Discourse and Coherence: Revisiting Specific Conventions of the Centering Theory

Ruslan Mitkov and Constantin Orasan
RIATec/HLSS
University of Wolverhampton
Stafford Street, Wolverhampton, WV1 1SB, United Kingdom
{R.Mitkov,C.Orasan}@wlv.ac.uk

Abstract
This paper discusses a corpus-based study whose aim is to evaluate specific conventions of the centering theory and to establish whether they should be revisited. In particular, the study explores the relation between discourse coherence and several parameters such as the definition of an utterance, the varieties of anaphora considered, the forms of the discourse entities and the type of genre.

1. Motivation
Centering theory (CT) characterises the local coherence of a text on the basis of its discourse entities and the way in which they are introduced (Grosz et al. 1995; Walker 1998). CT has received considerable interest in recent years but has been also criticised for its validity and significance (Kehler 1997). Other studies have proposed alternative centering models (Hahn and Strube 1997; Strube 1998; Walker 1998; Kibble 2001). More recently, in extensive corpus-based studies Poesio et al. (2000) and Poesio et al. (forthcoming) tested the validity of some of the claims/rules from the centering theory on naturally occurring data. In particular, Poesio et al. (2000) investigated the validity of the claim that each utterance has exactly one backward-looking center (Cb) (apart from the first utterance in the discourse segment) and of the claim stating that if any Cf (Un) is pronominalised in Un+1, then Cb (Un+1) must also be pronominalised, and found that both these claims were subject to frequent violation. The authors experimented with different definitions of utterances (Kameyama 1998, Suri and McCoy 1994) such as sentences or finite clauses, and also treating adjuncts as embedded utterances. They also allowed a discourse entity to serve as a Cb of an utterance even if it was only indirectly referred to by a bridging reference. This led to fewer violations of the first claim but to more of the second. The study concluded that texts could be coherent even if the above claims did not hold since coherence could be achieved by other means such as rhetorical relations.

The present study is related to the work carried out by Poesio et al. (2000) and Poesio et al. (forthcoming) in that it uses corpus evidence to establish the suitability of some centering conventions. It particular, it seeks to establish the relation between specific definitions/conventions of the centering theory and the discourse coherence. Our study has to be regarded as preliminary and does not claim to be as comprehensive as the corpus study carried out by Poesio et al. (2000, forthcoming) but we believe that it yet provides different perspectives into the empirical validation of centering. It employs Karamanis’s (2001) transition scores to compute discourse coherence and investigates under what conditions (type of utterances, types of anaphoric links, type of pronouns) and for which genre (encyclopaedic, newswire) the discourse is rated as more coherent from the point of view of centering.

In computing the transitions and subsequently the discourse coherence, we have used definitions adopted in (Walker et al. 1998) and presented in Figure 1 below. These definitions are somewhat different from the ones introduced in (Grosz et al. 1995) in that continuation and retaining hold only when Cb (Un+1) = Cb (Un) (and Cb (Un+1) = Cp (Un+1)), but also in cases when Cb (Un+1) = Cb (Un) or Cb (Un) = Cp (Un+1) or Cb (Un) = Cp (Un+1)).

2. Coherence and utterances
The definition of an utterance (or at least the practical decision on what an ‘utterance’ is), has been always of interest to researchers working on centering. In most early works, the utterance has been set to a sentence with exception of Kameyama (1998) and Suri and McCoy (1994) who argued in favour of a clause. The present work seeks to establish which definition of clause provides more coherence of the discourse from the point of view of the centering transitions. In this project we

1 We computed the discourse coherence for the Grosz et al.’s (1995) definitions of transition as well and found that the scores are higher for Walker et al.’s (1998) definitions of transitions. We have not reported these results due to space constraints.

2 This distinction was proposed first in (Brennan et al. 1987).

3 Kameyama’s definition of utterances as tensed or finite clauses has recently been questioned Miltsakaki (1999) who maintained that utterances are best identified with sentences and that only the main clause should be considered for the ranking.
Coherence and indirect anaphora

In addition to the results reported in the previous section, for each definition of utterance we conducted two alternative kinds of analysis. In the first analysis only cases of direct anaphora were considered, whereas in the second one also indirect anaphora was counted. We considered the indirect anaphoric relationships part-of (Although the store had only just opened, the food hall was busy and there were long queues at the tills) and set membership (When Take That broke up, the critics gave 

