RST Revisited: Disentangling Nuclearity[1]

Manfred Stede

Universität Potsdam (Germany)

Abstract

The paper discusses the notion of *nuclearity* as it had been put forward by Rhetorical Structure Theory as a general principle of text organization. On the basis of an inquiry into different kinds of *salience* phenomena in texts, several problems with the purported role of nuclearity are identified. It is argued that RST trees conflate too much information from different realms of description in a single structure. As an alternative for more detailed investigation of coherence phenomena, an approach toward *multi-level analysis* and annotation of text is outlined, which also keeps the various sources of salience separate from each other.

1. Introduction

Among the various theories of discourse structure, Rhetorical Structure Theory (RST, Mann/Thompson 1988) can probably claim to be the most *empirical*: It has been developed at the time on the basis of thorough analyses of a variety of texts, and it has since been used for quite different purposes and for different types of text. To characterize it very briefly[2], the idea of RST is to postulate some 20 coherence relations, defined in terms of the effect they are meant to cause in the reader's beliefs and attitudes, and to claim that a

---

[1] For their very helpful comments on earlier versions of this paper, I am grateful to Bonnie Webber, John Bateman and two anonymous reviewers.
[2] A good general introduction to RST, also discussing other work that was inspired by the theory, is given by Taboada and Mann (2006). See also the website www.sfu.ca/rst

coherent text can be analysed in terms of these relations. This would lead to a tree structure that recursively connects all the „minimal units" and the resulting larger text spans. Following popular practice in later publications, we henceforth call the minimal units *EDUs*, for "elementary discourse units". Mann and Thompson acknowledge that coherence relations are "pre-realizational" in the sense that they connect mental representations of text spans – but for the purposes of text analysis one regards them as connecting actual portions of surface text. As for the syntactic shape of the EDUs, the authors deliberately do not go into much detail but characterize them as being typically clauses. Mann and Matthiessen (1991: 234) elaborate on this decision: "The units of an RST analysis are chosen to fit the purposes of the analysis, and are not theoretically prescribed. Our usual practice in analyzing a text is to regard clauses as the realizations of units, but to merge restrictive relative and complement clauses with their parents and to treat elliptical clauses as if they were non-elliptical." In the following, we use "text segments" and "text spans" interchangeably as referring either to EDUs or to larger compositions of them.

A relation may only join adjacent text spans, and any span can only have one „parent node" in the tree. In the relation set postulated by Mann and Thompson (1988), 21 of the 24 coherence relations join a *nucleus* and a *satellite* segment, where the latter is less "important" (to be clarified in the next section). The remaining three relations are *multinuclear*, i.e. they join segments of equal importance. An important claim is another partitioning of relations into two groups: Instances of *presentational* relations are supposed to change the reader's mind in some way, such as motivate her to take a particular action, enable her to perform an action, or encourage her to believe a certain proposition. S*ubject-matter* relations, on the other hand, describe kinds of connections between events happening in the world; here, the effect is merely that the reader „recognize" the relationship between the two units. – As an example, consider the following definition of the (presentational) relation Evidence:

Evidence
Constraints on nucleus (N): Reader might not believe N to a degree satisfactory to writer

Constraints on satellite (S): Reader believes S or will find it credible
Constraints on the N+S combination: Reader's comprehending S increases Reader's belief of N
The effect: Reader's belief of N is increased
Locus of the effect: N

The empirical foundation was the main reason for RST's considerable popularity among researchers concerned with authentic discourse. And yet, the very fact that the same rhetorical relations can be applied to so many different pairs of text spans implies the danger that definitions become so vague that at many points in a text analysis, several relations can be assumed to hold, which equally well conform to the definitions. Ambiguity is not *a priori* problematic for text analysis – it is not surprising that portions of text and their relationships can have different readings – but it should be possible for the assumed representations to take notice of such genuine ambiguities rather than to hide them under the covers of a single tree representation. In earlier work (Grote et al. 1997; Stede 2004b), colleagues and myself argued that both the *Concession* and the *Contrast* relation in RST are a victim of the tendency toward too unspecific definitions; they conflate a variety of different ways in which two states of affairs can be contrasted with one another. In this paper, I try to make essentially the same point for RST's notion of *nuclearity.* Mann and Thompson, in a nutshell, characterized the nucleus of a relation as being „more central to the writer's purposes", and my point will be that being central to one's purposes can be a very different thing from case to case, and that it is worth distinguishing certain notions that occasionally give rise to conflicting predictions on nuclearity. As a consequence, it will be questioned whether it is helpful to assume a single hierarchical structure as an account of text coherence.

The paper proceeds as follows. In order to do justice to the original proposals on nuclearity by Mann, Matthiessen and Thompson in the late 1980s, their position will first be summarized in Section 2. Also, some interesting extensions proposed by other authors will be mentioned there. I then turn to a more general discussion of ‚salience' in text and investigate different ways in which certain segments can be more important/prominent/salient than others (Section 3). This will be my starting point for a critique of RST-nuclearity in

Section 4, which will be followed in Section 5 by a proposal for a multi-level approach to discourse structure, which confines different types of information to distinct layers.

2. Nuclearity in RST

The original idea: Mann, Matthiessen, and Thompson

Towards the end of the most-cited original RST paper, Mann and Thompson (1988) devote a section to clarifying their notion of nuclearity, which in the previous sections they had used more on an intuitive basis when defining the various coherence relations. The trigger for proposing nuclearity as a general principle of text organization was the observation that all but three (Sequence, Contrast, and Joint) of their coherence relations were quite obviously *asymmetric*. A clear example is Evidence: When A is evidence for B, then B cannot be evidence for A. In carving out this asymmetry for all the relations, Mann and Thompson noticed three commonalities, which for them warranted the general division into nucleus and satellite: (i) Often, one segment depends on the other for the discourse to be comprehensible – for Evidence, the nuclear claim is necessary for the satellite to be understood correctly; without the nucleus, it would be a non-sequitur. Consider the following example, where the relation holds between nucleus [a] and satellite [b], and [a] is more central for the text function than [b]:

(1)   [a] *Tomorrow the weather will be nice.* [b] *I listened to the latest forecast on the radio.* [c] *So we can go to the picnic in the afternoon.*

(ii) Often, the satellite is prone to substitution – for Evidence, one might replace the satellite with a different piece of evidence, without changing the overall purpose of the text. (iii) Often, the nucleus is more essential to the writer's purposes than the satellite.

The much-cited criterion for making the distinction between nuclei and satellites is Mann and Thompson's *deletion test*. The first argument concerns the overall text function: We can determine the "most nuclear" minimal unit of a text by tracing down from the root node to the nucleus at each level all the

way to the minimal units. For illustration, see Figure 1[3], which gives the RST analysis of a short text that extends example (1). Nuclei are attached with straight lines to their mother nodes, whereas satellites are linked to their nuclei with curved lines. Moving down from the root node (1-6) to a leaf node (i.e., an EDU) without ever following a satellite branch leads to unit (3), the "most nuclear" one. The deletion test predicts that when removing the most nuclear unit, the overall message of the text typically becomes quite difficult, if not impossible, to infer. This diagnostic is then extended to all minimal units that serve as nuclei in the tree (in Figure 1, these are 1, 3, 4, and 5): When they are removed from the text, the result is an incoherent sequence of utterances. On the other hand, with removal of those minimal units that have been assigned satellite status (2 and 6 in Figure 1), the remainders (even though lacking some cohesion) still convey an idea of what the text is about and often even manage to acceptably communicate the main idea.

