Optimizing Referential Coherence in Text Generation

Rodger Kibble∗
University of London

Richard Power†
University of Brighton

This article describes an implemented system which uses centering theory for planning of coherent texts and choice of referring expressions. We argue that text and sentence planning need to be driven in part by the goal of maintaining referential continuity and thereby facilitating pronoun resolution: Obtaining a favorable ordering of clauses, and of arguments within clauses, is likely to increase opportunities for nonambiguous pronoun use. Centering theory provides the basis for such an integrated approach. Generating coherent texts according to centering theory is treated as a constraint satisfaction problem. The well-known Rule 2 of centering theory is reformulated in terms of a set of constraints—cohesion, salience, cheapness, and continuity—and we show sample outputs obtained under a particular weighting of these constraints. This framework facilitates detailed research into evaluation metrics and will therefore provide a productive research tool in addition to the immediate practical benefit of improving the fluency and readability of generated texts. The technique is generally applicable to natural language generation systems, which perform hierarchical text structuring based on a theory of coherence relations with certain additional assumptions.

1. Overview

A central task for natural language generation (NLG) systems is to produce text that is *coherent*, in the sense in which (1a) is noticeably more coherent than (1b):

1. a. Elixir is a white cream.
        It is used in the treatment of cold sores.
        It contains aliprosan.
        Aliprosan relieves viral skin disorders.
    b. Elixir contains aliprosan.
        Viral skin disorders are relieved by aliprosan.
        Elixir is used in the treatment of cold sores.
        It is a white cream.

We can observe various ways in which text organization influences coherence: the sequence in which certain facts are presented, the order in which entities are mentioned in a clause, and the possibilities available for identifying the intended reference of pronouns. Generally, (1a) seems to conform better to a reader's expectations of what will be referred to next and of how to resolve underspecified referring expressions,

---

∗ Department of Computing, Goldsmiths College, University of London, London SE14 6NW, U.K. E-mail: r.kibble@gold.ac.uk

† Information Technology Research Institute, University of Brighton, Brighton BN2 4GJ, U.K. E-mail: Richard.Power@itri.brighton.ac.uk

in particular pronouns. These are issues which the well-known centering theory (CT) of Grosz, Joshi, and Weinstein (1995; henceforth GJW) is concerned with. Previous algorithms for pronominalization such as those of McCoy and Strube (1999), Henschel, Cheng, and Poesio (2000), and Callaway and Lester (2002) have addressed the task of deciding whether to realize an entity as a pronoun on the basis of given factors such as its syntactic role and discourse history within a given text structure; what is essentially novel in our approach is that we treat referential coherence as a *planning* problem, on the assumption that obtaining a favorable ordering of clauses, and of arguments within clauses, is likely to increase opportunities for nonambiguous pronoun use. Centering theory provides the basis for such an integrated approach.[1]

Of course coherence of a text depends on the realization of **rhetorical relations** (Mann and Thompson 1987) as well as referential continuity, and the latter is to an extent a byproduct of the former, as clauses that are rhetorically related also tend to mention the same entities. However, even when a set of facts is arranged in a hierarchical RST structure, there are still many possible linear orderings with noticeable differences in referential coherence. This article concentrates on the influence of referential continuity on overall coherence and describes a method for applying CT to problems in text planning and pronominalization in order to improve the fluency and readability of generated texts. This method is applicable in principle to any system which produces hierarchically structured text plans using a theory of coherence relations, with the following additional assumptions:

- There is a one-to-one correspondence between predicates and verbs, so that the options for syntactic realization can be predicted from the argument structure of predicates. Such "shallow" lexicalization appears to be standard in applied NLG systems (Cahill 1999).

- Pronominalization is deferred until grammatical relations and word order have been determined.

Our exposition will refer to an implemented document generation system, ICONOCLAST, which uses the technique of **constraint satisfaction** (van Hentenryck 1989; Power 2000; Power, Scott, and Bouayad-Agha 2003) with CT principles implemented among a set of soft constraints. The ICONOCLAST system allows the user to specify content and rhetorical structure through an interactive knowledge-base editor and supports fine-grained control over stylistic and layout features. The user-determined rhetorical structure is transformed into a **text structure** or a set of candidate text structures which respect various text formation rules encoded as hard constraints. Not all of the resulting text structures will give rise to stylistically acceptable documents, and of those which may be judged acceptable, some will be noticeably preferable to others. The text-structuring phase is followed by an evaluation of the candidate structures in which they are ranked according to a set of preferences encoded as soft constraints. Centering preferences are weighted along with other stylistic constraints to fix the preferred final ordering both of propositions in the text and of arguments within a clause.

It is not our primary aim in this short article to provide an empirical assessment of the claims of CT, for which we refer the reader to the relevant papers, such as

---

1 Callaway and Lester (2002) note that CT-based pronominalization algorithms "assume that the discourse tree was constructed with Centering theory in mind" (page 91); in our case this assumption is justified.

those collected in Walker, Joshi, and Prince (1998a) as well as Poesio et al. (2002) and other works cited there. We report elsewhere (Kibble and Power 2004) on two ongoing empirical studies: A paired-comparison study of judgments by naive subjects indicates that centering constraints make an appreciable difference to the acceptability of texts, and a corpus study using what we believe to be a novel technique involving **perturbations** provides clear evidence of preferences between the different constraints. One of the strengths of our framework is that it can be used as a research tool for the evaluation of variants of CT, as different realizations of an input sequence can be generated by varying control parameters, and one can very quickly see the results of alternative choices.