Table 1: Statistics from the corpus

<table>
<thead>
<tr>
<th></th>
<th>No. of texts</th>
<th>Words</th>
<th>Sentences</th>
<th>Clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newswire</td>
<td>6</td>
<td>2,080</td>
<td>96</td>
<td>110</td>
</tr>
<tr>
<td>Encyclopaedic</td>
<td>28</td>
<td>4,156</td>
<td>155</td>
<td>190</td>
</tr>
</tbody>
</table>

Figure 1: Centering transitions

Table 1: Statistics from the corpus

<table>
<thead>
<tr>
<th></th>
<th>No. of texts</th>
<th>Words</th>
<th>Sentences</th>
<th>Clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newswire</td>
<td>6</td>
<td>2,080</td>
<td>96</td>
<td>110</td>
</tr>
<tr>
<td>Encyclopaedic</td>
<td>28</td>
<td>4,156</td>
<td>155</td>
<td>190</td>
</tr>
</tbody>
</table>

experimented with alternative units of texts being classed as utterance. The following definitions have been considered so far (i) the utterance is defined as a sentence and (ii) the utterance is defined as a finite clause in cases of coordinate clauses which are part of a complex sentence, i.e. clauses, joined by and, or, but, then.

The first definition, adopted in a number of previous experiments has the practical advantage that the utterance boundaries are identified in a straightforward manner. On the other hand, there have been a limited number of studies only which have looked at utterances as something less than a sentence (e.g. Kameyama 1998, Suri and McCoy 1994, Poesio et al. 2000, Poesio et al. forthcoming). It is important to acknowledge that the establishment of what constitutes a clause is itself a taxing task. The decision to experiment with different definitions of utterance was partly influenced by the results of our corpus analysis which showed that in at least 10 percent of the complex sentences the clauses were sufficiently independent to be utterances of their own right. In future work we would like to experiment with the notion of semantic clause to represent units which express explicitly an action irrespective whether they feature a verb or not (see section 8 for more details).

In order to quantify the coherence of a discourse we used the metrics proposed in (Karamanis 2001) based on assigning different scores to each centering transition (continuation=5, retaining=3, smooth shift=2, rough shift=0, no transition=-1) and computing the composite score for each discourse segment. The higher the composite score, the more coherent the discourse segment was deemed to be. To judge to what extent each definition of utterance accounts for coherence, we computed the global composite scores across the whole corpus for each genre. Similarly to Poesio et al. (2000) we computed the coherence in cases where indirect realisation/indirect anaphora have also been counted (see section 3). The results and the discussion are presented in section 6.

4. Coherence and discourse entities

Another convention that has been a subject of the present study is the definition of a discourse entity. In most centering studies, a discourse entity has been represented by NPs in the form of lexical noun phrases (definite descriptions or proper names) and personal pronouns. Whereas certain centering studies for Japanese and Korean (e.g. Walker et al. 1994) have covered zero pronouns, in this project we marked zero pronouns in English coordinate constructions as in the case of ‘Willie pulled the sock up quickly’. We also annotated possessive pronouns with a view to establishing whether revisiting the convention would result in better discourse coherence.

We computed the coherence in four different scenarios counting together with the noun phrases

(i) personal, possessive and zero pronouns,
(ii) personal and possessive pronouns only,
(iii) personal and zero pronouns only and
(iv) personal pronouns only.

The results of these experiments and a related discussion are presented in Section 6.

5. Coherence and genres

The purpose of this experiment was to establish whether centering is equally suitable (or unsuitable) for different types of texts. The study has been limited in that we used corpora representing two distinct genres only, but it was hoped that it would provide some indication whether some genres were more suitable from the point of view of applying centering theory than others. While the Newswire texts were much longer and were characterised of frequent turns of direct and indirect speech, the encyclopaedic texts used indirect speech only. We annotated a small corpus of 4,156 words encyclopaedic texts and a small corpus of 2,080 words of newswire texts. Table 1 provides more details in terms of number of sentences and clauses for each annotated sample.

We assessed the coherence for both genres separately as we conjectured that centering theory might not apply equally well to all genres. As the genres were represented

Robbie Williams no chance of success); generalisation, specialisation or synonymy were considered as cases of direct anaphora. As a result, a total of 4 different analyses were produced. Using the scoring methodology described in the previous section, we evaluated the coherence in cases where indirect anaphora was counted.

In order to quantify the coherence of a discourse we used the metrics proposed in (Karamanis 2001) based on assigning different scores to each centering transition (continuation=5, retaining=3, smooth shift=2, rough shift=0, no transition=-1) and computing the composite score for each discourse segment. The higher the composite score, the more coherent the discourse segment was deemed to be. To judge to what extent each definition of utterance accounts for coherence, we computed the global composite scores across the whole corpus for each genre. Similarly to Poesio et al. (2000) we computed the coherence in cases where indirect realisation/indirect anaphora have also been counted (see section 3). The results and the discussion are presented in section 6.