Fig. 1: RST analysis of a sample text
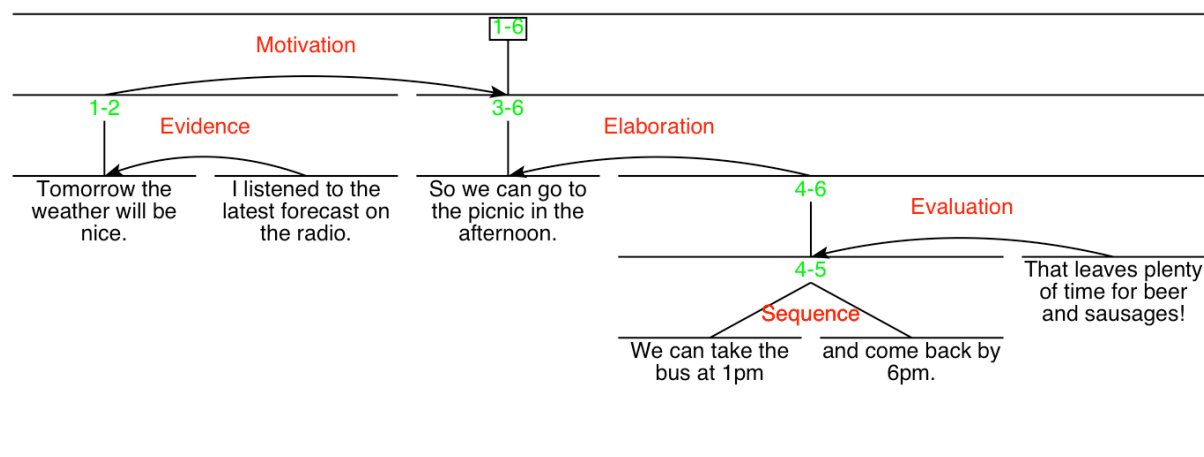
As for the relationship between nuclearity and linguistic structure, Mann and Thompson mention the idea that the nucleus/satellite distinction directly correspond to the syntactic phenomenon of hypotactic clause combining: „Grammars in many languages draw a distinction between hypotactic and

---

[3] The figure was created with the RSTTool software
(http://www.wagsoft.com/RSTTool).

main clauses because of the nucleus-satellite distinction in discourse." (p. 269) We do not elaborate here on the even stronger proposition conveyed in that quotation – that nuclearity might be *responsible* for grammars having evolved hypotactic clause combining; instead we will later discuss merely the (weaker) *correlation* that is proposed to hold between grammar and discourse structure. The claim is expanded in more detail by Matthiessen and Thompson (1988), who argue for a direct mapping between subordinate clauses and satellites of coherence relations.

Mann and Thompson acknowledge two cases where the nucleus/satellite distinction is rather meaningless: in „enveloping structures" such as conventionalized beginnings and endings of letters, and in parallel structures as they had been discussed by Fries (1981). His examples involve repeated comparisons and contrasts between the same entities. The shape of such structures comparing *A* and *B* can be sketched in a simplified way as follows: *While A is X1, B is rather Y1. A has a X1, and B has a X2. X3 is highly relevant for A, whereas X4 is more important for B.*

Regarding the functional interpretation of nuclearity, Mann and Thompson make another distinction between two groups of relations, based on their notion of *locus of effect* of the relation, which is part of each definition and states the segment that the described effect should arise from:[4]

- When the locus of effect is only the nucleus (as in Evidence), nuclearity represents the qualitative difference in role between the essential and the inessential, thought and afterthought. The satellite supports the nucleus but does not contribute to it, and the writer intends the reader to notice the distinction introduced by nuclearity.
- When the locus of effect is both nucleus and satellite (as in Condition or Elaboration), the structural difference between nucleus and satellite represents some distinction in the organization of the subject matter. The distinction is presented as important to the reader, and the

[4] This is a part of the definitions that in fact did not always receive enough attention by other authors referring to RST relations and applying them to sample texts. One reason might be that the "locus of effect" field is not present in the definitions that were chosen for the RST website (www.sfu.ca/rst).

significance of the satellite tends to be found in the nucleus – both nucleus and satellite contribute to the result.

The role of locus of effect is discussed in more detail in a less-cited paper by Mann and Matthiessen (1991). Interestingly, here the authors propose a more balanced view of nucleus and satellite in subject-matter relations, for which "it seems quite reasonable that the locus of effect should be the nucleus plus the satellite. The purpose of such relations is to represent the kind of connection prevailing between the nucleus and the satellite, not just to achieve the function of the nucleus" (p. 244). Mann and Matthiessen proceed to state that the division of relations according to locus of effect and that of presentational versus subject-matter orientation is in fact identical. This minimally diverges from the definitions given in (Mann/Thompson 1988), where Concession is in conflict with this view – maybe not surprisingly, because Concession can indeed be seen as a somewhat problematic presentational relation (Grote et al. 1997).

Nuclei on the run: extensions by other authors

The role of nuclearity for discourse structure has been strengthened further by Marcu (2000) who proposed the „strong nuclearity hypothesis": When a relation is postulated to hold between two spans of text, then it should also hold between the nuclei of these two spans. This move ensures the „upward compatibility" of the idea of nuclearity, from minimal units to arbitrary text spans. At the same time, it generates new constraints for nuclearity decisions at the lower levels: The annotator is encouraged to assign nuclearity status to EDUs in such a way that the combinations of larger units work out correctly. Thus, if in doubt about assigning a relation with a particular nucleus/satellite distribution on a lower level, considering the wider context will often assist in making the decision. Mann and Thompson had argued their case on the basis of small examples with elementary units only; the upward extension was implicitly alluded to but not explicitly proposed. Marcu took this step and supported it, *inter alia*, with the argument that the application of automatic text summarization would benefit from it: The nuclear portions of a text are assumed to be the more important ones that should become part of an (extractive) summary of the text. In (Marcu 1999), he suggests to turn the notion of nuclearity into a *scalar* one when considering complete texts; the

measure for assigning a degree of nuclearity to an EDU considers to what extent it is embedded in larger nuclei and satellites, which results in a partial ordering of all the units according to their „importance". For example, the aforementioned "most nuclear" EDU(s) will have only nuclear segments along the path from the EDU to the root node of the tree; others also appear in a number of satellites, which determines their degree. For our sample tree in Figure 1, this measure would assign the same degree to EDUs 1, 4 and 5, as they all involve one satellite connection when walking up to the root node.

Another notable promotion of the role of nuclearity was suggested by Cristea et al. (1998). In their „veins theory", they argue that antecedents to anaphoric expressions are more likely to be found in nuclear material of the preceding context rather than in satellite material. This appears to echo the distinction between *subordinating* and *coordinating* coherence relations in Segmented Discourse Representation Theory (Asher and Lascarides 2003), which was to a good extent motivated by constraining anaphoric accessibility – an idea that probably originated with Polanyi's (1988) „right-frontier constraint". While it had originated in this fashion as a linguistically-motivated criterion, SDRT in more recent work seems to regard it as a close relative of nuclearity in RST (cf. Danlos 2007).

An interesting move was made in the annotation guidelines for the RST Treebank (Carlson et al. 2003). The number of coherence relations used in that corpus is 78, which includes 25 multinuclear ones. This amounts to 32% of all relations, compared to 12.5% in the original Mann/Thompson relation set (where only 3 out of 24 relations are multinuclear). 17 of those 25 multinuclear relations have mononuclear counterparts, i.e., relations with identical definitions except for the assignment of one or two nuclei, respectively. In the RST Treebank, annotators made equal use of both versions: They chose the multinuclear version of a relation 1905 times (51.4%) and the mononuclear one 1801 times (48.6%).[5] Moreover, according to the guidelines, 8 of the mononuclear relations occur in both variants of nucleus/satellite distribution (as Mann and Thompson had proposed merely

---

[5] These numbers of annotations in the treebank were provided by Bonnie Webber (personal communication).

9

for the Cause and Result relations). This indicates that Carlson et al. were acutely aware of the problem that annotators often experience when asked to decide which of two segments should be labelled nuclear. Carlson et al.'s overall solution is to work with a much larger set of coherence relations; to my mind, this hides the fact that coherence arises from distinct sources, which also should be represented separately, and that nuclearity as a purportedly "general" phenomenon is largely orthogonal to the description of the specific coherence relations.