1.1 Related Work

Other researchers have applied CT to generation, though to our knowledge none have applied it to text planning, sentence planning, and pronominalization in the integrated way that we present in this article. This general approach is anticipated by McKeown's (1985) text-planning system, in which referential coherence is taken to be one of the factors determining fluency, though McKeown's work predates RST and centering. Mittal et al. (1998) apply what we term **salience** to sentence planning, with the goal of realizing the Cb as subject, though the text planner does not have a goal of attempting to maintain the same Cb. We regard Cheng's (2000) work on the interaction of centering preferences and aggregation in text planning as complementary to our enterprise. Karamanis (2001), Kibble (2001), and Beaver (2004), have argued for a *ranking* of the centering principles as opposed to weighting, and indeed Beaver provides a unified formulation of the centering rules and constraints as a ranked set of OT constraints. However, we believe that such a ranking stands in need of empirical justification, and Beaver's data actually provide little evidence for strict ranking as opposed to weighting of constraints (see Kibble 2003). Constraint satisfaction search was applied by Marcu (1996, 1997) to the far harder task of constructing RST trees given a set of facts and a repertoire of rhetorical relations; Mellish et al. (1998) argue that this approach may not scale up to the generation of larger texts and propose an alternative using **stochastic search**. We address the issue of computational complexity in section 4; however we do not face the same problems as Marcu, since the task for our text planner is to convert a given RST tree into a (possibly singleton) set of text structures rather than to build the RST tree from scratch.

2. Centering Parameters

We assume some familiarity with the basic concepts of CT. In this section we briefly and informally summarize the main assumptions of the theory and explain how we have interpreted and applied these assumptions:

1. For each utterance in a discourse there is said to be at most one entity that is the center of attention or **center** (Constraint 1). The center in an utterance U_n is the most highly ranked entity realized in U_{n-1}, which is also realized in U_n (Constraint 3). This is also referred to as the **backward-looking center** or **Cb**. (The set of entities mentioned in an utterance U_n is defined by Constraint 2 as the set of **forward-looking centers** or **Cfs**.) It is not entirely clear whether Constraint 1 is to be taken as an empirical claim or as a *stipulation* that some entity must be designated as Cb, if necessary by constructing an indirect anaphoric link.

2. There is a preference for consecutive utterances within a discourse segment to keep the same entity as the center and for the center to be realized as the highest-ranked entity or **preferred center** (**Cp**). Kibble (1999) dubbed these principles **cohe-**

Table 1
Centering transitions.

Continue	Cohesion and Salience both hold; same center (or $Cb(U_n)$ undefined), realized as Cp in U_{n+1}
Retain	Cohesion only; that is, center remains the same but is not realized as Cp in U_{n+1}
Smooth Shift	Salience only; center of U_{n+1} realized as Cp but not equal to $Cb(U_n)$
Rough Shift	Neither cohesion nor salience holds

sion and **salience**, respectively. Combinations of these preferences provide the familiar canonical set of **transitions** shown in Table 1, ranked in the stipulated order of preference first set out as Rule 2 by Brennan, Friedman, and Pollard (1987) and adopted by Walker, Joshi, and Prince (1998b).

3. The center is the entity which is most likely to be pronominalized: GJW's Rule 1 in its weakest form states that if any entity is referred to by a pronoun, the Cb must be.

As Poesio et al. (2002) point out, CT can be viewed as a "parametric" theory in that key notions such as **utterance** and **previous utterance**, **realization** of entities, and **ranking** are not given precise definitions by GJW, and subsequent applied studies have had to begin by fixing particular instantiations of these notions.

2.1 Ranking
Since Brennan, Friedman, and Pollard (1987), a ranking in terms of grammatical roles (or **obliqueness**) has become standard; for example: SUBJECT > DIRECT OBJECT > INDIRECT OBJECT > OTHERS.

We have simplified matters somewhat for the purposes of this implementation. First, we assume that syntactic realization serves only to distinguish the Cp from all other referents, which are ranked on the same level: Thus effectively SUBJECT > OTHERS. Secondly, we assume that the system already knows, from the argument structure of the proposition, which entities can occur in subject position: Thus in realizing a proposition *ban(fda, elixir)*, both arguments are potential Cps because active and passive realizations are both allowed; for *contain(elixir, gestodene)*, only *elixir* is a potential Cp because we disallow *Gestodene is contained by Elixir*.

