A Reformulation of Rule 2 of Centering Theory

Rodger Kibble*
Goldsmiths College

The standard preference ordering on the well-known centering transitions Continue, Retain, Shift is argued to be unmotivated: a partial, context-dependent ordering emerges from the interaction between principles dubbed **cohesion** *(maintaining the same center of attention) and* **salience** *(realizing the center of attention as the most prominent NP). A new formulation of Rule 2 of centering theory is proposed that incorporates these principles as well as a streamlined version of Strube and Hahn's (1999) notion of* **cheapness***. It is argued that this formulation provides a natural way to handle "topic switches" that appear to violate the canonical preference ordering.*

1. What Is Centering?

Centering theory (henceforth CT) is a theory of local discourse structure that models the interaction of referential continuity and salience of discourse entities in the internal organization of a text. The main assumptions of the theory as presented by Grosz, Joshi, and Weinstein (1995) (GJW) and Walker, Joshi, and Prince (1998) can be summarized as follows:

1. For each utterance in a discourse there is precisely one entity that is the center of attention.

2. There is a preference, formalized as **Rule 2**, (1) for consecutive utterances within a discourse segment to keep the same entity as the center of attention, and (2) for the entity most prominently realized in an utterance to be identified as the center of attention.

3. The center of attention is the entity that is most likely to be pronominalized: this preference is formalized as **Rule 1**.

These principles are more precisely explicated in Section 2.

An entity may become prominent as a referential link between successive utterances, or it may deliberately be made prominent, for example, through a less oblique grammatical role or by being mentioned early in a sentence. An attraction for natural language processing practitioners has been that CT's predictions are based on easily computable structural properties of utterances rather than costly content-based reasoning.

Standard expositions of CT hold that there is a total preference ordering over types of transitions between utterances defined in terms of the tests mentioned in Point 2 above and formulated as Rule 2 in Table 1. I argue in what follows that the standard

---

∗ Department of Mathematical and Computing Sciences, Goldsmiths College, University of London, London SE14 6NW, UK. E-mail: R.Kibble@gold.ac.uk

Table 1
Centering constraints and rules (adapted from Walker, Joshi, and Prince [1998, pages 3–4]).

Constraints

C1. There is precisely one Cb.

C2. Every element of $Cf(U_n)$ must be realized in U_n.

C3. $Cb(U_n)$ is the highest-ranked element of $Cf(U_{n-1})$ that is realized in U_n.

Rules

R1. If some element of $Cf(U_{n-1})$ is realized as a pronoun in U_n, then so is $Cb(U_n)$.

R2. Continue is preferred over Retain, which is preferred over Smooth Shift, which is preferred over Rough Shift.

account of CT both over- and undergenerates. On the one hand, the stipulated preference for Retain over Smooth Shift has not been confirmed by empirical evidence, and cannot be naturally incorporated in standard generation architectures. On the other hand, there is no mechanism within the theory to predict specific cases where a Retain or a Shift may be preferred over a Continue transition, as in the Retain–Shift pattern that has been argued to signal the introduction of a new discourse topic (Brennan, Friedman, and Pollard 1987 [BFP]; Strube and Hahn 1999). I aim to overcome these difficulties under an analysis that gives a partial ordering of the classic transitions, incorporating a "streamlined" version of Strube and Hahn's notion of "cheapness" to handle the Retain–Shift pattern. I do not claim to offer new empirical results; the aim is rather to consolidate existing results in a more economic and principled formulation of Rule 2 itself.

Finally, given that CT addresses *local* rather than *global* coherence, we need to consider the question, How local is "local"? Two possible notions of local coherence are (1) overall coherence of a multi-utterance discourse segment (as in the original GJW model, which stipulates preferences for coherent *sequences* of transitions)—this has been called "not psychologically plausible from a speaker's perspective" (Brennan 1998, page 231); (2) coherence between immediately adjacent utterances, as in the BFP algorithm, which replaces the original preference for sequences of transitions with a preference ordering on transitions. In this paper, I explore an intermediate position put forward by Strube and Hahn (1999), which is a preference over *pairs* of transitions or triples of utterances, which may or may not cross segment boundaries.