Other concepts of coherence relations

When considering other proposals that explain text coherence in terms of relations, it turns out that not many of them include a notion that corresponds to nuclearity as discussed in RST. Neither the work of Hobbs (1979), Sanders et al. (1992), or Kehler (2002) systematically assigns different prominence to the segments of relations. On the other hand, Bateman and Rondhuis (1997), who more explicitly build upon RST, propose to regard nuclearity not as a static aspect of the relation definitions, but as a distinct feature that can combine with a set of other features to dynamically create a coherence relation (similar to the decompositional approach of Sanders et al. (1992)). That is, they transfer the idea underlying the Cause/Result relation pair of Mann and Thompson to the entire set of relations, so that any relation can assign nucleus status to either span.[6] This amounts to a more radical solution to the above-mentioned problem noted by Carlson et al. (2003), whose response was to partition the overall relation set according to the distribution of nuclearity.

Another approach that borrowed nuclearity from RST is „relational discourse analysis" (Moser and Moore 1995), which aims at synthesizing RST with the approach of Grosz and Sidner (1986). These authors use the terms *core* and *contributor* for nucleus and satellite, respectively. A recent approach is that of Wolf and Gibson (2005) who draw a distinction between *directed* and *undirected* relations, and while they mention RST as related work, they do not

---

[6] Similarly, a recent proposal of a relation set by Golebiowski (2006) states that the majority of her relations "are basically neutral, and thus able to function both paratactically and hypotactically." (p. 261)

seem to regard their directionality as a general principle of text organization. Rather, they treat it as one feature of the semantics of the individual relations. The same holds for the more surface-oriented *conjunctive relations* proposed by Martin (1992), which will be mentioned again in Section 5.

In summary, the idea of nuclearity as a principle of discourse organization based on "centrality for the writer's purposes" is largely confined to RST and approaches that directly build upon it; a range of other theories of discourse coherence does not postulate an equivalent notion.

3. Salience in text

Having stated the RST position on nuclearity, we now turn to a somewhat broader viewpoint and examine different kinds of relative salience of segments in text – where *salience* is now meant as a neutral cover term for a variety of phenomena to be investigated. It is probably beyond doubt that a reader of a text perceives different portions as more or less salient; the crucial question is whether this is due to a single, underlying notion – nuclearity – that should be postulated as an elementary principle of text organization.

In this section, I will challenge this idea with observations that resulted from our annotation work with the Potsdam Commentary Corpus (Stede 2004), a collection of German newspaper commentaries, 175 of which have been annotated with RST trees by students trained on the basis of (Mann/Thompson 1988). From time to time, we asked the annotators on what grounds they had made their decision of nucleus assignment between adjacent text spans. Adding then some of our own observations leads to the following list of reasons that annotators can put forward when labelling a text segment as a nucleus. To illustrate the phenomena, in several cases we refer to a sample text, the beginning of the introductory section of a research paper, shown in Figure 2. Sentences are numbered, but note that most sentences contain more than one „minimal unit" of an RST-style analysis.

Fig. 2: Excerpt from D. Litman et al.: „Spoken versus typed dialogue tutoring.“ Int'l Journal of Artificial Intelligence in Education 16(2), 2006, p. 146. Sentence numbers inserted by MS.

(1) It is widely believed that the best human tutors are more effective than the best computer tutors, (...). (2) A major difference between human and computer tutors is that human tutors use face-to-face spoken natural language dialogue, whereas computer tutors typically use menu-based interaction or typed natural language dialogue. (3) This raises the question of whether making the interaction more natural, such as by changing the modality of the computer tutoring to spoken natural language dialogue, would decrease the advantage of human tutoring over computer tutoring.

(4) In fact, as will be detailed below, several potential benefits of spoken tutorial dialogue with respect to increasing learning have already been hypothesized in the literature. (5) One hypothesis is that spoken dialogue may be better at eliciting student behaviors that are believed to accelerate learning, such as student knowledge construction. (6) A second hypothesis is that speech allows tutors to infer a more accurate student model, which similarly is believed to accelerate learning. (7) A third hypothesis is that speech primes a more social interpretation of the tutorial environment, which again is hypothesized to accelerate learning.

(8) It is thus important to test whether a move to spoken dialogues is likely to yield increased benefits with respect to learning and other performance measures. (9) Furthermore, if the addition of speech can indeed increase learning gains, it is also important to understand why spoken dialogue accelerates learning.

a) One segment turns out to be more supportive of the text's overall purpose. Constructed example:

(2)   [a] *The Labour Party supported the strike.* [b] *The Liberals strongly voted against it.* [c] *Thus we have seen another example of the weakness of the Labour.*

Considering [a] and [b] in isolation, none of the two appears to deserve nuclearity status more than the other. But [c] indicates that the text is meant to encourage readers to support the position of the Liberals, thus rendering [b] nuclear.

b) One segment supports the intention of the other segment. This is the „classical" situation of RST relations such as Evidence or – in this example – Enablement:

(3)   *Open the printer cover carefully. The lever is located at the lower right of the backside.*

In the sample text in Figure 1, sentence 8 states the main question the paper is set to answer, and 4-7 are meant to support the claim that the question is actually a relevant one.

c) Recurrence: Material in one of the two spans is taken up later in the text, and hence appears to be more important for the development of the text:

(4)   [a] *Jim was opening a wine bottle when* [b] *his mother rushed in.* [c] *She held a letter in her hands.*

Choosing a relation joining [a] and [b] can amount to the decision which proposition is meant to be in the foreground and which in the background; since the mother appearing in [b] is also the subject of [c], this local configuration suggests that [b] be more nuclear than [a].

d) Repetition: When longer material is repeated, as is the case in both the second clauses of sentences 6 and 7 of the sample text, it does not provide new information but merely reminds the hearer – which renders it less nuclear than a potentially more informative adjacent unit (cf. the deletion test mentioned in Section 2).

e) Digression: When a text segment constitutes a temporary digression from the main topic of the text, it is likely to be only a satellite – which is also supported by the deletion test.

f) Meta-discursive elements: A segment can appear to be decidedly non-nuclear when it is a text-internal directive. An example is the clause *as will be detailed below* in sentence 4 of the sample text.

g) Connective: Sometimes, the connective chosen by the author can mark the nucleus/satellite distinction, as with the German „zwar (sat) aber (nuc)", or „nicht nur (sat) sondern (nuc)." Example:

(5)   *Der Bürgermeister hat zwar keinen Erfolg beim Bürokratieabbau, aber er ist ein eifriger Wirtschaftsförderer.*
         (,The mayor has no success in cutting red tape, but he is a vivid supporter of the local economy.')

Similarly, punctuation marks can indicate parenthetical, and hence satellite, information.

h) Other lexical marking of salience: Authors can use lexical items to explicit assign „weight" to utterances, as the phrase *it is important to* does in sentences 8 and 9 of the sample text.

i) Syntactic structure: A main clause often seems more nuclear than an associated subordinate clause, as mentioned in Section 2.

(6)   *While the speaker struggled through his manuscript, the audience gradually turned to sleep.*

Several examples can also be found in the sample text, where subordination correlates with other criteria, e.g. with (f) in sentence 4.

j) RST definition: The relation is stated by the RST definition to assign nuclearity in a specific way. For instance, Sentence 9 of the sample text expresses a Condition relation, which according to Mann/Thompson (1988) consists of a condition-satellite and a consequence-nucleus. Similarly, Carlson et al. (2003) had to fix the nucleus assignment for the relation Attribution used in the RST Treebank. It relates a statement to its source, as in *The spokesman reported that the company lost two important contracts last year*. Or consider sentence 1 of the sample text, where the source is not an individual but "public opinion". Rather than having annotators decide whether the source's act of reporting or the content of the report is more important, Carlson et al. defined the content portion of such sentences generally to be the nucleus.