2.2 Realization
GJW's original formulation distinguished between "direct" realization, or **coreference**, and "indirect" realization, which corresponds to **bridging reference**. As an example, in (1a) the terms *cold sores* and *viral skin disorders* are not strictly coreferential and so do not count as direct realizations of the same entity, but if we allow indirect realization, then there is the potential for one of these to be identified as Cb, in a sequence such as *Elixir is used to treat cold sores. Viral skin disorders are relieved by aliprosan*. Again, we keep things simple at this stage by treating nominal expressions as realizations of the same entity only if they strictly corefer. As Poesio et al. (2002) observe, under this interpretation of realization, a number of utterances will lack an identifiable Cb, so we have to allow for a "no-Cb" transition in addition to the canonical transitions listed in Table 1.[2]

---

2 Of course, even with indirect realization we would still have to allow for the possibility of no-Cb transitions.

2.3 Utterance and Previous Utterance

Two different approaches to the realization of "utterance" have become associated with the work of Kameyama (1998) and Suri, McCoy, and DeCristoforo (1999). To simplify somewhat: Kameyama argued that the local focus is updated in a linear manner by tensed clauses rather than by sentences, while Suri, McCoy, and DeCristoforo present evidence that the subject of the main clause in a complex sentence is likely to be the preferred antecedent for a subject pronoun in an immediately following sentence, winning out over candidates in an intervening subordinate clause, as in example (2):

2. Dodge^i was robbed by an ex-convict^j the other night.
   The ex-convict$_j$ tied him$_i$ up because he$_i$ wasn't cooperating.
   Then he$_j$ took all the money and ran / #he$_i$ started screaming for help.

In fact we would argue that Suri, McCoy, and DeCristoforo's analysis does not establish whether the accessibility effects are due to the syntactic or the rhetorical structure of utterances. The examples they present all involve sentences of the form *Sx because Sy* corresponding to the rhetorical pattern *nucleus–connective—satellite*. Their results are therefore consistent with the hypothesis that the nucleus of a preceding segment is more accessible than the satellite. We allow the user of our system to choose between two strategies: a linear, Kameyama-style approach or a hierarchical approach in which the utterance is effectively identified with a rhetorical span. Our approach is more general than that of Suri, McCoy, and DeCristoforo as it covers cases in which the components of a complex rhetorical span are realized in different sentences. Veins theory (Cristea, Ide, and Romary 1998) provides a possible formalization of the intuition that some earlier propositions become inaccessible as a rhetorical boundary is crossed. The theory could be applied to centering in various ways; we have implemented perhaps the simplest approach, in which centering transitions are assessed in relation to the nearest *accessible* predecessor. In many cases the linear and hierarchical definitions give the same result, but sometimes they diverge, as in the following schematic example:

3. *ban(fda, elixir)* since *contain(elixir, gestodene)*.
   However, *approve(fda, elixirplus)*.

Following Veins Theory, the predecessor of *approve(fda, elixirplus)* is *ban(fda, elixir)*; its linear predecessor *contain(elixir, gestodene)* (an embedded satellite) is inaccessible. This makes a considerable difference: Under a hierarchical approach, *fda* can be the Cb of the final proposition; under a linear approach, this proposition has no Cb.

2.4 Transitions versus Constraints

Kibble (1999, 2001) argued for a decomposition of the canonical transition types into the principles of cohesion and salience, partly on the architectural grounds that this makes it easier to apply CT to the generation task, and partly on the empirical grounds that the preference ordering assumed by GJW is not strongly supported by corpus evidence and that transitions are better seen as **epiphenomenal**, emerging in a partial ordering from the interaction of more fundamental constraints. We follow this general approach, including among the constraints the principle of **continuity**: Each utterance should have at least one referent in common with the preceding utterance, which is effectively a restatement of GJW's Constraint 1. If we assign a weight of 1 each to cohesion and salience and 2 to continuity, we obtain a partial ordering over the

canonical transitions as follows:

$$0 : \text{CONTINUE} > 1 : \{\text{RETAIN} \mid \text{SMOOTH SHIFT}\} > 2 : \{\text{ROUGH SHIFT} \mid \text{NO CB}\}$$

Any relative weighting or ranking of coherence over salience would need to be motivated by evidence that Retain is preferred over Smooth Shift, and we are not aware of any conclusive evidence of this in the literature (see Kipple [1999] for further discussion).

This approach also means that Strube and Hahn's (1999) principle of **cheapness** can be naturally incorporated as an additional constraint: This is a requirement that $Cp(U_{n-1}) = Cb(U_n)$. The principle of cheapness effectively cashes out the informal definition of the Cp as "represent[ing] a prediction about the Cb of the following utterance" (Walker, Joshi, and Prince, 1998b, page 3). In classic variants of centering theory, this happens only indirectly as a result of transition preferences, and only following a Continue or Smooth Shift, since the Cp is also the Cb and Rule 2 predicts that the preferred transition will maintain the same Cb. However, the prediction is not entailed by the theory following a Retain, Rough Shift, or no-Cb transition or indeed for the first sentence in a discourse, when there is effectively no prediction concerning the Cp. Strube and Hahn claim that the cheapness principle is motivated by the existence of Retain-Shift patterns, which are evidently a common means of introducing a new topic (see also Brennan, Friedman, and Pollard 1987 [henceforth BFP]). To summarize, our system incorporates the following constraints:

