether sd. So π tries to derive relative to her assumptions that sd. Given Γ_π, π won't succeed unless she decides to communicate. The following is a derivation for π of sd that involves both communication and observation. It highlights how information possessed by δ and observations that only π is able to perform in the situation that we described are combined to obtain a proof for sd:

(15)

$$\dfrac{\dfrac{\dfrac{ht \in \mathcal{O}_\pi \quad obs(\pi, ht)}{[\pi]\,\emptyset \vdash ht}\,(3)\,(obs.)}{[\delta]\,\emptyset \vdash ht}\,(2)\,(tr) \qquad \dfrac{[\delta]\,ht \rightarrow sd \in \Gamma_\delta}{[\delta]\,\emptyset \vdash ht \rightarrow sd}}{\dfrac{[\delta]\,\emptyset \vdash sd}{[\pi]\,\emptyset \vdash sd}\,(1)\,(tr)}$$

SIDE CONDITIONS: (1) $\Gamma_\pi := \Gamma_\pi \cup \{sd\}$, (2) $\Gamma_\delta := \Gamma_\delta \cup \{ht\}$, (3) $\Gamma_\pi := \Gamma_\pi \cup \{ht\}$.

As a result of the proof construction, $\Gamma_\pi = \{ht, sd\}$ and $\Gamma_\delta = \{ht \rightarrow sd, ht\}$. Note that although $sd \notin \Gamma_\delta$, δ can now infer sd without recourse to observation or communication. From the proof tree, we can read off the moves of dialogue 16. The dialogue contains a well-known conversation analytical structure, i.e., the insertion sequence (the subdialogue consisting of 3 and 4):

(16)
1. π: Do I need to see a doctor?
2. δ: Do you have a temperature?
3. π: Wait a minute [π checks her temperature], yes, I do.
4. δ: Then you do need to see a doctor.

7 Related Work

In this section we contrast the approach described in this paper with other related approaches. Firstly, note that the extended Natural Deduction calculus that we employed required a number modifications to the standard calculus employed by, for instance, Sundholm (for specifying inferential roles). We introduced a distinction between temporary and persistent assumptions and used the member rule to access both types of assumptions. Persistent assumptions were introduced to collect premises from sources other than inference (i.e., communication and observation). Another crucial extension was the explicit relativization of judgements and assumptions to agents.

Secondly, our approach differs in a number of respects from extant models of dialogue. Here we compare our approach with two representative classes of alternatives. Firstly, there is a body of work based on the idea that dialogues can be characterized in terms of information states in combination with update and generation rules. The difference with our approach is that we try to explain dialogue coherence in terms of independently motivated inferential roles of logical constants Compare this with, for example, (Beun, 2001) who introduces special purpose generation rules to achieve the same effect as our intro and elim rules for '$\rightarrow$', the conversational procedures in the pioneering work by Power (1979), the up- and downdating rules for the partially ordered questions under discussion in Ginzburg (1996), and the generic framework for information state-based dialogue modelling described in Traum & Larsson (2003). For a comparison of some of these existing approaches see Pulman (1999).

Secondly, there is a dialogue game approach going back to the work of Lorenzen – see Lorenzen & Lorenz (1978) – where the logical constants are defined in terms of their role in rational debates. There the order of explanation is from (a) formal winning strategies for *adversarial* dialogues (debate) to (b) valid patterns of reasoning involving the logical constants.[5] In contrast, we proceed from (b) valid patterns of reasoning involving the logical constants to (c) coherent *cooperative* dialogue. The undertakings are complementary and raise the, rather surprising, prospect of an account of cooperative dialogue based on adversarial dialogue (debate); that is, an account from (a) to (c) via (b).

[5]Hamblin (1971) also explores derivations in this direction: *from* a specification of legal dialogue – though his dialogues are information-oriented, rather than adversarial – *to* semantic properties of locutions. Furthermore, the game-theoretical semantics that has been developed by Hintikka and collaborators (Saarinen, 1979) has some central features in common Lorenzen's dialogue games.

8 Limitations and Further Research

The aim of this paper is to provide the foundations for a generative logic-based model of dialogue coherence. The generic framework is described in section 4, whereas specific systems are developed in sections 3, 5 and 6. The purpose of these systems was to demonstrate that the type of analysis advocated here can account for certain dialogue structures. The systems are, however, limited in a number of ways. In this section we identify these limitations, and provide some suggestions on how to address them.

