Chapter 3

ZEROS and Coherence in Japanese Discourse

3.1 Discourse coherence

3.1.1 Key concepts

In order to discuss discourse coherence as dealt with in this section, let us begin with defining some key concepts. A discourse refers to a set of language forms that are produced and interpreted as people communicate with each other. As such, it cannot be independent of the purposes or functions that those forms are designed to serve in human activities. A discourse may be written or spoken, and usually consists of two or more sets of utterances that are coherently linked and situated in context.¹

The term utterance, likewise, is defined to be an expression uttered or written by a particular speaker or writer at a particular time and at a particular location for a particular purpose. Utterances thus contrast with possible linguistic constructs, such as sentences or clauses, which do not reside in any specific time and space. That is, there is no utterance possible without a context.

A discourse is considered to be segmented. Factors that determine the boundaries of discourse segments (DSs) have been of great interest to many discourse analysts (e.g., Brown and Yule, 1983; Polanyi, 2001); a variety of segmentation algorithms, each focusing usually on a single factor, have been proposed by computational linguists (e.g., Morris and Hirst, 1991; Kozima, 1993; Nomoto and Nitta, 1994; Hearst, 1997; Passonneau and Litman, 1997). The determination of segment boundaries is so complex that researchers have reached little agreement about it. For a written discourse, paragraphing is sometimes employed as a conventional indicator of a

¹ A piece of discourse in context can also comprise as little as one or two words, as in ‘Stop’ or ‘No Smoking’ (Celce-Murcia and Olshtain, 2000, page 4).
A discourse model is an internal representation held by a discourse participant that links linguistic forms (referring expressions) to some referents, e.g., particular individuals, objects or events in some real world. When an addresser uses a referring expression, he or she is specifying, for example, a specific individual in his or her discourse model, with the intention of having the individual introduced or identified in the addressee’s discourse model.

A discourse model contains a set of discourse entities that are elicited in the discourse, along with their properties and their relations with other entities. Discourse entities are the real, abstract, or imaginary objects introduced by the discourse, and contrast with referring expressions that are linguistic mentions of the discourse entities throughout the text. One may have in mind a particular person, and may refer to this person in one context as ‘John,’ in another as ‘the man who won the race,’ in yet another as ‘he’ or whatever is linguistically possible (Chafe, 1976). Discourse entities may be evoked by the discourse via explicit (or implicit) linguistic mention; otherwise, entities can be inferred within the discourse model due to generic or specific knowledge of entities and relations holding among them. Entities may also be situationally evoked (Prince, 1981).

Several taxonomies for discourse entities have been proposed in the literature. The given (or old)-new distinction proposed by Chafe (1976) is one of the first classifications that consider psychological or cognitive status of entities. He defines “given” as what the speaker believes is in the hearer’s consciousness, and “new” as what the speaker believes is not. Prince (1981) elaborates on this distinction and, by adapting a discourse (rather than hearer) centric view, defines entities, when first introduced in discourse, as “discourse-new,” including “brand-new” when the hearer must create a new entity in his/her discourse model, and “unused” when the hearer already knows of this entity. Evoked entities that are already in the discourse are considered to be “discourse-old,” and are further classified into “textually evoked” and “situationally evoked.” In a similar vein, Gundel, Hedberg, and Zacharski (1993) claim, attending to the cognitive status of discourse entities, that their accessibility is reflected by linguistic forms.

A discourse is not a mere sequence of utterances. For a set of utterances to be a discourse, it must exhibit coherence. Coherence, however, is a cognitive state; it is not in the language itself, but is rather perceived by the language users, who unite utterances into a coherent representation of discourse. These utterances may contain linguistic devices that help the perceiver (including the speaker/writer and the hearer/reader) in establishing coherence. Speakers/writers utilize such linguistic devices, called cohesion, and hearers/readers recognize them, to establish coherence in discourse, often

---

2 “Discourse entity” is a term first introduced by Webber (1979); this term is equivalent to “discourse referent” as used by Karttunen (1976).
supplemented with their knowledge of the world and so on.