cohesion: $Cb(U_{n-1}) = Cb(U_n)$

salience: $Cp(U_n) = Cb(U_n)$

cheapness: $Cp(U_{n-1}) = Cb(U_n)$

continuity: $Cfs(U_{n-1}) \cap Cfs(U_n) \neq \emptyset$

2.5 Preferences: Transitions, Pairs, or Sequences?

The original version of GJW's Rule 2 specified that *sequences* of Continue transitions are preferred over sequences of Retains, and so on; in BFP's implementation, however, transitions are evaluated incrementally and the preference applies to individual transitions such as Continue versus Retain rather than to sequences. Strube and Hahn (1999) take an intermediate position: In their formulation, pairs of transitions $\langle \langle U_i, U_j \rangle, \langle U_j, U_k \rangle \rangle$ are preferred that are cheap, that is, $Cp(U_j) = Cb(U_k)$. Strube and Hahn intended the preference for cheap transition pairs to replace GJW's Rule 2 in toto, which seems a rather weak requirement. On the other hand the original GJW formulation is difficult to verify, since as Poesio et al. (2002, page 66) found, sequences of multiple occurrences of the same transition type turn out to be relatively rare. Our position is a little more complex, as we do not directly aim to generate particular transitions or sequences of transitions but to minimize violations of the constraints *continuity*, *cohesion*, *salience*, and *cheapness*. Violations are computed on individual nodes and summed for each candidate text structure, so we may expect that the candidate with the fewest violations will have a preponderance of the preferred transitions. The system is certainly more slanted toward global optimization than BFP's incremental model but may be said to achieve this in a more natural way than a strategy of trying to produce uniform sequences of transitions.

2.6 Pronominalization

GJW's Rule 1 is rather weak as a guide to pronominalization decisions in general, as it only mentions the Cb and gives little guidance on when or whether to pronomi-

nalize non-Cbs. An important consideration for NLG is to minimize the possibility of ambiguity, and so we adopt a cautious strategy: The user can choose between invariably pronominalizing the *Cb* or using a fairly simple algorithm based on parallelism of grammatical roles. A possible future development is to supplement our CT-based text planner with a more sophisticated pronominalization algorithm as proposed by Henschel, Cheng, and Poesio (2000) or Callaway and Lester (2002).

3. Generation Issues

CT has developed primarily in the context of natural language interpretation, focussing on anaphora resolution (see, e.g., Brennan, Friedman, and Pollard 1987). As stated above, the novel contribution of this article is an integrated treatment of pronominalization and planning, aiming to determine whether the principles underlying the constraints and rules of the theory can be "turned round" and used as planning operators for generating coherent text. We have assumed some familiarity in the foregoing 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):

Text planning/content determination: deciding the content of a message and organizing the component propositions into a text structure (typically a tree)

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 realization: surface details such as agreement and orthography

Reiter observed that these functions can often be identified with discrete modules in applied NLG systems and that a de facto standard had emerged in which these modules are organized in a **pipeline** such that data flows only in one direction and only between consecutive modules.

Breaking down the generation task in this way makes it evident that there are various ways the distinct principles of CT can be incorporated. Continuity and cohesion naturally come under text planning: respectively, ordering a sequence of utterances to ensure that each has a backward-looking center and maintaining the same entity as the center within constraints on ordering determined by discourse relations. Salience and cheapness, on the other hand, would come under sentence planning, since in each case a particular entity is to be realized as subject. However, we encounter an apparent paradox in that identifying the center itself depends on grammatical salience as determined by the sentence planner: for example, choice of active or passive voice. Consequently, the text planner appears to rely on decisions made at the sentence-planning level, which is incompatible with the fact that "pipelined systems cannot perform general search over a decision space which includes decisions made in more than one module" (Reiter 2000, page 252).

We can envisage three possibilities for incorporating CT into a generation architecture:

1. "Incremental" sentence-by-sentence generation, in which the syntactic structure of U_n is determined before the semantic content of U_{n+1} is planned. That is, the text planner would plan the content of U_{n+1} by aiming to realize a proposition in the knowledge base which mentions an entity which is salient in U_n. We are not aware
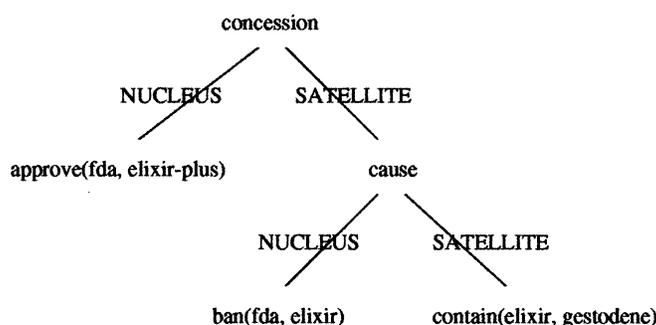
407

Figure 1
Rhetorical structure.

of any system which performs all stages of generation in a sentence-by-sentence way, and in any case this type of architecture would not allow for global planning over multisentence sequences, which we take to be essential for a faithful implementation of centering.

2. A pipelined system in which the "topic" or "theme" of a sentence is designated independently as part of the semantic input and centering rules reflect the information structure of a discourse. Prince (1999) notes that definitions of *topic* in the literature do not provide objective tests for topichood and proposes that the topic should be identified with the center of attention as defined by CT; however, what would be needed here would be a more fundamental definition that would account for a particular entity's being chosen to be the center of attention in the first place.

3. The solution we adopt is to treat the task of identifying Cbs and Cps as an optimization problem. We assume that certain options for syntactic realization can be predicted on the basis of the argument structure of predicates, which means that centering constructs can be calculated as part of text planning *before* syntactic realization takes place, so that the paradox noted above is resolved. Pronominalization decisions are deferred until a point at which grammatical relations and word order have been fixed.