Notice that these factors, except for (j), are not meant to be hard-and-fast rules for making the nuclearity decision; rather, they are evidential factors. In a suitable context, any one of them could be cited by an annotator as „the" decisive factor for labelling one of the adjacent text segments as nuclear. (a)-(j) are obviously not mutually exclusive, but – and that is the point here – they are largely independent from each other: Any one of them alone can in some specific context be taken as the decisive evidence.

The observation that the factors are not mutually exclusive suggests that they operate on somewhat different levels of description. (a) and (b) are intention-based criteria: When the intention of one segment „dominates" that of the adjacent one, or when the segment intention supports the overall text intention, we have evidence for nuclearity. (c)-(f) in different ways have to do with the thematic development of the text. In (c), coreference and topic-continuity trigger the decision on the level of minimal units, and in (e) topic dis-/continuity, among minimal or larger units, is responsible. Repeating information in (d) and supplying text-organizing meta-information in (f) can be seen as ways of interrupting the „main" flow of information. Criteria (g)-(i) are the most surface-oriented ones, and they are likely to co-occur with some of the others. Underneath the surface, (h) again is related to intentions: The author, for whatever reason, chooses to emphasize some portion of the

text. In contrast to (a) and (b), however, this is not relational but a feature of a single text unit – the reader shall recognize that the author regards the contents of this unit as central for her purposes. This might co-occur with (a) – the unit can also support the intention of a larger segment – but it need not necessarily do so. Finally, (j) is of an entirely different kind: It is an inherent necessity postulated by the underlying theory to assign nuclearity in one particular way once a specific relation has been found applicable.

The fact that such a multitude of factors can influence nuclearity assignment is not *per se* problematic. The question is, however, whether those distinct factors indeed are reflexes of a common underlying notion, a „general principle of text organization". In the next section I will argue that we have reason to be sceptical on this issue.

4. Problems with nuclearity in RST

We now proceed to synthesize the exposition of RST-nuclearity and our inventory of types of salience into a critique and then to an outline of a different approach towards investigating discourse phenomena. To that end, this section critically reviews three central aspects of the RST view of nuclearity: its relationship to syntactic subordination, its direct association with coherence relations, and the purported pervasiveness of nuclearity.

Syntactic subordination and nuclearity

Undoubtedly, the idea to link syntactic subordination to discourse nuclearity has intuitive appeal, and moreover it is not difficult to accumulate evidence in support of such a link: Quite often, a syntactically subordinate clause is in fact „less central to the writer's purposes" than the matrix clause is. But on the other hand, one can also gather evidence for the opposite position – the writer's purposes running against syntactic subordination, which altogether suggests that matters are more complicated, and that the view accentuated especially in Matthiessen and Thompson (1988) generalizes too boldly from a mere tendency to a principle (see Blühdorn, this volume). For one thing, we have pointed out in the previous section that in the absence of any over-arching intentional structure, it is the *continuation* of discourse – in terms of

referential and/or thematic continuity – which often governs nuclearity assignment. If a discourse referent retains prominence in the subsequent text, the clause containing its original introduction has a good reason to claim nuclearity status. This clause, however, might quite well be a subordinate one. This point is also made by Golebiowski (2006: 262), who uses the sentence *While I was feeling really well on Saturday, I was ill on Sunday* to illustrate that in a suitable context, either both clauses can be of equal functional prominence (7a) or one can be more important (7b):

(7) [a] *While I was feeling really well on Saturday, I was ill on Sunday. On Saturday, I had a lot of energy and I (…) managed to have a beautiful meal with my friends and even spent an enjoyable evening with my family. On Sunday, however, I woke up with a splitting headache. In spite of taking some pain relief tablets, my head did not clear up till the evening. I had to cancel all arrangements made for the day and ended up spending most of Sunday in bed.*

(7) [b] *While I was feeling really well on Saturday, I was ill on Sunday. I woke up with a splitting headache. In spite of taking some pain relief tablets, my head did not clear up till the evening. I had to cancel all arrangements made for the day and ended up spending most of Sunday in bed.*

It seems worth to relate this point to a different issue raised in Mann/Thompson (1988), that of a „standard ordering" of the spans of relations. The authors state that many relations have a clear tendency as to the linear ordering of their nucleus und satellite; for example, both Antithesis and Condition would normally have their satellites occur before their nuclei. Notice that this statement in conjunction with frequent linguistic realization of such relations („*although* A, B" and „*if* A, *then* B", respectively) indeed lends support to the purported link between satellite and subordinate clause. But, of course, this „standard" ordering is by no means mandatory – „B, *although* A" and „B, *if* A" are perfectly natural. In these cases, however, the subordinate A segments (i.e., the satellites qua relation definition) due to their position at the end of the sentence can play different roles for the continuation

of the discourse; for the case of temporal clauses, ordering has been investigated by de Swart (1999) and Schilder/Tenbrink (2001).

In summary, for text portions where intentions provide the structural scaffolding, subordination seems to usually mirror the relative weights of such intentions:

(8)   *While Smith's approach to the free-will problem was not entirely wrong, the general direction of his philosophy has been completely refuted by now.*

But where intentions do not play the critical role[7], *linear order* takes over as an important force in assigning different degrees of prominence to text segments. Syntactic subordination also plays a role here, but the interactions become complicated, and RST does not offer the means to account for such effects.

In addition to the more difficult relationship between subordinate clauses and nuclearity, it is also not the case that, conversely, co-ordinating clauses would naturally lead to multinuclearity. It is well-known that in clauses joined by the contrastive conjunction *but* the second clause typically carries more „weight" than the first one. An illustrative example was given by Elhadad and McKeown (1990):

(9) [a] *He   failed   the   exam,   but   he   is   smart.   Let's   hire   him.*
(9) [b] * *He is smart, but he failed the exam. Let's hire him.*

Nuclearity tied to relations

For the seven presentational, or intention-based, coherence relations in RST, nuclearity plays the role of „domination" between discourse segment purposes in the theory of Grosz and Sidner (1986). RST generalized this

---

[7] This does not mean that intentions can be entirely absent; it is to be understood in terms of the distinction drawn between *subject-matter* and *presentational* relations by Mann and Thompson (and quite similarly by other authors), as sketched at the very beginning of the paper.

notion to the 16 subject-matter relations as well and defined for each relation which elements would play the roles of nucleus and satellite, respectively. The one exception is the set of causal relations, where RST distinguishes Cause from Result, with the only difference being the mapping from cause/effect to nucleus/satellite. Mann and Thompson thus acknowledge that describing a causal relation between two events does not inherently make either cause or effect more salient than the other. But is such an inherent ascription warranted for the other subject-matter relations? Given the list of factors contributing to nuclearity decisions in Section 3, it is not clear that for any instance of a relation such as Interpretation or Evaluation, the interpreted or evaluated segment would consistently be more important than the interpreting/evaluating one. Consider again our example text in Figure 1. If the text were continued with a sentence like *For me, eating and drinking have always been the major attraction at picnics*, the weight distribution for the preceding three segments would change, because now it turns out that the beer-and-sausage sentence is more relevant for the text function than the Sequence segment giving the timing of the trip. Correspondingly, for the Interpretation relation, it is not difficult to conceive contexts that render the interpretation more prominent than the interpreted.

With other relations, predefining the nuclearity assignment is equally problematic. A case in point (noted also by Bateman and Rondhuis (1997)) is the Purpose relation, where according to the RST definition the underlying goal of the activity is the satellite and the activity itself is the nucleus. But, given a sentence like (10), why would buying the sports car and impressing the girl-friend not equally be able to play the role of nucleus?

(10)   *Jim bought a red sports car to impress his new girl-friend.*

Adopting again the argument of continuation possibilities, both elements can easily be picked up by a subsequent sentence – even more so when the linear order of the Purpose sentence is in line with the continuation. Or, independent of a particular continuation, either the sports car or the necessity of impressing the girl-friend can be the main topic of the larger discourse unit, so that criterion (a) from Section 3 would favour it as nucleus. It is evident that the

elements of a Purpose relation are clearly distinguishable on semantic grounds, but this should not entail a strict mapping to a nuclearity assignment.