(i) Inconsistencies between participants' informational states are avoided by the use of a minimal logic that lacks negation. In future work, we plan to add inference rules for negation, and investigate the implication of such an extension, in particular, with regards to interactions with the observation rule. (ii) We assume that dialogue participants always successfully perform speech acts: the communication channel is perfect (no misperception) and the language is fully shared and free of ambiguous expressions. (iii) Communication is mostly direct: speakers express what they mean by saying it, rather than by Gricean implicature (Grice, 1975). We intend to address implicature by extending the system with non-standard patterns of inference (e.g., default reasoning) without changing the inferential roles of the logical vocabulary. We intend to achieve this by means of accommodation rules operating on the sets of persistent assumptions; *cf.* chapter 2 of Piwek (1998). (iv) The current systems only deal with information-oriented dialogues. In future, we would like to investigate application of the current framework to task-oriented dialogues that involve imperatives and actions. (v) The current systems are based on proposition logic. We plan to develop further systems that incorporate more expressive logics, in particular, the predicate calculus. For this purpose, we will build on an implementation of a system for natural deduction for predicate and higher order logics (Piwek, 2006), and the work on consistency maintenance in type theory-based

natural deduction systems by Borghuis and Nederpelt (2000). (vi) Beun (2001) points out that his system needs to be extended with less elegant rules to prevent generation of dialogues with loops (e.g., by not allowing an agent to ask the same question twice). Our system is infected with the same problem. To avoid classifying repetitive dialogue structures as coherent, we need a rule that prevents an agent from transferring the same proof goal to the same agent more than once. (vii) Currently, our framework is set up so that proof (search) trees are produced first and then mapped to dialogue (structures). This provides us with a theoretically clean and transparent framework for relating inference systems to dialogue structure. The work also has practical potential, for example, as a framework for generating information presentations in dialogue form; see the discussion of dialogue as discourse in Piwek & Van Deemter (2002). Nevertheless, there is also scope for investigating how the mapping rules can be integrated with proof search, thus making it possible to use the resulting system in human-computer dialogue.

9 Conclusion

The current paper is foundational in nature. We show how to model dialogue coherence in terms of generative systems that rely on an extended calculus of Natural Deduction. At the core of this account is the standard Natural Deduction calculus which has been motivated independently. The paper presents extensions of the calculus with rules for communication and observation, and describes a mapping from proof (search) trees to dialogue structures. We hope that the current paper will stimulate discussion about the role of logic and sentence semantics in understanding dialogue coherence.

Acknowledgements

I would like to thank the three anonymous reviewers of the ESSLLI workshop on "Coherence in Dialogue and Generation" for helpful comments and suggestions.

References

R.J. Beun. 2001. On the Generation of Coherent Dialogue: A Computational Approach. *Pragmatics & Cognition*, 9(1).

T. Borghuis and R. Nederpelt. 2000. Belief Revision with Explicit Justifications: An Exploration in Type Theory. CS-Report 00-17, Eindhoven University of Technology.

R. Brandom. 1994. *Making It Explicit: reasoning, representing, and discursive commitment*. Harvard University Press, Cambridge, Mass.

R. Craig and K. Tracy, editors. 1983. *Conversational Coherence: Form, Structure and Strategy*. Sage Publications, Beverly Hills.

G. Gentzen. 1934. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, 39:176–210, 405–431.

J. Ginzburg. 1996. Dynamics and the Semantics of Dialogue. In *Language, Logic and Computation*, volume 1. CSLI, Stanford.

H.P. Grice. 1975. Logic and conversation. In Peter Cole and Jerry Morgan, editors, *Syntax and Semantics 3: Speech Acts*, pages 64–75. Academic Press, New York.

C.L. Hamblin. 1971. Mathematical Models of Dialogue. *Theoria*, 37:130–155.

P. Lorenzen and K. Lorenz. 1978. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt.

B. Mann. 2002. What is dialogue coherence? Memo available at http://www-rcf.usc.edu/ billmann/WMlinguistic/dcoherence.htm, June.

P. Piwek and K. van Deemter. 2002. Towards automated generation of scripted dialogue: Some time-honoured strategies. In *EDILOG 2002: Proceedings of the sixth workshop on the semantics and pragmatics of dialogue*, pages 141–148. The University of Edinburgh, September.

P. Piwek. 1998. *Logic, Information and Conversation*. Ph.D. thesis, Eindhoven University, The Netherlands.

P. Piwek. 2006. The ALLIGATOR Theorem Prover for Dependent Type Systems: Description and Proof Sample. In *Proceedings of the Inference in Computational Semantics Workshop (ICoS-5)*, Buxton, UK.

R. Power. 1979. The organisation of purposeful dialogues. *Linguistics*, 17.

S. Pulman. 1999. Relating Dialogue Games to Information States. In *Proceedings of the European Speech Communication Association workshop on Dialogue and Prosody*, pages 17–24, De Koningshof, The Netherlands.

E. Saarinen, editor. 1979. *Game-theoretical Semantics*. D. Reidel, Dordrecht.

D. Sudnow, editor. 1972. *Studies in Social Interaction*. The Free Press, New York.

G. Sundholm. 1986. Proof Theory and Meaning. In *Handbook of Philosophical Logic*, volume III, pages 471–506. D. Reidel.

D. Traum and S. Larsson. 2003. The Information State Approach to Dialogue Management. In *Current and New Directions in Discourse and Dialogue*, pages 325–353. Kluwer Academic Publishers.

D. Walton and E. Krabbe. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press, New York.

✶ ✶ ✶