2. Transition Rules

The main claims of CT are formalized in terms of Cb, the backward-looking center; Cf, a list of forward-looking centers for each utterance U_n; and Cp or preferred center, the most salient candidate for subsequent utterances. $Cf(U_n)$ is a partial ordering on the entities mentioned (or "realized") in U_n, ranked by grammatical role; for example, SUBJ $>$ DIR-OBJ $>$ INDIR-OBJ $>$ COMP(S) $>$ ADJUNCT(S). $Cb(U_n)$ is defined as the highest-ranked member of $Cf(U_{n-1})$ that is realized in U_n. $Cp(U_n)$ is the highest-ranked member of $Cf(U_n)$, and is predicted to be $Cb(U_{n+1})$.

The ranking of Cf by grammatical role has been widely adopted in the literature following BFP, though it is questioned by some researchers including Strube and

Table 2
Centering transitions (Walker, Joshi, and Prince 1998, page 6).

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

Hahn (1999), who propose a ranking based on "functional information structure," a combination of degrees of "givenness" and left-to-right order. They note that the BFP ranking is not appropriate for German, which they say is a free-word-order language (page 310); more accurately, relative order of NPs within a clause is not determined by grammatical role to the extent that it is in English. For the purposes of this paper, there is no need to commit to either BFP's or Strube and Hahn's rankings, or to go into the details of the latter's "functional centering" model, as both make the same predictions for the examples considered. Strube and Hahn themselves (page 334) state that the grammatical and functional analyses achieved consistent results for all examples in GJW. I adopt the ranking by grammatical role for purposes of exposition.

2.1 "Salience" and "Cohesion"
Transitions are defined in terms of two tests: whether the Cb stays the same ($Cb(U_n) = Cb(U_{n-1})$), and whether the Cb is realized as the most prominent NP (grammatically or otherwise): $Cb(U_n) = Cp(U_n)$. For the sake of convenience and concision, I refer to the first of these tests as **cohesion** and the second as **salience**; it is important to keep in mind that I use the terms in these defined and limited ways, and to disregard (for the time being) other uses of the terms in the literature. There are four possible combinations, which are displayed in Table 2, resulting in the named transitions Continue, Retain, Smooth Shift, and Rough Shift. The optimal case, where both salience and cohesion obtain, is Continue; the least preferred is Rough Shift. Walker, Joshi, and Prince (1998), following BFP, stipulate that Retain is preferred over Smooth Shift, which implies that cohesion is a stronger requirement than salience. However, corpus analyses reported by di Eugenio (1998, page 127), Hurewitz (1998, pages 280ff.), and Passoneau (1998, pages 338ff.) do not support this claim. In fact, all these researchers found a higher percentage of Smooth Shifts than Retains. In a spoken corpus, Passoneau found more Shifts than Continues.

A preponderance of Shifts over Continues may reflect the domain and content of a text rather than the author's organizational goals. In fact, it can be seen that sequences of Smooth Shifts are rather natural in certain kinds of narrative or descriptive texts; see Example 1 (adapted from a pharmaceutical leaflet).

Example 1
    a.   The name of your medicine[i] is Compound X.

    b.   It[i] contains budesonide[j].
       (Continue)

c. This$_j$ is one of a group of medicines called corticosteroids^k.
    (Smooth Shift)

d. These$_k$ can help to relieve the symptoms of hay fever or rhinitis.
    (Smooth Shift)

This does not appear to be an incoherent text, but there is no way that the content could be rearranged to turn the Shifts into Continues. However, we can see that the author has maintained centering coherence as far as the content allows.

We may conclude that not only does corpus evidence fail to confirm the canonical ordering, but in fact corpus analysis itself is not sufficient to evaluate the claims of CT without taking into account the underlying semantic content of a text. That is, statistics about the relative frequency of occurrences of different transition types do not in themselves tell us much about which transitions are preferred in particular situations since they do not take account of the choices available to an author.[1] A more promising approach is that of Brennan (1998), who gave subjects the controlled narrative task of providing a running commentary on a video recording of a basketball game, and used the videotape itself to construct a "propositional script" listing the sequence of events and their participants, and identifying players who were likely to continue as the center of attention over a sequence of incidents.

2.2 Rule 2 Applied to Generation

Reiter (1994) claimed that existing generation systems converged on a "consensus," generic natural language generation (NLG) architecture consisting of the following tasks:

- **Content determination/text planning:** deciding the content of a message and organizing the component propositions into a text tree;

- **Sentence planning:** aggregating propositions into clausal units and choosing lexical items corresponding to concepts in the knowledge base, including referring expressions;

- **Linguistic realization:** taking care of surface details such as agreement and orthography.