4. Coherence and discourse entities

Another convention that has been a subject of the present study is the definition of a discourse entity. In most centering studies, a discourse entity has been represented by NPs in the form of lexical noun phrases (definite descriptions or proper names) and personal pronouns. Whereas certain centering studies for Japanese and Korean (e.g. Walker et al. 1994) have covered zero pronouns, in this project we marked zero pronouns in English coordinate constructions as in the case of ‘Willie pulled the sock up quickly’. We also annotated possessive pronouns with a view to establishing whether revisiting the convention would result in better discourse coherence.

We computed the coherence in four different scenarios counting together with the noun phrases

(i) personal, possessive and zero pronouns,
(ii) personal and possessive pronouns only,
(iii) personal and zero pronouns only and
(iv) personal pronouns only.

The results of these experiments and a related discussion are presented in Section 6.

5. Coherence and genres

The purpose of this experiment was to establish whether centering is equally suitable (or unsuitable) for different types of texts. The study has been limited in that we used corpora representing two distinct genres only, but it was hoped that it would provide some indication whether some genres were more suitable from the point of view of applying centering theory than others. While the Newswire texts were much longer and were characterised of frequent turns of direct and indirect speech, the encyclopaedic texts used indirect speech only. We annotated a small corpus of 4,156 words encyclopaedic texts and a small corpus of 2,080 words of newswire texts. Table 1 provides more details in terms of number of sentences and clauses for each annotated sample.

We assessed the coherence for both genres separately as we conjectured that centering theory might not apply equally well to all genres. As the genres were represented
by texts of different lengths/words, we computed for each genre what we call ‘average coherence score per utterance’ defined as the global coherence score divided by the number of utterances.

### 6. Results and discussion

Table 2 presents the results of our experiments. It should be pointed out that the values in the table represent average scores for the texts, whereas the values in the brackets - the average coherence score per utterance. The results show that with regard to the types of utterances and in the case of direct anaphora, clauses account for better coherence, scoring better on most individual files and featuring a higher average per text on both genres (4.29 vs. 3.96 for encyclopaedic texts and 28.83 vs. 21.17 for newswire texts). In the case of indirect anaphora, clauses provide better coherence for newswire texts (30 vs. 23.17) but not for encyclopaedic texts (6.21 vs. 7.07). For indirect anaphora in general, the coherence is higher for both types of utterances and genres than direct anaphora (6.21 vs. 4.29 for clauses and 7.07 vs. 3.96 for sentences for encyclopaedic texts; 30 vs. 28.83 for clauses and 23.17 vs. 21.17 for sentences for newswire texts).

The results also show that with regard to the different types of discourse entities, the best coherence is obtained when all three types of pronouns (personal, possessive and zero) are counted. Ignoring possessive or zero pronouns, or both, resulted in lower coherence. While in newswire texts the average coherence score computed after counting all three pronoun types was 21.16 for sentences and 23.17 for clauses in the case of direct anaphora, these figures dropped to 19.17, 22.17, 21.17, and 20.83 when discounting possessive pronouns, and to 19.17, 22.17, 21.17, and 20.83 when discounting zero pronouns. For encyclopaedic texts counting all three types of pronouns resulted in an average score of 3.96 for sentences and 4.29 for clauses in the case of direct anaphora and 7.07 and 6.21 in the case of indirect anaphora, but the scores dropped to 2.86, 2.79, 6.43 and 5.21 respectively when discounting possessive pronouns, and to 3.86, 2.25, 6.05 and 4.43 when discounting zero pronouns.

Ignoring both possessive and zero pronouns and counting only personal pronouns which has been standard practice in centering studies so far, resulted in the lowest average coherence score of all. While for newswire texts the coherence figures were as low as 19.17 for sentences and 16 for clauses in the case of direct anaphora and 22.17 and 18.17 in the case of indirect anaphora, and for encyclopaedic texts these figures dropped to 3, 1.21, 6.11 and 3.89 respectively.

In order to compare different genres we had to compute a coherence score per utterance. As for the different types of genres, we initially conjectured that centering would not generally account well for coherence of newswire texts, one of the reasons being the frequent switches from direct speech to indirect speech. In fact, identifying clauses in newswire texts proved to be an extremely challenging task and we even questioned whether centering was suitable for discourse analysis of such texts at all. However, the evaluation results proved that newswire texts were much more coherent than encyclopaedic texts (for direct anaphora the average clause coherence score for newswire text was 1.57 as opposed to 0.63 for encyclopaedic texts; the sentence average score for newswire text was 1.32 as opposed to 0.71 for encyclopaedic texts; for indirect anaphora these figures were 1.63 vs. 0.91 and 1.44 vs. 1.27 respectively).

