

An investigation into the use of Centering transitions for summarisation

Laura Hasler

Research Group in Computational Linguistics
University of Wolverhampton
Stafford Street
Wolverhampton, WV1 1SB, UK
L.Hasler@wlv.ac.uk

Abstract

This paper investigates the usefulness of Centering Theory, a theory of local coherence and salience, for automatic summarisation. The transitions computed for utterances in different types of texts are examined for patterns which relate to the coherence of summaries. These transitions are also employed in preliminary experiments as an evaluation method by which to judge extracts as readable and coherent.

1 Introduction

Whilst the quality of summaries produced automatically may be improving, they are still some way off the standard of those produced by humans. Although there are methods which can select some of the most important information in a text, and work has been done on coherence and cohesion in summarisation, automatic extracts still are not as coherent and easy to read as they could be. Therefore, we need to investigate and develop new means by which to improve and evaluate this aspect of the quality of automatically produced summaries.

As Centering Theory (Grosz et al., 1995) is a theory which attempts to explain local coherence and salience within a discourse, it is a prime candidate with which to experiment for the improvement and evaluation of summaries and summarisation methods, both manual and automatic. To date, very little work on summarisation from the perspective of Centering Theory has been completed ((Orăsan, 2003) uses the continuity principle for local coherence) and it seems that a first step in this direction should be to investigate the Centering transitions which appear in human-produced extracts and the transitions for the extracted sentences in their corresponding full texts to determine whether Centering really can be

useful for summarisation. Section 2 gives a more detailed description of Centering Theory.

This paper presents an investigation into a new way to try to improve the readability of extracts via an attempt to identify regular sequences of Centering transitions in human-produced extracts which could then be exploited in a system to improve the coherence in automatic extracts, as well as for the evaluation of summarisation methods. A preferred set of transitions for extracts could help in the selection of the most coherent summary out of several, meaning that the best one can be produced and presented to the user. This transition set can also be used in evaluation, to judge summaries produced by different automatic methods.

The rest of the paper is structured as follows: section 2 discusses Centering Theory and relates it to issues in summarisation. The next section details the texts used and their annotation from the Centering perspective. Results for Centering transitions in different types of texts are described in section 4. Section 5 presents further discussion of these results, followed by conclusions and future work.

2 Centering Theory (CT)

This section describes Centering Theory and discusses the basic ideas behind it, along with its rules and constraints, and the parameters which are most suitable for work on the theory within the area of summarisation. It also addresses the question of why Centering transitions can be useful for summarisation.

Centering Theory (Grosz et al., 1995) is a theory of discourse structure encompassing local coherence and salience, the development of which started in the 1980s (Grosz et al., 1986) but was not fully published until

1995. Before publication, the theory prompted research by several other authors, extending and challenging the basic notions and parameters of Centering (for example, (Brennan et al., 1987; Gordon et al., 1993; Suri and McCoy, 1994)). Since then, there have been numerous other works concerning Centering Theory (some of the most influential being (Strube and Hahn, 1996; Kameyama, 1998; Walker, 1998; Kibble, 1999; Poesio et al., In preparation)).

Whilst the most popular application for Centering Theory in the past has been anaphora resolution, the suitability of Centering variations for other computational tasks has been shown in recent years by its application in Natural Language Generation (Kibble and Power, 1999; Karamanis, 2001) and automatic summarisation (Orăsan, 2003).

2.1 Basic concepts and notions

The main concepts and presuppositions introduced in the earliest versions of Centering Theory, (Grosz et al., 1995; Brennan et al., 1987), are as follows. As the theory is one of local coherence and salience, only two utterances (although what constitutes an utterance was not fully defined in the original formulation of the theory) are considered at any one time (U_n and U_{n+1}), and each utterance in a text introduces a number of *forward-looking centers* (*Cfs*), which are NPs in U_n . These *Cfs* are partially ranked according to grammatical function, and the more highly ranked a *Cf* is, the more likely it is to be the *Cb* of the next utterance (the *backward-looking center*; the link between the two utterances). The most highly ranked *Cf* of an utterance is known as the *Cp* (the *preferred center*).