I have argued elsewhere (Kibble 1999) that if CT is to be implemented in an NLG system, the principles I call "salience" and "cohesion" belong to different tasks within this scheme: "salience" is a matter for sentence planning, choosing a verb form or some other construction that makes the *Cb* prominent within a clause or sentence, while "cohesion"—ordering propositions in a text to maintain referential continuity—is a matter for text planning. So there may be no single point in the generation process where the system has a choice between Retain and Shift, for instance: rather, the terms *retain* and *shift* describe the composite results of choices made at different stages of the generation task. This point is discussed in more detail in the cited paper. Referential continuity as determined by CT is only one of a number of factors determining the fluency and acceptability of generated text; see Kibble and Power (2000) for further discussion.

---

1 Corpus studies have also tended to be flawed by imprecise notions of "coreference"; see van Deemter and Kibble (2000).

3. The "Cheapness" Principle

The Cp or preferred center has a dual role in CT: in optimal transitions, where $Cp = Cb$, it highlights the center of attention of the current utterance, and it is also intended to signal the center of attention of the following utterance:

> The *preferred center* represents a prediction about the Cb of the following utterance. (Walker, Joshi, and Prince 1998, page 3)

It turns out that this informally stated aspect of the Cp is not actually made explicit in the rules and constraints of CT: transitions $\langle U_n, U_{n+1} \rangle$ are defined in terms of the Cp of U_{n+1} and the Cbs of U_n and U_{n+1}, but no definition mentions the Cp of U_n. Strube and Hahn's "cheapness" principle can be seen as the "missing link" of CT, making explicit the prediction represented by the Cp. They question the canonical ordering of transitions, partly on the grounds that this ordering fails to predict the Retain–Shift pattern that has been claimed by some researchers to signal the introduction of a new "discourse topic." The principle of "cheapness" is intended to capture the intuition that a Retain is naturally followed by a Smooth Shift and is defined as follows:

> A transition pair is *cheap* if the backward-looking center of the current utterance is correctly predicted by the preferred center of the immediately preceding utterance, i.e., $Cb(U_i) = Cp(U_{i-1}) \ldots$ (Strube and Hahn 1999, page 332)

Cheapness is claimed to minimize the inferential costs of processing sequences of utterances, and is proposed as a constraint on pairs of successive transitions as a replacement for the canonical orderings in Rule 2, which is restated as follows:

> **Rule 2′**
> Cheap transition pairs are preferred over expensive ones. (Strube and Hahn 1999, page 334)

This claim is supported by analysis of a variety of German texts. It turns out that although cheapness appears to be a sensible principle, it does not neatly partition the types of transition pairs; in particular, this principle does not necessarily hold of all Retain–Smooth Shift sequences. Strube and Hahn propose to rectify this by redefining the transitions, with an additional test $Cp(U_i) = Cp(U_{i-1})$ to subdivide Continue and Smooth Shift, resulting in new "expensive" transitions Exp-Continue and Exp-Smooth Shift. Strube and Hahn (1999, page 333) provide a table (not reproduced here) of 36 transition pairs, labeled as "cheap," "expensive," or "-" (not occurring).

In fact, it seems that the way this principle is presented is unnecessarily complicated, and on closer examination it appears to be rather weak. First, if cheapness is the only criterion considered, CT would have nothing to say about texts such as Example 1 that have *no* cheap transition pairs. So it appears unwise to simply abandon the canonical ordering altogether. Second, the constraint on transition pairs can be stated more economically in terms of *triples* of utterances. If it is the preferred case that for every transition pair in a discourse $\langle \langle U_{n-1}, U_n \rangle, \langle U_n, U_{n+1} \rangle \rangle$, $Cb(U_{n+1}) = Cp(U_n)$, then this equation also holds for each triple $\langle U_{n-1}, U_n, U_{n+1} \rangle$ and vice versa. Note also that if $Cp(U_n)$ is mentioned at all in U_{n+1}, it is by definition the Cb of U_{n+1}; so the requirement can be stated more simply as $Cp(U_n) \in Cf(U_{n+1})$. I propose that the cheapness principle should *supplement* rather than replace the principles of salience and cohesion.

A consequence of this is that the choice of Cp for an utterance U_n has to look backward to U_{n-1} (to identify the current Cb) and forward to U_{n+1}. In general, the question of which principles take precedence in cases of conflict cannot be settled in this short paper, but I adopt the following working hypothesis: the optimal case will be the one where both cheapness and salience obtain; the normal case will be the one where at least one of them is satisfied, which may be at the expense of cohesion between the current and the subsequent utterance. If the Cp is part of a "cheap" sequence, correctly predicting the upcoming Cb, but does *not* identify the *current Cb*, this will normally be signaled by a *nonanaphoric* nominal form.