<table>
<thead>
<tr>
<th>Genre</th>
<th>Sentence direct (score)</th>
<th>Sentence indirect (score)</th>
<th>Clause direct (score)</th>
<th>Clause indirect (score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newswire</td>
<td>21.16 (1.32)</td>
<td>23.16 (1.44)</td>
<td>28.83 (1.57)</td>
<td>30 (1.63)</td>
</tr>
<tr>
<td>Encyclopaedic</td>
<td>3.96 (0.71)</td>
<td>7.07 (1.27)</td>
<td>4.29 (0.63)</td>
<td>6.21 (0.91)</td>
</tr>
<tr>
<td>Newswire</td>
<td>21.16 (1.32)</td>
<td>23.16 (1.44)</td>
<td>19.66 (1.07)</td>
<td>20.83 (1.13)</td>
</tr>
<tr>
<td>Encyclopaedic</td>
<td>3.85 (0.69)</td>
<td>6.5 (1.17)</td>
<td>2.25 (0.33)</td>
<td>4.42 (0.65)</td>
</tr>
<tr>
<td>Newswire</td>
<td>19.16 (1.19)</td>
<td>22.16 (1.38)</td>
<td>26.16 (1.42)</td>
<td>28.33 (1.54)</td>
</tr>
<tr>
<td>Encyclopaedic</td>
<td>2.85 (0.51)</td>
<td>6.42 (1.15)</td>
<td>2.78 (0.40)</td>
<td>5.21 (0.76)</td>
</tr>
<tr>
<td>Newswire</td>
<td>19.16 (1.19)</td>
<td>22.16 (1.38)</td>
<td>16 (1)</td>
<td>18.16 (1.13)</td>
</tr>
<tr>
<td>Encyclopaedic</td>
<td>3 (0.54)</td>
<td>6.10 (1.10)</td>
<td>1.21 (0.17)</td>
<td>3.89 (0.57)</td>
</tr>
</tbody>
</table>

Table 2: The coherence scores for different experiments

Independence from the UK was approved in 1960, with constitutional guarantees by the Greek Cypriot majority to the Turkish Cypriot minority. In 1974, a Greek-sponsored attempt to seize the government was met by military intervention from Turkey, which soon controlled almost 40% of the island. In 1983, the Turkish-held area declared itself the “Turkish Republic of Northern Cyprus,” but it is recognized only by Turkey. UN-led direct talks between the two sides to reach a comprehensive settlement to the division of the island began in January 2002 and will reach a culmination when a referendum of all Cypriots is held on 21 April 2004, just 10 days before the Greek part of Cyprus is scheduled to join the EU.
We looked at the texts to establish why the results were so much different to our initial expectations. The encyclopaedic texts turned out to be brief and trying to cover too many facts in a short paragraph, which in fact meant that they often changed the topic (see Figure 2 for an example). It would be interesting to establish whether the results for other, ‘longer’ encyclopaedic texts would be different.

It is probably not relevant to compare our results with the results reported in (Poesio et al. 2000) as the authors of that study do not compute discourse coherence scores and as the class of our finite clauses is a subset of theirs. Nevertheless, if we use the number of utterances with no Cbs as an indirectly proportional measure to discourse coherence, we see that the results are slightly different. Poesio et al. (2000) report that when taking sentences as utterances, there are less violations of constraint 1 of the centering theory (there are less utterances with no Cb) than when regarding clauses (both in Suri and McCoy’s and Kameyama’s senses) as utterances. However, if in our data we ignore zero and possessive pronouns, and count personal pronouns only (which has been typical of most centering studies), we conclude (similarly to Poesio et al. 2000) that sentences do account for better coherence than clauses. Finally, Poesio et al. (2000) establish that there are less violations of constraint 1 when counting bridging (indirect anaphora), which in our case also accounts for better coherence in most cases.

Table 3 displays the statistics related to the centering transitions. Due to space constraints we did not include pairs of transitions. However, it is worthwhile mentioning the interesting observation that contrary to the traditional belief of the Centering Theory there were a number of cases (XXX) when Retain was followed by Continuation.