4. Generation as Constraint Satisfaction

In this section we give an overview of our text-planning component in order to set the implementation of CT in context. The methodology is more fully described by Power, Scott, and Bouayad-Agha (2003).

The text planner was developed within ICONOCLAST, a project that investigated applications of constraint-based reasoning in natural language generation using as subject matter the domain of medical information leaflets. Following Scott and de Souza (1990), we represent rhetorical structure by graphs like Figure 1, in which nonterminal nodes represent RST relations, terminal nodes represent propositions, and linear order is unspecified. The task of the text planner is to realize the rhetorical structure as a text structure in which propositions are ordered, assigned to textual units (e.g., sentences, paragraphs, vertical lists), and linked where appropriate by discourse connectives (e.g., *since, however*). The boundary between text and sentence planning is drawn at the realization of elementary propositions rather than at the generation of individual sentences. If a rhetorical subtree is realized as a complex sentence, the effect

is that "text planning" trespasses into the higher-level syntax of the sentence, leaving only the elementary propositions to be realized by "sentence planning."[3]

Even for a simple rhetorical input like figure 1, many reasonable text structures can be generated. Since there are two nucleus-satellite relations, the elementary propositions can be ordered in four ways. Several discourse connectives can be employed to realize each rhetorical relation (e.g., concession can be realized by *although, but,* and *however*). At one extreme, the text can be spread out over several paragraphs, while at the other extreme, it can be squeezed into a single sentence. With fairly restrictive constraint settings, the system generates 24 text structure patterns for figure 1, including the following (shown schematically):

A. Since *contain*(*elixir, gestodene*), *ban*(*fda, elixir*).
   However, *approve*(*fda, elixirplus*).

B. *approve*(*fda, elixirplus*), although since *contain*(*elixir, gestodene*),
   ban(*fda, elixir*).

The final output texts will depend on how the propositions are realized syntactically; among other things, this will depend on centering choices within each proposition.

In outline, the procedure that we propose is as follows:

1. Enumerate all text structures that are acceptable realizations of the rhetorical structure.

2. For each text structure, enumerate all permissible choices for the Cb and Cp of each proposition.

3. Evaluate the solutions, taking account of referential coherence among other considerations, and choose the best.

For the example in figure 1, centers can be assigned in four ways for each text structure pattern, making a total of 96 solutions.

As will probably be obvious, such a procedure could not be applied for rhetorical structures with many propositions. For examples of this kind, based on the relations *cause* and *concession* (each of which can be marked by several different connectives), we find that the total number of text structures is approximately 5^{N-1} for N propositions. Hence with $N = 5$, we would expect around 600 text structures; with perhaps five to ten ways of assigning centers to each text structure, the total number of solutions would approximate to 5,000. Global optimization of the solution therefore becomes impracticable for texts longer than about five propositions; we address this problem by a technique of **partial optimization** in which a high-level planner fixes the large-scale structure of the text, thus defining a set of local planning problems, each small enough to be tackled by the methods described here.

Stage 1 of the planning procedure is described in more detail by Power, Scott, and Bouayad-Agha (2003). A brief summary follows, after which we focus on stages 2 and 3, in which the text planner enumerates the possible assignments of centers and evaluates which is the best.

---

3 See Power, Scott, and Bouayad-Agha (2003) for detailed motivation of this concept of text structure as a level of representation distinct from both rhetorical structure and syntactic structure.

4.1 Generating and Evaluating Text Structures
A text structure is defined in ICONOCLAST as an ordered tree in which each node has a feature named TEXT–LEVEL. Values of TEXT–LEVEL are represented by integers in the range $0 \ldots L_{max}$; these may be interpreted in various ways, but we will assume here that $L_{max} = 4$ and that integers are paired with descriptive labels as follows:

0   text phrase
1   text clause
2   text sentence
3   paragraph
4   section

Informally, a text structure (TS) is well-formed if it respects the hierarchy of textual levels, so that sections are composed of paragraphs, paragraphs of text sentences, and so forth. An example of an ill-formed structure would be one in which a text sentence contained a paragraph; such a structure can occur only when the paragraph is indented—a possibility we are excluding here. As well as being a well-formed text structure, a candidate solution must realize a rhetorical structure (RS) "correctly," in a sense that we need to make precise. Roughly, a correct solution should satisfy three conditions:

1.   The terminal nodes of the TS should express all the elementary propositions in the RS; they may also contain discourse connectives expressing rhetorical relations in the RS, although for some relations discourse connectives are optional.

2.   The TS must respect rules of syntax when it combines propositions and discourse connectives within a text clause; for instance, a conjunction such as *but* linking two text phrases must be coordinated with the second one.

3.   The TS must be structurally compatible with the RS.

The first two conditions are straightforward, but what is meant by "structural compatibility"? We suggest the crucial criterion for such compatibility should be as follows: Any grouping of the elementary propositions in the TS must also occur in the RS. In other words, the text structurer is allowed to eliminate groupings, but not to add any. More formally:

- If a node in the TS dominates terminal nodes expressing a set of elementary propositions, there must be a corresponding node in the RS dominating the same set of propositions.