In addition, each utterance except the first has precisely one *backward-looking center* (the *Cb*), which is the link between one utterance and the previous utterance in the text. This *Cb* is the most highly ranked *Cf*(U_n) which is realised in U_{n+1} . In later papers, a weaker version of this constraint is posited, that each utterance has at most one *Cb*. If an entity within an utterance is pronominalised, it is most likely to be the *Cb*.

The relationships between *Cfs* and *Cbs* of utterances result in *transitions* between utterances, which have a definite order of preference, meaning that texts demonstrating

certain transitions are considered to be more coherent than those demonstrating others. *Continue* is preferred over *retain*, which is preferred over *smooth shift*, which is in turn preferred over *rough shift*. This ordering of transitions reflects the idea that it is preferable for consecutive utterances to have the same *Cb* and also for the most salient entity in one utterance to be the *Cb* of the next utterance. Tables 1 and 2 summarise these features of Centering Theory as described in the earlier work and presented in (Kibble, 1999), in their formalisations of rules and constraints, and describe the possible transitions.

2.2 Suitable parameters for summarisation

The original version of Centering Theory (Grosz et al., 1995) presupposes certain properties of utterances within a text, and suggests (not always fully defining) parameters which have since been found problematic by other researchers. The parameters adhered to in this paper are mainly consistent with the earlier versions of Centering Theory (Grosz et al., 1995; Brennan et al., 1987) for several reasons. However, this is not to say that these parameters are the most appropriate or best for Centering Theory applications in general, but this is not the concern here. A comprehensive overview of the different parameters which result in different versions of the theory, and advantages and disadvantages associated with them can be found in (Poesio et al., In preparation).

The parameters chosen for this study are those which best fit with current practices in automatic summarisation. An utterance is considered to be a sentence, as most summarisation systems¹ operate at sentence-level so it would not be practical to equate an utterance with only a finite clause, for example. It is also not feasible to use indirect realisation between a *Cb* and a *Cf* due to the amount of information which would need to be processed, and the fact that it is not easy to define what exactly indirect realisation could encapsulate. For these reasons, only direct

¹Including the AHRB-funded computer-aided summarisation tool, CAST, being developed at the University of Wolverhampton and related to this paper (Orăsan et al., 2003).

Constraint 1	Each utterance has precisely 1 <i>Cb</i> (Weaker version: Each utterance has at most 1 <i>Cb</i>)
Constraint 2	Every element of $Cf(U_n)$ must be realised in U_n
Constraint 3	$Cb(U_{n+1})$ is the highest-ranked element of $Cf(U_n)$ which is realised in U_n
Rule 1	If some element of $Cf(U_n)$ is realised as a pronoun in U_n , then so is $Cb(U_{n+1})$, (Strong version: if $Cb(U_{n+1}) = Cb(U_n)$, a pronoun should be used)
Rule 2	In transitions, <i>Continue</i> is preferred over <i>Retain</i> , which is preferred over <i>Smooth Shift</i> , which is preferred over <i>Rough Shift</i>

Table 1: Centering Theory rules and constraints

	$Cb(U_{n+1}) = Cb(U_n)$ or $Cb(U_n)$ undefined	$Cb(U_{n+1}) \neq Cb(U_n)$
$Cb(U_{n+1}) = Cp(U_{n+1})$	Continue	Smooth Shift
$Cb(U_{n+1}) \neq Cp(U_{n+1})$	Retain	Rough Shift

Table 2: Centering Theory transitions

realisation between entities is used to identify the *Cb* of an utterance.

The grammatical ranking of the *Cf* list for each utterance is employed: the subject of the sentence is most preferable, followed by the direct object, indirect object, and finally any other noun phrases in the sentence. Because in real life sentences can be complex and do not always comprise only one clause, where there is a main and subordinate clause, the grammatical classes in the main clause appear higher in the *Cf* list, with those in the subordinate clause coming later. This reflects the assumption that the most important information in a sentence tends to be presented in its main clause. Where there are two or more co-ordinate clauses constituting a sentence, the *Cf* list ranking reflects the linear organisation of the sentence. The weak version of Constraint 1 (each utterance has at most one *Cb*) is used, which allows two utterances to display a *no transition* between them, in cases where there is no *Cb*.