(11)   *If I win the lottery next week, the first thing I'll buy is a red sports car.*

The same point can be made for the Condition relation. In (11), the two clauses clearly express propositions of different epistemic status, and thus the relation is ‚asymmetric' – but it is not clear in what sense the two clauses should be distinguished on the grounds of a general notion of nuclearity. Example (9) above, at least, demonstrates a difference in salience that is clearly not the same as that found in such Condition examples. And besides, the „continuation argument" applies to Condition just as well: The following clause can either elaborate the chances of winning the lottery, or the prospective purchase of a sports-car (or neither one).

Mann and Thompson apparently saw these problems and pointed out that the definitions of relations such as Condition and Purpose would specify as „locus of effect" both the nucleus and the satellite, and therefore the deletion test would not produce satisfactory results for them (which is correct). Resorting to a „locus of effect", however, brings a new parameter into play and in effect undermines the general nature of the nuclearity idea. It seems more appropriate to question the strict association between (subject-matter) relations and nuclearity assignment altogether.

Finally, while Condition and Purpose are clearly marked at the linguistic surface and thus do not create ambiguity for annotators, the situation is different with the aforementioned Interpretation and Evaluation. We pointed out that their fixed nucleus assignment can create conflicts in the annotation process. Notice however, that these relations are much less "visible" at the surface and hence rely more on context-driven interpretation. Unfortunately, when the definition of Interpretation or Evaluation is met for a text span, chances are that other relation definitions are also met, e.g. the notorious Background or the rather vague Circumstance. These two offer a way to turn the nucleus assignment around, and thus annotators can select a "weaker" relation rather than the specific Evaluation/Interpretation, merely in order to achieve a suitable nucleus assignment. This might serve to get the *overall*

discourse structure right (in terms of degrees of nuclearity of segments; see above), but it is clearly not optimal for describing *local* coherence.

Enforced nuclearity

One source of complaints from RST annotators is the requirement that nucleus assignment is a „must" for every pair of segments, unless one wants to resort to the multinuclear Joint relation, which is void of any specific content. On the level of minimal units, such cases are encountered, for example, in descriptive texts that characterize various aspects of a topic in an enumerative way. Among those aspects, two may be causally related, but there need not be a consequence of the causality for the writer's intentions, nor for thematic continuity. The relationship can be marked by a causative adverbial, leaving the surface structure „neutral"[8]:

(12)   *The courtyard was very quiet that day. Two teenagers read their books in one corner, and a new car was parked beside them. The trees still had all their leaves, and thus the flowers were blooming in a cool shade. A little bench stood in the middle of the lawn. (...)*

Teenagers, car, trees, flowers and bench all contribute to the description. The marked causal relation between the third and the fourth clause leaves no doubt as to the „semantic directionality" from cause to effect, but from a textual perspective, it appears to be an artificial move to grant either of the two clauses a nuclear status, for on what grounds would it be more central to the writer's purposes? Accordingly, an annotator would struggle here to decide between RST's Cause and Result relations. Recall that the relation set used by Carlson et al. (2003) employs many more multinuclear relations than the original RST set, which indicates that those researchers encountered the same problem (for our specific example, a multinuclear Cause-Consequence relation is provided by Carlson et al.).

---

[8] In German, where a causative adverbial such as *deshalb* can float quite freely throughout the clause, the impression of perceived „neutrality", or the lack of a structure-induced weighting, might be somewhat clearer than in English.

As indicated earlier, Mann and Thompson (1988) had acknowledged that envelope structures (conventionalized beginnings and endings) and certain parallel structures of comparison and contrast are to be exempt from the need to assign nuclei. My suggestion here is that this list needs to be extended quite a bit, which questions the explanatory power of the nuclearity notion. Besides some problematic cases on the level of minimal units as just illustrated, the problem is also prominent higher up in the tree structure, between larger segments, where structural clues indicating a salience distribution are often absent. The severity of this problem differs along the dimension of *text type* (in the sense of, e.g., Werlich (1975)), or what Smith (2003) calls *discourse mode*. For *argumentative* text, the overall structure is often characterized quite well by the RST presentational relations and their nuclearity assignments, which reflect the relative weight and hierarchical structure of the points argued by the author. For *narrative*, presentational relations are much less relevant, and all the problems with subject-matter relations, as discussed in the previous section, become especially prominent. A central aspect in understanding narrative is reconstructing the temporal structure of reported events – which cannot be characterized sufficiently with RST relations. And relative "centrality to writer's purposes" is often not an issue, as illustrated with example (12) above. A similar judgement applies to *expository* text, whose function is to explain some (concrete or abstract) entity to the reader, by outlining its decomposition into parts, its function, and the like. Here, the reader is typically invited to form a mental image not of a temporal event sequence but of a spatial configuration of objects (even if they are abstract). As with narrative, the linearity of the exposition is very important for ease of processing, while internal structure according to "importance" is only occasionally relevant; nuclearity thus does not explain nearly as much as it does in argumentative text – where the linear order is much less critical.[9]

---

[9] We conducted many re-ordering experiments with the texts in the Potsdam Commentary Corpus, which indicate that the text function is relatively independent from the linear order of the segments, as long as referential chains are not broken. Roughly speaking: In presenting your argument, you have quite a bit of freedom in sequencing and arranging your points.

5. Multi-level discourse representation and annotation

Researchers investigating the different text types (cf. the citations above) point out that on the one hand the types can be distinguished by (relative frequencies of) individual surface-linguistic features, and on the other hand, coherence is created in different ways. For example, both Lötscher (1987) and Smith (2003) characterize type-specific strategies of thematic development. And we noted above that the overall function of a text varies according to the type: convey the temporal structure of events; convince the reader by presenting an argument; enable the reader to form a mental image of some, possibly complicated, object or state of affairs; etc. And along with the text function varies the shape of the functional *description* of the text, which is supposed to capture the phenomena that make the reader experience the text as coherent. To some extent, this can be explained in terms of RST relations: While argumentative text can be characterized by presentational relations, expository text features much fewer instances of those and instead is rich in Elaboration, as the text moves from (sub-) topic to (sub-) topic. In narrative, we might expect an abundance of Sequence relations, but of course the temporal structure can be much more complicated and require different means of representation (see, e.g., (Mani/Pustejovsky 2004)). Thus, different text types favour different means of describing text structure.

However, unlike *genre*, text type is not a category that texts typically belong to in their entirety. While for most texts we can clearly state that it *is* a recipe, a news report, an instruction manual, etc. (genre), it is relatively rare for a text to be an instance of just one type. Narratives typically contain descriptive, expository or other portions; a largely argumentative text may contain a short narration of events that are then being commented on, and so forth. Therefore, a single type-specific description will normally not be sufficient to sufficiently cover a text. Instead, we need arrangements of descriptive devices that are able to characterize the various dimensions of text coherence simultaneously. One undisputed facet of this notion is the long-recognized fact that both relation-based and reference-based means play their role in creating coherence. Then, taking the step from coreference to the more abstract level of thematic development, Knott et al. (2001) voiced their discomfort with RST's mixing the more "standard" coherence relations with Elaboration (in particular, the

object-attribute variant), which does not convey a relationship between propositions, but instead indicates a shifting focus of attention. Accordingly, Knott et al. dismissed that coherence relation and introduced a different representational device, which leads to 'entity chains': Individual portions of the text are captured with standard RST trees, and these portions are linked by a focus shift indicator.

My proposal is to take a further step and to seek a text representation that distinguishes various dimensions of coherence from one another. Whereas Knott et al. use a single representation that alternates between RST relations and topic shifts, it seems more appropriate to consider the dimensions in parallel: After all, some referential phenomena are bound to occur also *within* an RST-portion of an entity chain in the style of Knott et al. – thematic development and RST-like relations are by no means mutually exclusive. Similarly, temporal relations can easily overlap with non-temporal ones:

(13)   *Tim's performance at the piano was impeccable. When he had walked onto the stage, people seemed sceptical, but as soon as he finished the sonata, the whole audience burst into applause.*

Understanding (13) involves *both* reconstructing the right order of events (on the basis of verb tense and with the help of world knowledge) and grasping the contrast between the 'before' and 'after' situations.