### Table 3: Statistics about different types of transitions

<table>
<thead>
<tr>
<th></th>
<th>Continuation</th>
<th>Retain</th>
<th>Smooth shift</th>
<th>Rough shift</th>
<th>No transition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SD</td>
<td>SI</td>
<td>CD</td>
<td>CI</td>
<td>SD</td>
</tr>
<tr>
<td>All pronouns</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>No zero pronouns</td>
<td>11</td>
<td>13</td>
<td>12</td>
<td>14</td>
<td>32</td>
</tr>
<tr>
<td>No possessive pronouns</td>
<td>11</td>
<td>13</td>
<td>20</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>No possessive and no zero pronouns</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>28</td>
<td>30</td>
</tr>
</tbody>
</table>

7. Annotation scheme, annotation tool and annotation issues

In order to perform the experiments described in this paper, a very rich annotation had to be applied to the texts. Before we could establish which is the best annotation scheme several experiments were carried out. In a first instance we required our annotators to perform four different annotation on the same text (sentence/clause, direct/indirect realisation), each reflecting one version of the theories investigated in this paper. This type of annotation was easier, but because in some cases the differences between different types of annotation were minor, it was very easy to introduce errors.

In light of this, we switched to a more complicated annotation in which all the elements necessary to apply the centering theory were marked. The actual testing of different versions of the theory was performed automatically on the basis of the information provided in the annotation.

We decided not to use the annotation proposed in (Poesio et al. forthcoming) because we found it far too detailed and time consuming. Instead we marked a minimal set of elements necessary to run the experiments. The annotation process can be briefly described as follows:

1. Identify all the entities in the text and the relations between them. This process also included marking the zero pronouns
2. Mark the utterances (i.e. sentences and finite clauses)
3. Assign to each utterance a preferred center

For annotation we used PALinkA (Orasan, 2003) a multipurpose annotation tool. In this instance the tool proved very appropriate because it facilitated the annotation process by hiding unnecessary details.

The subordinate sentences were not regarded as utterances on their own, but as modifying a discourse entity in the main clause. In most cases utterances were centred around a verb or a verbal element. On the other hand, sentences containing semi-column or column were analysed as featuring utterances joined by semi-column or column in the case where the utterance was not set to the whole sentence. This is so because the clauses in such sentences are characterised by relative autonomy.

For the purpose of the experiments with different types of pronouns, we marked up in the corpus all instances of zero pronouns. Possessives were also annotated and in such cases first the possessive itself was marked as a discourse entity, different from the discourse entity represented by the possessed entity. Pleonastic pronouns were not assigned discourse entities.

In newswire texts (which due to the frequent changes of topics and switches from indirect to direct speech made the identification of clause boundaries very difficult), we did not regard expressions such as he said or she said as being an utterance on its own and were analysed as forming one utterance with the subordinate clause that they introduced.
8. Future work

We propose the notion semantic clause to account for the cases where nominalisations or gerunds represent a specific action as explicitly as a verb and therefore, could act equally well as utterances. To illustrate what we regard as a non-finite verb semantic clause, consider the sentence ‘Presidential and legislative elections held in October and December 2000 provoked violence due to the exclusion of opposition leader Alassane Ouattara’. In this example the exclusion of opposition leader Alassane Ouattara is analysed as a semantic clause featuring one single discourse entity (opposition leader) Alassane Ouattara. As a further illustration, in the text ‘UN-led talks on the status of Cyprus resumed in December 1999 to prepare the ground for meaningful negotiations leading to a comprehensive settlement.’ the phrase meaningful negotiations leading to a comprehensive settlement is regarded as a semantic clause featuring meaningful negotiations and a comprehensive settlement as discourse entities. Our forthcoming work will experiment with this definition of a clause and will seek to establish whether it will contribute to achieving better discourse coherence from centering perspective.

9. Conclusions

The results obtained in this study point to a number of interesting conclusions which merit further investigation. To start with, the centering theory considers a discourse more coherent in English, when possessive pronouns and zero pronouns are also counted as discourse units. In this case defining the utterance as a coordinate clause would by and large result in better coherence, than if only sentences were considered. In most cases counting indirect realisations yielded better coherence. The study also found that the sample of newswire texts was more coherent than the sample of encyclopedic texts. Whereas the study carried out and the results obtained can be regarded as preliminary, we believe that they still give an indication that it might be worthwhile revisiting some of the current conventions in centering theory – e.g. what would be the best definition of a utterance or a discourse entity, and whether indirect realisation should be counted as well.

10. Acknowledgments

We would like to thank Sarah Oxford, Aixa Perez, Rafaela Favata, Boriana Bratanova and Yarmila Daskalova for their assistance in annotating the corpus used in the present study.

11. References