2.3 Why Centering transitions for summarisation?

As mentioned above (section 2), Centering is a theory of local discourse coherence and salience. Whilst both of these are important in summarisation, this paper is more concerned with possible ways of improving and evaluating the readability (through local coherence) of summaries. Although the transitions are based on both *Cfs* and *Cbs* from utterances, which

contribute to salience as well as coherence, Centering Theory associates more coherent texts with those displaying certain types of transitions, which are preferred over others deemed to give a less coherent text (see section 2.1).

In addition, as the aim of summarisation is to select the most important information from a text in order to present a condensed version to the reader, presumably the most salient information from the full text should already be present in the extract (although this is not always the case). By examining transitions in different sets of summaries of the same texts (human and automatically produced), and comparing both the sets of transitions and the summaries, it should be possible to identify transitions typical of good summaries, which can be used to produce more locally coherent and readable summaries. These transitions can also be used to evaluate different summaries of the same text and select the best one.

3 The annotation process

In this section, the annotation process which enabled us to make a comparison between the Centering transitions in different types of extracts and texts is discussed.

In order to assess the usefulness of the transitions for summarisation, a selection of texts from a newswire corpus already annotated for summarisation (Hasler et al., 2003), these extracted sentences in their full texts and baseline (tf.idf) summaries of the same texts

	Human extract	Full text	Baseline (tf.idf)	Total human extract
Sport	10	10	10	15
Politics	10	10	10	22
Finance / business	10	10	10	48
Total	30	30	30	85

Table 3: Breakdown of annotated texts

were further annotated from the perspective of Centering Theory, adhering to the parameters described above (section 2.2). Using *PALinkA*, a multi-purpose annotation tool developed at the University of Wolverhampton, the list of *Cfs* (*forward-looking centers*), the *Cp* (*preferred center*) and the *Cb* (*backward-looking center*) were marked for each utterance in a text in order to find the transition, which was then also added to the annotation.

To date, there are 85 human-produced extracts annotated for Centering. By human-produced extracts, we mean the sentences in each newswire text marked by a human annotator as worthy of inclusion in a summary. The extracts fall into three distinct categories or sub-genres: sport (15 texts), politics (22 texts) and finance/business (48 texts). In order to compare the transitions in human-produced extracts with transitions for the same utterances in their corresponding full texts, we annotated the source texts of the first 10 human-produced extracts in each of the three sub-genres of the corpus. Tf.idf summaries (baseline summaries) of these texts were also annotated so that the transitions of automatic baseline texts could be compared to the transitions in the human-produced summaries. Table 3 summarises the texts annotated.

In addition, 20 extracts were annotated for evaluation purposes (to assess Centering Theory’s ability to distinguish better extracts). These texts comprised two versions of 10 different extracts produced by two different people, divided into 10 extract pairs in order for a comparison to be made.

4 Results

This section describes the results of the annotation of the different kinds of texts

	Sport	Politics	Finance / Business	Total
Cont.	26	30	47	103
Ret.	12	29	37	78
SS	1	7	9	17
RS	2	10	9	21
No Tran.	59	77	206	342
Total	100	153	308	561

Table 4: Transitions for all human-produced extracts

detailed above (section 3). These results will be compared and discussed further in section 5.

4.1 Human-produced extracts

These results are for the numbers of different types of transitions found in human-produced extracts and are presented in Table 4.

The results show that while the most frequently found transition by far was *no transition*, there were also substantial numbers of *continue* and *retain* transitions. This is due to the fact that extracts are generally concerned with one or two main topics of the full text and so most of the extracted sentences should be about those topics.

Although for all three sub-genres of the corpus (sport, politics and finance/business) the highest number of any one transition was *no transition*, there were some genre-dependent preferences for *continue* and *retain* transitions, with sports texts having more *continue* than *retain*, and politics texts displaying an even distribution of these two transitions. There was also a slight preference for *continue* transitions in finance/business texts.