Motivated by related observations, Wolf and Gibson (2005) responded to the multi-faceted nature of local coherence by dismissing RST's assumption that text structure be represented by a tree; instead, they allow coherence relations to quite freely connect non-adjacent segments, which can also lead to cross-dependencies. Importantly, though, they hold on to coherence relations as the single representational device, merely substituting a graph for the constrained tree structure. It seems to me that incrementally adding more relations as phenomena are being recognized, is not a promising direction. Rather, one should explicitly distinguish the various dimensions of discourse structure, so that, for example, the phenomena listed in Section 3 can be attributed to their respective realms. Incidentally, in one RST paper, Mann and Matthiessen (1991) point into this direction when they relate RST relations to the

metafunctions of systemic-functional linguistics[10] (SFL, (Halliday 2004³)), but they neglect the possibility of allowing different realms of relationships to operate *simultaneously* – which however is the underlying idea of characterizing sentence structure in SFL. Bateman and Rondhuis (1997) took the step to decompose rhetorical relations systematically along the three metafunctions (interpersonal, ideational, textual). A more cautious step into this direction had also been taken by Moore and Pollack (1992) with their suggestion that text needs to be analyzed with subject-matter and presentational relations in parallel, which can possibly lead to two divergent tree structures. However, the thesis that both types of relations *always* apply simultaneously appears to be too strong, unless one includes in the relation set some very unspecific relations that can be used in cases where just one of the two groups is clearly the appropriate one.

In order to make progress with defining an explanatory set of representation levels that accounts for the relevant dimensions of coherence, I propose a data-oriented approach that emphasizes systematic annotation of authentic text, so that ideas can be tested immediately with "real" data and be revised when necessary.

Annotation framework

Research on discourse (in particular: text) structure can be broadly divided into two kinds of approaches:

- Work that aims at extending existing theories of sentence syntax and/or semantics to the discourse level. Emphasis is on rigorous formalization and faithfulness to the theories. Usually, such approaches work with constructed examples and are gradually extended to cover more phenomena and, eventually, gain a certain coverage of „real" data (corpora of actual language use).

---

[10] The authors correlate presentational relations with the interpersonal, and subject-matter relations with the ideational metafunction; as for the textual metafunction, they vaguely indicate that the order of text spans in a relation functions textually.

- Work that aims at handling „real" data with as much coverage as possible, thus emphasizing an empirical foundation. These accounts pay the price that the theories are much less formalized and definitions are often vague.

Representatives of the first approach are SDRT (Asher and Lascarides 2003), D-LTAG (Webber et al. 2003), and LDM (Polanyi 1988). For the second approach, RST has probably been the most influential; recently, an alternative corpus-oriented approach has been proposed by Wolf and Gibson (2005). In general, recent years have seen a surge of interest in annotating text corpora with discourse-related information. One example is the Penn Discourse TreeBank (Prasad et al. 2004) that builds on a popular *Wall Street Journal* corpus and provides annotations of connectives and their arguments. For RST, the *RST Treebank* (Carlson et al. 2003) has been made available, which also covers *Wall Street Journal* text. These are important steps, because discourse research urgently needs an impetus similar to that which sentence-related annotations (so-called sentence treebanks) have given to syntax research. Such investment of annotation efforts should be carefully reflected, though. Text is a highly multi-faceted means of transmitting information (in the broadest sense), and our discussion of various salience phenomena was meant to illustrate that. Annotations according to RST are a useful starting point for exploring coherence phenomena with „real" data, but they clearly have their limitations, as the discussion in the preceding sections demonstrated. Building RST trees is a very complex task that asks annotators to make a range of quite different decisions (segmentation, nuclearity assignment, relation choice) whose individual results are then amalgamated into a single representation. In our experience, annotators are quite often unhappy after finishing a text analysis because of the many ambiguities encountered along the way: In many places, one could just as well have decided otherwise, and the alternative result would have been equally defendable in terms of the RST definitions. The trouble is that those choice points are all hidden in the tree, which does not allow for inspecting the various individual micro-decisions and for considering possible consequences of alternative answers to them.

In order to make text annotations maximally useful for research purposes, it therefore seems more helpful to separate the various phenomena from each

other and annotate them individually, in the shape of a systematic multi-level annotation (henceforth *MLA*, see (Stede 2007)). These levels can then be „mined" for correlations and inform the search for more comprehensive theories of discourse representation and processing. Until recently, this was extremely cumbersome to do (if possible at all), but nowadays with the advances in XML-based software technology, creating and querying MLA corpora is a realistic and fruitful research avenue. One possibility is the architecture developed at Potsdam University, with a generic data exchange format for diverse, task-specific annotation tools and a database that reads and aligns the individual annotations, and allows for manual cross-level analyses, as well as statistical evaluations.[11] Importantly, our MLA framework is flexible enough to allow for new levels to be added incrementally to an existing corpus. In the following, I sketch one possible configuration of annotation levels, which is at present being implemented for the Potsdam Commentary Corpus (PCC, Stede 2004), and then I relate this to RST.

Levels of annotation

Sentence syntax. A representation of syntactic structure serves as building blocks for the subsequent levels. In PCC, we use the TIGER schema (Brants et al. 2002), which is designed to be as theory-neutral as possible. It includes a constituent structure as well as information on grammatical functions. The dedicated annotation tool that allows for efficient, semi-automatic construction of syntax trees is Annotate[12].

Referential structure. Besides discourse relations, coreference is generally taken to be the most important source of coherence in text. For the time being, our annotations are restricted to links between anaphors (pronouns and definite descriptions) and nominal antecedents; both „bridging" relations and event anaphora are excluded. The units related here – referring expressions – must be licensed by the level of sentential syntax. Technically, the syntax trees are automatically mapped to the input format for the annotation tool

---

[11] http://www.ling.uni-potsdam.de/~stede/MLA.html. Technical details on the database can be found in Dipper et al. (2004).
[12] http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/annotate.html

(MMAX2[13], specifically designed for co-reference) so that the range of possible „markables" is already presented to the annotator.

Thematic structure. As a step of abstraction over the referential chains, we partition the text into zones that annotators identify as dealing with distinct topics. These zones can be embedded, and they can bear subtopic relations to one another, which the annotators assign. This builds on the idea of the 'focus stack' used in the attentional structure representation of Grosz and Sidner (1986). In addition, a specific label „Formulaic" is reserved for meta-utterances whose topic is the text itself (e.g., the first clause of sentence 4 of the text in Fig. 1) or that contain other, text-sort-specific, conventionalized formulae – in other words, for utterances that do not have a standard, „content" topic. In future work, the level of thematic structure will be linked to our work on annotating information structure in sentences (Dipper et al. 2007).

Conjunctive relations. Along the lines of (Martin 1992), this level identifies the ‚minimal units' of higher-level discourse structure, as well as the connectives linking those units in the text. From the relations proposed by Martin, we chose 10 abstract ones, and hence a set much smaller than that of RST, for labelling the type of connection; following Martin, this does not include any salience/nuclearity assignment, but relations are in general directed to reflect the (semantically) different roles of the units.[14] The range of possible minimal units is not arbitrary but constrained by constituents built on the sentential syntax level.[15] For every connective (as defined by the criteria in (Pasch et al. 2003: 331), and adding certain prepositions), annotators can mark its semantic scope (similar to the PDTB, Prasad et al. 2004), but they can also leave one or both segments unspecified, if they find it hard to decide on the scope of an adverbial connective such as *so*. Technically, we have implemented

---

[13] http://www.eml-research.de/english/research/nlp/download/mmax.php

[14] Bateman (2001) proposed a discourse representation combining RST and Conjunctive Relations, where the former is in charge of capturing abstract intentions and content, and the latter serves to model local cohesion and thematic development; hence, Bateman's goal is somewhat similar to ours.