Example 2 (adapted from GJW, page 217) illustrates the Retain–Shift pattern, though it does not provide unambiguous support for the proposal sketched above. The sequence (c-d-e) seems preferable to (c-d′-e′) even though the latter apparently scores better according to the canonical ranking. In both sequences, cheapness is satisfied wherever it is applicable, but the apparently less preferred sequence scores higher on salience and cohesion.

Example 2

    a.   John has had trouble arranging his vacation.

    b.   He cannot find anyone to take over his responsibilities.

    c.   He called up Mike yesterday to work out a plan.
        (Continue)

    d.   Mike has annoyed him a lot recently.
        (Retain)

    e.   He called John at 5 A.M. on Friday last week.
        (Smooth Shift)

    d′. He has been pretty annoyed with Mike recently.
        (Continue)

    e′. He got a call from him at 5 A.M. on Friday last week.
        (Continue)

From an interpretation perspective, we can address this apparent discrepancy by looking again at the interaction between Rule 1 and Rule 2. Rule 1 states that if anything is pronominalized, the Cb will be; so in (d), for example, *Mike* cannot be interpreted as the Cb because the sentence contains a pronoun in addition. So in fact the Retain transition (c-d) is maximally coherent given the options available to the reader: salience is not tested for because the Cp is not a pronoun, but both cohesion and cheapness obtain. This choice means that *Mike* is predicted to be the new Cb of (e) so cohesion will be unavailable for (d-e). In general, it appears that cheapness is most likely to be an available option—the expectation that the current Cp will be the next Cb is generally plausible—but salience and cohesion are not always available. Thus, if we take account of the options available to a reader at each stage, both versions of the discourse conform as far as possible to the principles of cohesion, salience, and cheapness.

From a production perspective the question remains, Why has the author chosen the "lumpy" sequence (c-d-e) rather than a "smooth" sequence of Continues? Some possible answers are these: the preferred sequence uses simple, canonical active verb forms, which may be easier to process; the sequence of clause-initial pronouns *He...* *He...*, and so on, in the variant sequence makes it appear "flat" and uninteresting;

584

Table 3
Reanalysis of Example 2 (Cb in **bold**).

	Co	**Sal**	**Ch**
a. John has had trouble arranging his vacation.	—	—	Y
b. **He** cannot find anyone to take over his responsibilities.	Y	Y	Y
c. **He** called up Mike yesterday to work out a plan.	**Y**	**Y**	**Y**
d. Mike has annoyed **him** a lot recently.	**Y**	N	**Y**
e. **He** called John at 5 A.M. on Friday last week.	N	**Y**	—
d′. **He** has been pretty annoyed with Mike recently.	**Y**	**Y**	**Y**
e′. **He** got a call from him at 5 A.M. on Friday last week.	**Y**	**Y**	—

the author is realizing a communicative goal to say something about John in (a-b-c) and something about Mike in (d-e). The bottom line is that from a generation point of view, *centering is not enough*. Maximizing coherent transitions will not in itself produce optimally fluent and readable text; instead, a number of other factors have to be taken into consideration in order to minimize the inferential load on the reader, hold the reader's interest, and reflect communicative intentions.

Both versions of the text are preferable to one where the last two sentences have different subjects: for example, (d) followed by *(e″) John....* The intuition is that once the topic has changed, the discourse must stay with the new topic rather than "flip-flop" between two entities. This intuition can be sharpened by noting that (d-e) form an identifiable embedded discourse segment, whose subject matter is not directly related to the main issue of John's vacation plans. I conjecture that absence of salience is *not* penalized in *segment-initial* utterances as long as cheapness holds.

The following restatement of Rule 2 is intended to bring out the Janus-like nature of CT, simultaneously looking backward and forward:

> **Rule 2″**
>
> **Cohesion:** Prefer transitions $\langle U_{n-1}, U_n \rangle$ where:
> $Cb(U_n)$ is defined and
>   if $Cb(U_{n-1})$ is defined, $Cb(U_n) = Cb(U_{n-1})$.
> **Salience:** Prefer transitions $\langle U_{n-1}, U_n \rangle$ where $Cp(U_n) = Cb(U_n)$
>   [if $Cp(U_n)$ is pronominal]
> **Cheapness:** Prefer sequences $\langle U_{n-1}, U_n, U_{n+1} \rangle$ where $Cp(U_n) \in$
> $Cf(U_{n+1})$
> **Conditions:**
>   1. In case of conflicts, the following ordering is
>   hypothesized: {**cheapness** | **salience**} >
>   **cohesion**
>   2. If U_n is segment-initial, **salience** is not required
>   if **cheapness** holds.