4.2 Human-produced extracts vs extracted sentences in their full texts

As the 30 human-produced extracts are taken from the 85 mentioned in section 4.1, it is not surprising that the results display a similar distribution of transitions. However, there was one main difference. With fewer texts to compute the transitions for, sports texts kept their preference for *continue* transitions over *retain*, but the politics extracts had a higher number of *retain* than *continue*. This means that a topic was continued, but not placed in the

	Sport	Politics	Finance / Business	Total
Cont.	20	11	11	42
Ret.	8	16	11	35
SS	0	4	4	8
RS	2	6	3	11
No Tran.	38	30	39	107
Total	68	67	68	203

Table 5: Transitions for 30 human-produced extracts

most prominent position in the next utterance. Again, the most popular type of transition was *no transition*, and the distribution of *smooth* and *rough shifts* was similar. Table 5 shows these results.

In order to examine potentially useful similarities and differences between extracts and full texts, these human-produced extract transitions were compared with the transitions exhibited by the extracted sentences in their full texts. The total number of transitions for the extracted sentences in their full texts is slightly higher because there can be one (as in the human-produced extract) or two transitions depending on the location of the first two extracted sentences in the full text. The most noticeable difference between the two sets of transitions is the much lower number of *retain* transitions in the full texts; about half the number of those present in the actual extracts. This illustrates the fact that the information about one topic in the full texts is more likely to be concentrated in one place.

Within the different sub-genres, sports texts displayed a similar distribution of transitions, whilst the number of *retains* in politics texts was reduced by half, and there was a big difference between the *retain* and *continue* transitions in finance/business texts whereas before they had been equal. Related to this, the number of *retain* and *smooth shift* transitions in this sub-genre were both very low and both the same. There were also fewer *smooth shifts* and *rough shifts*, and more *no transitions*. These results are presented in Table 6.

	Sport	Politics	Finance / Business	Total
Cont.	17	14	17	48
Ret.	7	8	2	17
SS	0	3	2	5
RS	1	4	1	6
No Tran.	43	39	47	129
Total	68	68	69	205

Table 6: Transitions for the sentences in 30 human-produced extracts in their corresponding full texts

	Sport	Politics	Finance / Business	Total
Cont.	20	11	11	42
Ret.	8	16	7	31
SS	0	4	3	7
RS	2	6	3	11
No Tran.	38	30	40	108
Total	68	67	64	199

Table 7: Transitions for 30 human-produced extracts, for comparison with baseline extracts

4.3 Human-produced extracts vs baseline (tf.idf) extracts

The 30 human-produced extracts discussed here contain the sports, politics and some of the finance/business texts already mentioned, and a few different finance/business texts. This results in a slightly lower number of total transitions and slightly fewer *retain* transitions in the finance/business texts, and consequently overall. These results can be found in Table 7.

Transitions in automatically produced baseline summaries were annotated to allow a comparison with those in human-produced extracts. The baseline summaries were automatically produced using tf.idf, a standard method for automatic summarisation. There is a small difference in the total number of transitions due to the fact that the human-produced extracts have a compression rate of 30% of the number of words in the full text, whereas the tf.idf baseline has a compression rate of 30% of the number of sentences in the full text.

Overall, the two sets of texts displayed the same number of *no transitions* and *continue*

	Sport	Politics	Finance / Business	Total
Cont.	22	7	13	42
Ret.	11	14	12	37
SS	3	3	2	8
RS	0	9	7	16
No Tran.	46	31	31	108
Total	82	64	65	211

Table 8: Transitions for 30 automatically-produced baseline (tf.idf) extracts

transitions. The baseline summaries had more *retains* and *rough shifts*, but this could be due to the higher number of overall transitions. The more noticeable differences are within the different sub-genres rather than the overall transitions, especially with regard to the distribution of *no transitions* within sport and finance/business, which were significantly higher and lower respectively than in the human-produced extracts. This is most significant in the finance/business texts as there are a similar number of total transitions for these in both text sets. Politics texts display fewer *contain* transitions, and both politics and finance/business have a higher number of *rough shifts*. Table 8 shows these results.

5 Further discussion of results

This section discusses the findings from an analysis of the transition results described in section 4, considering the transitions found in the different types of texts and the ways in which these can be interpreted and utilised within the area of summarisation.

5.1 Transitions, coherence and readability

The constraints and rules stated by Centering Theory give rise to the claim that texts which display certain transitions are more coherent than those which display others. The preference order for transitions, from most to least coherent, is detailed in section 2.1. As this is what the earlier work on Centering stipulates, this is what we shall take as our measure of better extracts: those which display most *continue* transitions.