- The converse does not hold: For instance, an RS of the form $R_1(R_2(p_1, p_2), p_3)$ can be realized by a paragraph of three sentences, one for each proposition, even though this TS contains no node dominating the propositions p_1 and p_2 that are grouped by R_2. However, when this happens, the propositions grouped together in the RS must remain consecutive in the TS; solutions in which p_3 comes in between p_1 and p_2 are prohibited.

Table 2
Examples of text-structuring constraints.

Name	Type	Description
Root domination $L_p > L_d$	Hard	The TEXT–LEVEL of the root node r must exceed that of any daughter d.
Parental domination $L_p \geq L_d$	Hard	The TEXT–LEVEL of a parent node p must be equal to or greater than the TEXT–LEVEL of any daughter d.
Sister equality $L_a = L_b$	Hard	If nodes a and b are descended from the same parent, they must have the same TEXT–LEVEL.
Sister order $O_a \neq O_b$	Hard	If nodes a and b are descended from the same parent, they must have different values of ORDER.
Connective	Hard	Governs choice of discourse connective.
Rhetorical grouping	Soft	Failure to express a rhetorical grouping can be treated as a defect.
Oversimple paragraph	Soft	A paragraph containing only one text sentence can be treated as a defect.
Centering	Soft	Constraints derived from centering theory.

Our procedure for generating candidate solutions is based on a technique for formulating text structuring as a constraint satisfaction problem (CSP) (van Hentenryck, 1989), using the ECLIPSE logic programming environment.[4] In general, a CSP is characterized by the following elements:

- a set of variables $V_1 \ldots V_N$

- For each variable V_i, a finite domain D_i of possible values

- a set of constraints on the values of the variables (for integer domains these often use "greater than" and "less than"; other domains usually rely on "equal" or "unequal".)

A solution assigns to each variable V_i a value from its domain D_i while respecting all constraints. For instance each node of the rhetorical structure is annotated with a TEXT–LEVEL variable with the domain $0 \ldots L_{\max}$ and an ORDER variable with the domain $1 \ldots N$, where N is the number of sisters. Depending on the constraints, there may be multiple solutions, or there may be no solution at all. We distinguish between **hard constraints**, which are applied during the enumeration phase, determining which candidate structures will be considered, and **soft constraints**, which apply during an evaluation phase in which the enumerated solutions are ordered from best to worst. Some examples of hard and soft constraints are shown in Table 2.

4.2 Choosing Centers
Given a text structure, we enumerate all permissible centering assignments as follows:

1. Determine the predecessor U_{n-1} (if any) of each proposition U_n.

2. List the potential Cbs and Cps of each proposition (henceforth denoted by ΣCb and ΣCp).

---

4 See http://www-icparc.doc.ic.ac.uk/eclipse/.

Table 3
Cbs and Cps for solution A.

U	Proposition	$\Sigma Cb(U)$	$\Sigma Cp(U)$
U_1	*contain(elixir, gestodene)*	[ ]	[*elixir*]
U_2	*ban(fda, elixir)*	[*elixir*]	[*fda, elixir*]
U_3	*approve(fda, elixir-plus)*	[*fda*]	[*fda, elixir-plus*]

3. Compute all combinations from ΣCb and ΣCp that respect the fundamental centering constraint that $Cb(U_n)$ should be the most salient candidate in U_{n-1}.

As stated earlier, two criteria for determining the predecessor have been implemented; the user can select one or the other criterion, thus using the NLG system to test different approaches. Following a **linear** criterion, the predecessor is simply the proposition that precedes the current proposition in the text, regardless of structural considerations. Following a **hierarchical** criterion, the predecessor is the most accessible previous proposition, in the sense defined by Veins Theory (Cristea, Ide, and Romary, 1998). For now we assume the criterion is linear.

$\Sigma Cb(U_n)$ (potential Cbs of proposition U_n) is given by the intersection between $Cf(U_n)$ and $Cf(U_{n-1})$—that is, all the referents they have in common. The potential Cps are those referents in the current proposition that can be realized as most salient. Obviously this should depend on the linguistic resources available to the generator; the system actually uses a simpler rule based on argument types within the proposition. Table 3 shows the potential Cbs and Cps for the proposition sequence in solution A presented at the beginning of this section. As stated earlier, our treatment of salience here simplifies in two ways: We assume that syntactic realization serves only to distinguish the Cp from all other referents and that the system already knows, from the argument structure of the proposition, which entities can occur in subject position. With these simplifications, the enumeration of centering assignments is straightforward; in the above example, four combinations are possible, since there are two choices each for $Cp(U_2)$ and $Cp(U_3)$.

4.3 Evaluating Solutions

The system evaluates candidate solutions by applying a battery of tests to each node of the text plan. Each test identifies whether the node suffers from a particular defect. For instance, one stylistic defect (at least for the rhetorical relations occurring in figure 1) is that of placing nucleus before satellite; in general, the text reads better if important material is placed at the end. For each type of defect, we specify a weight indicating its importance: In evaluating continuity of reference, for example, the defect "no Cb" is regarded as more significant than other defects. Other violations are recorded only in the case in which a Cb is present, so if all violations were weighted equally, this could result in a "no-Cb" transition's being treated as less serious than an "expensive" Smooth Shift, for example (violating cheapness and cohesion). Summing the weighted costs for all defects, we obtain a total cost for the solution; our aim is to find the solution with the lowest total cost.