Table 3 illustrates an analysis of Example 2 in terms of the interacting constraints. Note that the absence of salience against (d) is not penalized for reasons explained above.

585

This example suggests a need for optimization over sequences of more than two utterances. In a sequence Continue–Retain–Smooth Shift, the Shift is predicted in its local context but the Retain is not; although Retain is a cheap transition following Continue, another Continue would be cheap as well. The Retain is motivated as it allows a new topic to be introduced with a "cheap" Smooth Shift, and so we need to evaluate the whole sequence Continue–Retain–Smooth Shift. This illustrates that while a sequence that conforms to the cheapness principle may reduce the cognitive load on the hearer, it can actually *increase* the load on the speaker owing to the need to plan ahead beyond the current utterance. In fact, the proposals outlined here do not entail that speakers must plan the entire content and structure of sentences so far in advance. Rather, they entail that a speaker knows when uttering U_n that he or she intends to express a particular fact about a particular entity E in utterance U_{n+2}; and it entails that the speaker would do well to prepare the hearer for this by making E prominent in utterance U_{n+1}. The hypothesis is that speakers will make a degree of effort to help hearers to process their utterances smoothly, rather than opportunistically planning and realizing sentences one by one, but not to the extent of planning all the transitions in a discourse segment in advance of uttering anything.

4. Conclusion

Comparison of the standard preference ordering for centering transitions and Strube and Hahn's (1999) variant has established the following points:

1. The strict ordering of canonical transitions assumed by GJW and others has not (so far) been confirmed by corpus evidence and does not naturally fit into generation architectures. There is no mechanism to predict the Retain–Shift sequence to introduce a new discourse topic.

2. By reducing Rule 2 to a requirement for cheap transition pairs, Strube and Hahn weaken the predictive power of the theory while complicating the apparatus with two additional transitions and a 36-position table of cheap versus expensive pairs.

I have argued that in fact we can dispense with not only Strube and Hahn's two new transition types but the four old ones as well, retaining them only for descriptive convenience. The various different transitions can be seen to *emerge* in a partial, context-dependent ordering as a result of the interaction of cohesion, salience, and cheapness. Following established practice in empirical work such as that discussed in Section 2.1, centering coherence is applied to inter- as well as intrasegmental transitions. The modified proposal is still weaker than GJW's original formulation as a wider variety of texts is tolerated. This underscores the fact that referential continuity as specified by CT may play an essential part in computing the overall coherence of utterance transitions but it is only one of the determinants of discourse structure.

References
Brennan, Susan E. 1998. Centering as a psychological resource for achieving joint reference in spontaneous discourse. In

Marilyn Walker, Aravind K. Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*. Clarendon Press, Oxford, pages 227–249.

Brennan, Susan E., Marilyn Walker Friedman, and Carl Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 155–162.

di Eugenio, Barbara. 1998. Centering in Italian. In Marilyn Walker and Aravind K.

586

Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*. Clarendon Press, Oxford, pages 115–137.

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

Hurewitz, Felicia. 1998. A quantitative look at discourse coherence. In Marilyn Walker, Aravind K. Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*. Clarendon Press, Oxford, pages 273–291.

Kibble, Rodger J. 1999. Cb or not Cb? Centering theory applied to NLG. In *Proceedings of ACL Workshop on the Relation of Discourse/Dialogue Structure and Reference*, pages 72–81.

Kibble, Rodger J. and Richard D. J. Power. 2000. An integrated framework for text planning and pronominalisation. In *Proceedings of the 1st International Conference on Natural Language Generation*, pages 77–84.

Passoneau, Rebecca. 1998. Interaction of discourse structure with explicitness of discourse anaphoric noun phrases. In

Marilyn Walker, Aravind K. Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*. Clarendon Press, Oxford, pages 327–358.

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

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

van Deemter, Kees and Rodger Kibble. 2000. On coreferring: Coreference annotation in MUC and related schemes. *Computational Linguistics* 26(4):615–623.

Walker, Marilyn, Aravind K. Joshi, and Ellen Prince. 1998. Centering in naturally occurring discourse: An overview. In Marilyn Walker, Aravind K. Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*. Clarendon Press, Oxford, pages 1–28.