Later, we discuss whether this analysis is upheld on reading the extracts, and use the

Centering transitions to select better extracts out of a pair of different extracts of the same text, comparing this with a human judgement on what seems to intuitively be the better one (in this case, the easiest to read/most coherent).

We have decided against using *no transitions* for the evaluation of the extracts because the presence of this transition does not generally result in unreadable or incoherent texts. In addition, the sentences present in the extracts can be, and are, taken from different positions in the full texts, often from different discourse segments. This is one reason for the high number of *no transitions*. Another reason is that we only use direct realisation between entities in two consecutive utterances to compute the *Cb*, and many of the *no transitions* were due to an indirect realisation between entities. Of the 342 *no transitions* found in the 85 human-produced extracts, over one third (127) were due to the fact that we do not take into account indirect realisation. As mentioned above (section 2.2), it is unfeasible at present to use this type of realisation in automatic summarisation, and for this reason we do not pursue it any further.

Considering the results already discussed and the measure of coherence/better extracts to be those texts displaying the most *continue* transitions, it appears that there is not a striking difference between human-produced extracts, the extracted sentences in their corresponding full texts and the baseline (tf.idf) summaries (see Tables 5-8). Overall, the number of *continue* transitions in all the types of text are similar.

5.2 Sports texts

At a lower level, the sports texts display a much higher number than either politics or finance/business texts, particularly in the baseline summaries, and should therefore be more coherent according to Centering Theory.

In order to test this, a human judge read the 10 pairs of automatic baseline summaries and human-produced extracts of the sports texts and found that this was not the case. In all cases, the human judge selected the human-produced summary as the better one. The judge did not have access to the full texts, the judgement being given on the readability and coherence of the summary as a text in its own

right. They were allowed to read both extracts at the same time, in order to compare them. The judge noted that, when looking at the two extracts together, the last 1 or 2 sentences in the human-produced extract seemed more relevant to the rest of the text than those same sentences in the automatically-produced baseline summary. The automatic extracts also tended to be more repetitive. In addition, there were some problems with references in the baseline summaries, with the judge finding it difficult to assign referents to some pronouns. This problem was not apparent in the extracts produced by humans.

5.3 Politics texts

The results also suggest that the politics texts produced automatically should be the least coherent and easy to read, with just 7 *continue* transitions out of 64. A human judge read the automatic baseline summaries and the human-produced extracts of the politics texts and found that this was generally the case. The reasons given were similar to those described above regarding the sports texts, although the politics texts did not have any problems with references. They still tended to be more repetitive, and contain less information relevant to the rest of the text, particularly in the final 1 or 2 sentences.

However, with these sets of texts, the judge reported that in one pair the baseline seemed better than the human-produced extract because the text flowed more easily. When we consider this evaluation along with that of the sports texts, we realise it could be that the baseline summaries in general, regardless of the transitions they display, are not as coherent and readable as the human-produced summaries. This would be consistent with other reports in the automatic summarisation literature.

5.4 Evaluation of "better" summaries

In order to assess the usefulness of Centering Theory for evaluating summaries and selecting the better out of a pair of extracts, a human judged 10 pairs of extracts of the same text summarised by different people. Their assessment was then compared with what Centering Theory says is the better summary out of each pair (i.e. the one with the most *continue* transitions). If there are an equal

amount of *continue* transitions, we count *retain*, followed by *smooth shift* if these are still equal, and finally *rough shift* if there are the same number of *smooth shift* transitions. If there are no differences in the texts after considering all the transitions, we say that the texts have the same level of coherence.

The human judge and Centering Theory definitely agreed in 6 out of the 10 cases. There was 1 case where the human judged it as "very close" and the transitions were exactly the same in both texts (although the extracts themselves were different). There were disagreements over three pairs of texts. In the first, the judge said that there was not enough information given in the extract selected as best by Centering Theory. The judge's background knowledge may have affected their decision in the second case, and in the third case, the extract that the judge selected was "more concise". These results show that evaluation of extracts could be the most useful application for Centering Theory in summarisation.

6 Conclusions and future work

This section draws conclusions from the results and discussion already presented in this paper.