Regarding centering, the tests currently applied are as follows:

Salience violation: A proposition U_n violates salience if $Cb(U_n) \neq Cp(U_n)$. This defect is assessed only on propositions that have a backward-looking center.

Cohesion violation: A transition $\langle U_{n-1},\ U_n \rangle$ violates cohesion if $Cb(U_n) \neq Cb(U_{n-1})$. This defect is not recorded when either U_n or U_{n-1} has no Cb.

Cheapness violation: A transition $\langle U_{n-1},\ U_n \rangle$ violates cheapness if $Cb(U_n) \neq Cp(U_{n-1})$. This defect is assessed only on propositions that have a backward-looking center.

Continuity violation: This defect is recorded for any proposition with no Cb, except the first proposition in the sequence (which by definition cannot have a Cb).

Relative weightings for these defects can be chosen by the user; for the current examples we have chosen a neutral scheme with a weight of 3 for continuity violations and 1 each for the others, so that a no-Cb transition is ranked equally bad as an "expensive" Rough Shift.[5] Applied to the four solutions to text structures A and B presented in this section, these definitions yield costs shown in Table 4. According to our metric, solutions A1 and A2 should be preferred because they incur less cost than any others, with B3 and B4 the least preferred.

Although this article focuses on centering issues, it is important to remember that other aspects of text quality are evaluated at the same time: The aim is to compute a global measure so that disadvantages in one factor can be weighed against advantages in another. For instance, text pattern B is bound to yield poor continuity of reference because it orders the propositions so that U_1 and U_2 have no referents in common. Text pattern A avoids this defect, but this does not automatically mean that A is better than B; there may be other reasons, unconnected with centering, for preferring B to A. The constraints which have an effect on clause ordering include:

Satellite before nucleus: For nucleus-satellite relations, place the satellite before the nucleus.

Right-branching structure: If an elementary proposition is coordinated with a complex rhetorical structure, place the elementary proposition first.

Centering constraints: Penalize orderings which violate centering preferences.

Text pattern B is favored by "right-branching structure," but in this case the centering constraints will "conspire" with "satellite before nucleus" to favor pattern A overall.

5. Conclusion

We have described a technique for generating texts which will be coherent according to a reasonably faithful interpretation of centering theory. NLG systems need some principled means of deciding on the preferred orderings of clauses and of arguments within clauses, and CT appears a good candidate to provide a basis for these decisions, in tandem with other stylistic considerations. We have reported on a particular implementation in the ICONOCLAST document generation system, but the technique can be applied to other NLG systems that perform hierarchical text structuring based on a theory of coherence relations (with additional assumptions as detailed in Section 1):

- For systems which generate a single text plan, CT can determine the most coherent ordering of arguments within clauses.

---

5 See Kibble and Power (2004) for initial results of empirical research on constraint weightings.

Table 4
Realizations of text patterns A and B, with weights: cohesion | salience | cheapness = 1,
continuity = 3.

Version	Text	Cb	Cp	Defects	Sum
A1	Since Elixir contains gestodene the FDA bans Elixir. However, it approves Elixir+.	∅ elixir fda	elixir fda fda	none sal coh	2
A2	Since Elixir contains gestodene it is banned by the FDA. However, the FDA approves Elixir+.	∅ elixir fda	elixir elixir fda	none none coh, ch	2
A3	Since Elixir contains gestodene the FDA bans Elixir. However, Elixir+ is approved by the FDA.	∅ elixir fda	elixir fda elixir+	none sal sal, coh	3
A4	Since Elixir contains gestodene it is banned by the FDA. However, Elixir+ is approved by the FDA.	∅ elixir fda	elixir elixir elixir+	none none sal, coh, ch	3
B1	The FDA approves Elixir+ although since Elixir contains gestodene it is banned by the FDA.	∅ ∅ elixir	fda elixir elixir	none cont none	3
B2	Elixir+ is approved by the FDA although since Elixir contains gestodene it is banned by the FDA.	∅ ∅ elixir	elixir+ elixir elixir	none cont none	3
B3	The FDA approves Elixir+ although since Elixir contains gestodene the FDA bans Elixir.	∅ ∅ elixir	fda elixir fda	none cont sal	4
B4	Elixir+ is approved by the FDA although since Elixir contains gestodene the FDA bans Elixir.	∅ ∅ fda	elixir+ elixir elixir	none cont sal	4

Note: ch = cohesion, coh=cohesion, cont=continuity, sal=salience.

- For systems which generate multiple text plans, CT can be used to evaluate the different plans as well as to determine the optimal realization of any particular plan.

We have carried out empirical studies that provide clear evidence that centering features make a difference to the acceptability of texts and demonstrate one way to determine weightings (Kibble and Power 2004). It may turn out that different weight-

ings are appropriate for different text genres or for speech as opposed to "written" text. Our framework will facilitate detailed research into evaluation metrics and will therefore provide a productive research tool in addition to the immediate practical benefit of improving the fluency and readability of generated texts.

References

Beaver, David. 2004. The optimization of discourse anaphora. *Linguistics and Philosophy*, 27(1):3–56.

Brennan, Susan, Marilyn Walker Friedman, and Carl Pollard. 1987. A centering approach to pronouns. In *Proceedings of 25th ACL*, pages 155–162, Stanford, CA.