[15] See (Dinesh et al. 2005) for a discussion of possible mismatches between the units of syntactic analysis and discourse structure.

a dedicated annotation tool that allows for marking connectives and scopes semi-automatically (Stede and Heintze 2004).

Intentional structure. This most abstract level of representation largely corresponds to the „presentational" relations of RST, with exceptions such as our treatment of Evaluation (see below). Thus it is usually only a partial structure covering not the entire text but only those portions where indeed segments are related via 'dominance' to achieve a particular intention. Marking this can involve introducing new segments (larger than minimal units), which have not been found at the conjunctive-relation level because there is no connective present. Technically, we use RSTTool[16] for this step.

Example

For the short text given in Figure 1, the annotation would proceed as follows. After the syntactic analysis of the individual sentences, co-referential NPs are being identified, which in this case leads only to the chain *I* (sentence 2), *we* (3), *we* (4), elided *we* (5). The event anaphor *that* (6) and the bridging connections (*weather – forecast, afternoon – 1pm, picnic – beer and sausages*) are at present not being represented. On the thematic level, the text is divided into segments (1-2), dealing with *weather*, and (3-6), dealing with *picnic*. The latter has a subtopic segment (4-5), *timing*. The assignment of conjunctive relations is a flat, surface-oriented annotation, which here identifies the relations Addition (1, 2), Consequence (1-2, 3), Temporal-successive (4, 5) and Addition (4-5, 6). Finally, the intentional structure corresponds in part to the presentational relations of the RST tree: Evidence (2, 1) and Motivation (1-2, 3). Regarding RST-Evaluation, notice that in the example (and in many other cases), there is no clear 'dominance' relation between the evaluated and the evaluating segment. In this case, we would annotate 6 as dominating (4-5), in contrast to the RST nuclearity assignment.

---

[16] http://www.wagsoft.com/RSTTool

RST trees versus multi-level representation

For a short and "friendly" text such as that in Figure 1, MLA might not seem particularly advantageous when compared to the RST tree, which serves as a compact representation of the text structure and also indicates its function (by showing the "most-nuclear unit", here 3). But recall, for one thing, that MLA is richer in information (syntax, co-reference), which can be used for analyzing correlations between levels. Also, the partial intentional structure of MLA also conveys the most "important" segments: the roots of the individual intention-based trees, here 3 and 6. The fact that these two are disconnected in fact seems to do justice to the text, as neither of the two speech acts supports the other.

More importantly, the picture changes when we turn to the realistic case of longer texts. In general, when comparing an RST analysis with a representation following MLA, at first sight two kinds of information appear to be missing in the latter: the subject-matter relations, and the existence of a complete tree structure. As for the subject-matter relations, they are in part replaced by the conjunctive relations. These are explicitly intended to be less specific than the RST relations, which have been criticised for their essentially open-ended nature, for example by Grosz and Sidner (1986). After all, the range of possible semantic relationships between events in the world or within people's minds is not likely to be small. From the discourse perspective, seeking a "complete" inventory of relationships seems neither fruitful nor necessary: In terms of discourse function, reporting a semantic relation between two eventualities is not so different from reporting a single eventuality, so that a fine-grained inventory of semantic relations would not correspond to an equally fine-grained inventory of consequences for text coherence. In short, a dedicated inventory of subject-matter (or semantic) relations seems to belong to the realm *of domain knowledge* but not to that of text-oriented coherence relations.

The loss of a complete tree, which after all was supposed to be the main criterion for coherence postulated by RST, might look like a rather dramatic move. However, the completeness of an RST tree (or that of other discourse representation theories) is often achieved with the help of coherence relations

of a somewhat dubious status. RST offers the multinuclear Joint relation for cases where no more specific connection can be found; the main reason why Joint is relatively rarely used seems to be the wide applicability of the Elaboration relation, which, as annotators report, fits in very many situations. Elaboration, however, has the disadvantage of masking what exactly is being elaborated – typically it is not a relationship between the propositions or eventualities, but between individual entities taking part in those eventualities. As noted by Knott et al. (2001), this amounts to a regular shift of focus of attention, which should be represented as such – in our framework on the levels of referential structure (which entities exactly are coreferent?) and thematic structure (does the overall topic change or stay the same?). The latter has the advantage of explicitly modelling the flow of topics through the discourse, which cannot be read off an RST tree. And recall that we emphasized the need for representing focus shift and coherence relations *in parallel* (in contrast to the suggestion by Knott et al.), because focus shifts occur all over the text, and not just in places where *no* intentional or subject-mater relation is present.

In the version of MLA sketched above, the hierarchy information encoded in an RST tree is distributed to three levels: the partial intentional structure (where the role of the hierarchy is largely in analogy to that of an RST tree), the flat segment annotation on the conjunctive-relations level, and the partitioning into content zones, which allow for embedding. The claim here is that this tripartite representation is in fact an advantage, as RST annotators regularly report that the precise assignment of the single hierarchy is very difficult and time-consuming – and the result is typically ambiguous anyway.

When we proceed to merge the three levels (which have been annotated individually) into a single data structure, the result in principle is a graph that in various ways violates the constraints of a tree: There can be multiple roots, nodes can have multiple parents, and there may be crossing edges (which is disallowed in RST). To what extent these violations are either annotation artefacts or indeed reflections of phenomena in the linguistic structure, is subject to empirical investigation: From the corpus-oriented perspective, we should decide on the most appropriate formal shape of text structure not based on *a priori* considerations of elegance or processing efficiency, but based on a

comprehensive picture of the variety of structuring phenomena in discourse – given analyses of authentic data.

We argued that the main goal of the multi-level approach is to separate distinct kinds of information from one another, so that they can be individually annotated (resulting in easier coding procedures) and reviewed, and ultimately set into correspondence with one another. Obviously, the amount of work necessary to produce such annotations is larger than with an RST tree, but the gain in useful information should outweigh that additional effort. In other words, a smaller number of texts annotated with a multi-level approach are likely to be more useful for discourse research than a larger number of RST trees, which are prone to the problems indicated throughout this paper.

6. Conclusion: Salience on different levels of description

The possible relationships between distinct levels of discourse representation had been lively discussed in the computational linguistics community in the early 1990s; it had been moved forward especially by research in text generation, where the successive mapping from deep representations of knowledge to the surface text needs to be made explicit. This discussion ended without a clear result, as the attention shifted – in accordance with the general trend in computational linguistics – to decidedly surface-based methods of discourse processing, then focusing on the side of understanding rather than generation (e.g., Marcu 2000). Today, it is opportune to re-open the discussion of levels and give it a decidedly data-oriented twist: Developments in annotation and retrieval software over the past years have made it possible to systematically collect text annotations on different levels and relate them to one another, either manually, or by employing statistical techniques of pattern recognition. Annotated corpora (in particular the so-called treebanks) have tremendously helped sentence-oriented research over the past decade; on the text level, this development has only started. Amongst others, it will enable us to investigate research hypotheses on level interactions that have been put forward, for example, for rhetorical structure and

constituent ordering, rhetorical structure and coreference, or discourse structure and sentential information structure.

Finally, we need to consider what happens to the idea of nuclearity in the proposed MLA approach. To some extent, it is being preserved in its original flavour, namely on the level of intentional structure. Here, it should also be possible to identify the central units of a text (with respect to its overall purpose), as it has been mentioned as an advantage of RST trees. In text portions where intention-based nuclearity is absent, salience can arise from other levels: The referential chains indicate what are the prominent entities under discussion in the text, and in which units they are being mentioned; the thematic structure identifies meta-discursive statements or formulaic portions, as well as temporary topic switches (digressions); and the syntactic structure identifies subordinate clauses, which might generate a mood of non-salience in case none of the other levels points to a different effect. Nuclearity thus becomes an epiphenomenon resulting from different types of sources of salience and non-salience – but it might just as well be absent in certain regions of the text, where segments are equally salient to the author's purposes.