Whilst it is clear that the readability and coherence of an extract cannot always be explained or accounted for in terms of Centering Theory's transitions, and that we cannot present a set of "ideal transitions" for summaries, we can take something away from this investigation. Although we cannot say that an extract must display certain transitions for it to be coherent and readable, or prescribe a set of transitions which coherent extracts should strictly adhere to, we found that a human judge chose extracts which did exhibit the most *contain* transitions in our exercise involving selecting the most readable text from a pair. This evaluation of "better" summaries suggests that within automatic summarisation, Centering Theory could be employed to this end. Evaluation is an important topic in summarisation as there is often disagreement on what constitutes a "good" summary, so any scale which might allow summaries to be judged in a less subjective way is a step in the right direction, however preliminary it may be.

The next step is to conduct these evaluation

experiments on a larger scale, as only a preliminary investigation was carried out here. If the results prove to be conclusive, this can be used as a means of evaluating automatic summaries produced by different methods and systems. It would also be interesting to examine the effect of re-ordering the extracted sentences forcing the text to display the most coherence-giving transitions possible, and then to compare a human reader's judgement on whether the resulting extracts are actually easier to read.

7 Acknowledgement

This paper was written as part of an AHRB-funded project, "CAST - A Computer-Aided Summarisation Tool".

References

- Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. 1987. A Centering Approach to Pronouns. In *Proceedings of the 25th Annual Meeting of the ACL*, pages 155 – 162, Stanford, California, June.
- P. Gordon, B. Grosz, and L. Gilliom. 1993. Pronouns, Names and Centering of Attention in Discourse. *Cognitive Science*, 17(3):311 – 347.
- Barbara Grosz, Aravind Joshi, and Scott Weinstein. 1986. Towards a computational theory of discourse interpretation. Unpublished manuscript.
- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203 – 225.
- Laura Hasler, Constantin Orăsan, and Ruslan Mitkov. 2003. Building better corpora for summarisation. In *Proceedings of Corpus Linguistics 2003*, pages 309 – 319, Lancaster, UK, 28 - 31 March.
- M. Kameyama. 1998. Intracentential Centering: A Case Study. In M. Walker, A. Joshi, and E. Prince, editors, *Centering Theory in Discourse*, pages 89 – 112. Oxford, UK: Clarendon Press.
- Nikiforos Karamanis. 2001. Exploring entity-based coherence. In *Proceedings of CLUK4*, pages 18 – 26, Sheffield, UK, 10 - 11 January.
- Rodger Kibble and Richard Power. 1999. Using Centering Theory to plan coherent texts. In *Proceedings of the 12th Amsterdam Colloquium*, pages 187 – 192, Amsterdam, The Netherlands, 18 - 21 December.
- Rodger Kibble. 1999. Cb or not Cb? Centering Theory applied to NLG. In *Proceedings of the ACL workshop on Discourse and Reference Structure (ACL'99)*, pages 72 – 81, Maryland, USA, 21 June.
- Constantin Orăsan, Ruslan Mitkov, and Laura Hasler. 2003. CAST: a Computer-Aided Summarisation Tool. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL2003)*, pages 135 – 138, Budapest, Hungary, 12 - 17 April.
- Constantin Orăsan. 2003. An evolutionary approach for improving the quality of automatic summaries. In *Proceedings of the Multilingual Summarization and Question Answering - Machine Learning and Beyond Workshop*, pages 37 – 45, Sapporo, Japan, 11 July.
- Massimo Poesio, Rosemary Stevenson, Hua Cheng, Barbara di Eugenio, and Janet Hitzeman. In preparation. A Corpus-Based Evaluation of Centering Theory. Technical report, University of Essex. NLE Technical Note TN-02-01, April.
- Michael Strube and Udo Hahn. 1996. Functional Centering. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL'96)*, pages 270 – 277, Santa Cruz, California, USA, 24 - 27 June.
- L. Suri and K. McCoy. 1994. RAFT/RAPR and Centering: A comparison and discussion of problems related to processing complex sentences. *Computational Linguistics*, 20(2):307 – 317.
- Marilyn Walker. 1998. Centering, Anaphora Resolution and Discourse Structure. In M. Walker, A. Joshi, and E. Prince, editors, *Centering Theory in Discourse*, pages 401 – 435. Oxford, UK: Clarendon Press.