Cahill, Lynne. 1999. Lexicalisation in applied NLG systems. Technical Report ITRI-99-04, Information Technology Research Institute, University of Brighton.

Callaway, Charles B. and James C. Lester. 2002. Pronominalization in generated discourse and dialogue. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 88–95, Philadelphia.

Cheng, Hua. 2000. Experimenting with the interaction between aggregation and text planning. In *Proceedings of ANLP-NAACL*, pages 1–6, Seattle.

Cristea, Dan, Nancy Ide, and Laurent Romary. 1998. Veins theory: A model of global discourse cohesion and coherence. In *Proceedings of COLING/ACL'98*, pages 281–285, Montreal.

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

Henschel, Renate, Hua Cheng, and Massimo Poesio. 2000. Pronominalisation revisited. In *Proceedings of 18th COLING*, pages 306–312, Saarbrücken, Germany.

Kameyama, Megumi. 1998. Intrasentential centering: A case study. In Marilyn

Walker, Aravind Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*, pages 89–112. Clarendon, Oxford.

Karamanis, Nikiforos. 2001. Exploring entity-based coherence. In *Proceedings of Fourth CLUK*, pages 18–26, University of Sheffield, Sheffield, England.

Kibble, Rodger. 1999. Cb or not Cb? Centering theory applied to NLG. In *Proceedings of ACL Workshop on Discourse and Reference Structure*, pages 72–81, University of Maryland, College Park.

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

Kibble, Rodger. 2003. Towards the elimination of centering theory. In Ivana Kruijff-Korbayová and Claudia Kosny, editors, *DiaBruck 2003: Proceedings of the Seventh Workshop on the Semantics and Pragmatics of Dialogue*, Universität des Saarlandes, Saarbrücken, Germany.

Kibble, Rodger and Richard Power. 2004. Optimising referential coherence as a constraint satisfaction problem. Technical Report RK/2004/1, Department of Computing, Goldsmiths College, and ITRI-04-07, Information Technology Research Institute, University of Brighton.

Mann, William and Sandra Thompson. 1987. Rhetorical structure theory: A theory of text organisation. Technical Report ISI/RS-87-190, Information Sciences Institute, Los Angeles.

Marcu, Daniel. 1996. Building up rhetorical structure trees. In *Proceedings of AAAI-96*, pages 1069–1074, Portland, OR.

Marcu, Daniel. 1997. From local to global coherence: A bottom-up approach to text planning. In *Proceedings of AAAI-97*, pages 629–635, Providence, RI.

McCoy, Kathleen and Michael Strube. 1999. Generating anaphoric expressions: Pronoun or definite description? In *Proceedings of ACL Workshop on Discourse and Reference Structure*, pages 63–71, University of Maryland, College Park.

McKeown, Kathleen R. 1985. *Text Generation*. Cambridge University Press, Cambridge.

Mellish, Chris, Alistair Knott, Jon Oberlander, and Mick O'Donnell. 1998. Experiments using stochastic search for text planning. In *Proceedings of the Ninth International Workshop on Natural Language*

Generation, pages 97–108,
Niagara-on-the-Lake, Ontario.

Mittal, Vibhu, Johanna Moore, Giuseppe
Carenini, and Steven Roth. 1998.
Describing complex charts in natural
language: A caption generation system.
Computational Linguistics, 24(3):431–467.

Poesio, Massimo, Rosemary Stevenson, Hua
Cheng, Barbara di Eugenio, and Janet
Hitzeman. 2002. A corpus-based
evaluation of centering theory. Technical
Report TN-02-01/CSM-369, Natural
Language Engineering Group, University
of Essex.

Power, Richard. 2000. Planning texts by
constraint satisfaction. In *Proceedings of
COLING 2000*, pages 642–648,
Saarbrücken, Germany.

Power, Richard, Donia Scott, and Nadjet
Bouayad-Agha. 2003. Document structure.
Computational Linguistics, 29(2):211–260.

Prince, Ellen. 1999. How not to mark topics:
"Topicalization" in English and Yiddish.
Unpublished manuscript, Linguistics
Department, University of Pennsylvania.

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

Reiter, Ehud. 2000. Pipelines and size
constraints. *Computational Linguistics*,
26(2):251–259.

Scott, Donia and Clarisse de Souza. 1990.
Getting the message across in RST-based
text generation. In Robert Dale, Chris
Mellish, and Michael Zock, editors,
*Current Research in Natural Language
Generation*, pages 47–73. Academic Press,
London.

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

Suri, Linda, Kathleen McCoy, and Jonathan
DeCristofaro. 1999. A methodology for
extending focussing franeworks.
Computational Linguistics, 25(2):173–194.

van Hentenryck, P. 1989. *Constraint
Satisfaction in Logic Programming*. MIT
Press, Cambridge, MA.

Walker, Marilyn, Aravind Joshi, and Ellen
Prince, editors. 1998a. *Centering Theory in
Discourse*. Clarendon, Oxford.

Walker, Marilyn, Aravind Joshi, and Ellen
Prince. 1998b. Centering in naturally
occurring discourse. In Marilyn Walker,
Aravind Joshi, and Ellen Prince, editors,
Centering Theory in Discourse, pages 1–28.
Clarendon, Oxford.