References

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

Bateman, John. 2001. Between the leaves of rhetorical structure: static and dynamic aspects of discourse organization. *Verbum* 23(1): 31-58.

Bateman, John and Rondhuis, Jan-Klaas. 1997. 'Coherence relations': towards a general specification. *Discourse Processes* 24: 3-49.

Brants, Sabine, Dipper, Stefanie, Hansen, Silvia, Lezius, Wolfgang and Smith, George. 2002. The TIGER Treebank. In: Proceedings of the Workshop on Treebanks and Linguistic Theories, Sozopol,.

Carlson, Lynn, Marcu, Daniel and Okurowski, Mary Ellen. 2003. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In *Current Directions in Discourse and Dialogue*, Jan van Kuppevelt and Ronnie Smith (eds). Dordrecht: Kluwer Academic Publishers.

Cristea, Dan, Ide, Nancy and Romary, Laurent. 1998. Veins Theory: A model of global discourse cohesion and coherence. In: Proceedings of COLING/ACL'98, Montréal.

Danlos, Laurence. 2007. Strong generative capacity of RST, SDRT and dependency DAGs for Discourse. In *Constraints in Discourse*, Anton Benz and Peter Kühnlein (eds). Amsterdam: Benjamins, 2007.

Dinesh, Nikhil, Lee, Alan, Miltsakaki, Eleni, Prasad, Rashmi, Joshi, Aravind and Webber, Bonnie. 2005. Attribution and the (Non-) Alignment of Syntactic and Discourse Arguments of Connectives. In: Proceedings of the ACL Workshop "Frontiers in Corpus Annotation", Ann Arbor (MI).

Dipper, Stefanie, Götze, Michael, Stede, Manfred and Wegst, Tillmann. 2004. ANNIS: A Linguistic Database for Exploring Information Structure. In: .

Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS) 1, Shinichiro Ishihara, Michaela Schmitz and Anne Schwarz (eds), 245-279. Potsdam: Universitätsverlag.

Dipper, Stefanie, Götze, Michael and Skopeteas, Stavros (eds). 2007. *Information structure in cross-linguistic corpora. Interdisciplinary Studies on Information Structure (ISIS) 9*. Potsdam: Universitätsverlag.

Elhadad, Michael and McKeown, Kathy. 1990. Generating Connectives. In: Proceedings of the 13th International Conference on Computational Linguistics (COLING), Helsinki.

Fries, Peter. 1981. On the status of theme in English: Arguments from discourse. *Forum Linguisticum* VI(1): 1-38.

Golebiowski, Zofia. 2006. The distribution of discoursal salience in research papers: relational hypotaxis and parataxis. *Discourse Studies* 8(2): 259-278.

Grosz, Barbara and Sidner, Candace. 1986. Attention, Intention and the Structure of Discourse. *Computational Linguistics* 12(3):175-204.

Grote, Brigitte, Lenke, Nils and Stede, Manfred. 1997. Ma(r)king concessions in English and German. *Discourse Processes* 24(1): 87-118.

Halliday, Michael. 2004. *An introduction to functional grammar* (third edition). London: Arnold.

Hobbs, Jerry. 1979. Coherence and Co-Reference. *Cognitive Science* 1:67-82.

Kehler, Andrew. 2002. *Coherence, reference and the theory of grammar*. Stanford: CSLI Publications.

Knott, Alistair, Oberlander, Jon, O'Donnell, Michael and Mellish, Chris. 2001. Beyond elaboration: The interaction of relations and focus in coherent text. In *Text representation: linguistic and psycholinguistic aspects*, Ted

Sanders, Jost Schilperoord and Wilbert Spooren (eds). Amsterdam: Benjamins.

Lötscher, Andreas. 1987. *Text und Thema. Studien zur thematischen Konstituenz von Texten*. Tübingen: Niemeyer.

Mani, Inderjeet and Pustejovsky, James. 2004. Temporal discourse models for narrative structure. In: Proceedings of the ACL Workshop on Discourse Annotation, Barcelona.

Mann, William and Matthiessen, Christian. 1991. Functions of language in two frameworks. *Word* 42(3): 231-249.

Mann, William, and Thompson, Sandra. 1988. Rhetorical structure theory: Towards a functional theory of text organization. *Text* 8 (3): 243-281.

Marcu, Daniel. 1999. Discourse trees are good indicators of importance in text. In *Advances in Automatic Text Summarization*, Inderjeet Mani and Mark Maybury (eds), 123-136. Cambridge: MIT Press.

Marcu, Daniel. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge: MIT Press.

Martin, James. 1992. *English text – system and structure*. Amsterdam: Benjamins.

Matthiessen, Christian and Thompson, Sandra. 1988. The structure of discourse and "subordination". In *Clause Combining in Discourse and Grammar*, John Haiman and Sandra Thompson (eds). Amsterdam: Benjamins.

Moore, Johanna and Pollack, Martha. 1992. A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics* 18(4): 537-544.

Moser, Megan and Moore, Johanna. 1995. Using Discourse Analysis and Automatic Text Generation to Study Discourse Cue Usage. In: AAAI Spring

Symposium on Empirical Methods in Discourse Interpretation and Generation, Stanford University.

Pasch, Renate, Brauße, Ursula, Breindl, Eva and Waßner, Hermann Ulrich. 2003. *Handbuch der deutschen Konnektoren*. Berlin/New York: Walter de Gruyter.

Polanyi, Livia. 1988. A Formal Model of the Structure of Discourse. *Journal of Pragmatics* 12:601-638.

Prasad, Rashmi, Miltsakaki, Eleni, Joshi, Aravind and Webber, Bonnie. 2004. Annotation and Data Mining of the Penn Discourse TreeBank. In: Proceedings of the ACL Workshop on Discourse Annotation, Barcelona.

Sanders, Ted, Spooren, Wilbert and Noordman, Leo. 1992. Towards a taxonomy of coherence relations. *Discourse Processes* 15:1-35.

Schilder, Frank and Tenbrink, Thora. 2001. Before and after: sentence-internal and -external discourse relations. In: Proceedings of the Workshop "From Sentence Processing to Discourse Interpretation: Crossing the Borders", Utrecht.

Smith, Carlota. 2003. *Modes of discourse. The local structure of texts*. Cambridge: Cambridge University Press.

Stede, Manfred. 2004. The Potsdam Commentary Corpus. In: Proceedings of the ACL Workshop on Discourse Annotation, Barcelona.

Stede, Manfred. 2004b. Kontrast im Diskurs. *In Brücken schlagen - Grundlagen der Konnektorensemantik*, Hardarik Blühdorn, Eva Breindl and Hermann Ulrich Waßner (eds). Berlin: deGruyter.

Stede, Manfred. 2007. *Korpusgestütze Textanalyse. Eine Einführung in die Ebenen-orientierte Textlinguistik*. Tübingen: Narr.

Stede, Manfred and Heintze, Silvan. 2004. Machine-assisted Rhetorical Structure Annotation. In: Proceedings of the International Conference on Computational Linguistics (Coling-04), Geneva.

de Swart, Henriette. 1999. Position and meaning. Time adverbials in context. *In Focus. Linguistic, cognitive, and computational perspectives*, Peter Bosch and Rob van der Sandt (eds), 336-361. Cambridge: Cambridge University Press.

Taboada, Maite and Mann, William. 2006. Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies* 8(3): 423-459.

Webber, Bonnie, Joshi, Aravind, Stone, Matthew and Knott, Alistair. 2003. Anaphora and Discourse Structure. *Computational Linguistics* 29(4): 545-587.

Werlich, Egon. 1975. *Typologie der Texte*. Heidelberg: Quelle und Meyer.

Wolf, Florian and Gibson, Edward. 2005. Representing discourse coherence: A corpus-based analysis. *Computational Linguistics* 31(2): 249-287.