The recognition of cohesion in the linguistic input leads to a better perception of coherence or a more coherent mental representation of the discourse, and hence to better comprehension. However, this alone is not sufficient. Comprehension is a complex cognitive process that also involves extensive inferential processes drawn on knowledge of the world as well as on memory for the preceding discourse. Inferencing that takes place in the comprehension process is a second major mechanism in creating coherence, after cohesion. Inference can be defined as any piece of information that is not explicitly stated in a discourse, but is required to establish a coherent mental representation of the discourse. Not all inferences, therefore, are of the same sort. Researchers in psycholinguistics and discourse processing have proposed several typologies of inferences (e.g., Clark, 1977). The amount of inference required or the processing cost has also been of considerable interest in comprehension research (e.g., Shiro, 1994).

We have reviewed some important concepts for the present study of discourse coherence, which we assume are important in discussing ZEROS, i.e., invisible discourse entities, as well. Our major concerns are: (i) how much of a contribution ZEROS as a unique linguistic device make to discourse coherence, and (ii) how much inference cost they require in comprehending a certain discourse segment or a certain sequence of utterances that contain ZEROS.

All the terminology reviewed here (in bold above) is applicable to both written and spoken discourse. Our focus is, however, on written Japanese monologues, which we call “text” in this thesis. Spoken forms of discourse, such as dialog or conversation, will be excluded from our discussion for the remainder of the thesis.

3.1.2 Approaches to coherence

Discourse coherence concerns the way in which utterances are related to each other in a reasonably meaningful fashion, and many researchers have attempted to describe coherence in terms of the relations among utterances within a discourse.

Such attempts have been made, in the literature, from two major distinct views of coherence. One prominent approach is to characterize the possible ways in which successive utterances can be connected to form a coherent discourse representation, and to enumerate such characterizations in a list of “coherence relations” (e.g., Halliday and Hasan, 1976; Hobbs, 1979; Mann and Thompson, 1987; Kehler, 1995, 2002). Kehler (1995, 2002), for instance, presents a list that includes three major classes, “resemblance,” “contiguity,” and “cause or effect,” and the subclasses therein. This approach concerns the “relational coherence” of discourse.

The other approach, in contrast to relation-based theories of coherence, views coherence in regard to repeated reference to the same entity or event in a discourse.
One of the earliest studies in this approach was made by Kintsch and van Dijk (1978), who proposed that propositional information, abstracted from the incoming text, is connected to previous propositional information via “argument overlap.” In a similar vein, Givón (1983) argues for “topic continuity” as one aspect of the complex process of continuity in discourse. This type of approach focuses on “referential coherence” in a discourse, or “entity coherence” as it is called by Poesio, Stevenson, Cheng, Di Eugenio, and Hitzeman (2002), Poesio, Stevenson, Di Eugenio, and Hitzeman (2004), and Karamanis (2003). One important work in this strand presents Centering Theory (Grosz, Joshi, and Weinstein, 1995).

These two approaches are not exclusive of, but rather complementary to each other, as Kintsch and van Dijk (1978) have stated. It is probably a matter of priority or focus of discussion whether one takes the former approach or the latter. In this study, we take the latter approach and adopt the centering model as an explanatory tool that measures coherence of discourse, in relation to a specific referring expression, i.e., ZEROS. This approach makes sense because ZEROS that we are concerned with are “entities” in discourse.

3.2 Centering theory

We have overviewed some major concepts surrounding discourse coherence and approaches to coherence. We adopted, from among different approaches to discourse coherence, Centering Theory, which was officially formulated in Grosz, Joshi and Weinstein (1995; hereafter GJW95) and two previous works (Grosz, Joshi, and Weinstein, 1983, 1986; hereafter GJW83, 86).

The development of Centering Theory has been based mainly on two different strands of background work. Firstly, Grosz and Sidner (Grosz, 1977; Sidner, 1979; Grosz and Sidner, 1986) acknowledged the “attentional state” as a basic local-level component of discourse structure, and proposed that it consisted of two levels of focusing: global and local. In Grosz and Sidner’s account, centering delivered a model for monitoring utterance-by-utterance changes in the local focus of attention. Secondly, Joshi, Kuhn and Weinstein (Joshi and Kuhn, 1979; Joshi and Weinstein, 1981) proposed centering as a model of the complexity of inferencing required in discourse comprehension. They attempted to explicate how each utterance is integrated into the preceding discourse and is linked to the succeeding discourse, in relation to the inferential complexity involved.

The successful merger of these two lines of work resulted in the original version of the centering model, which accounts for the attentional state factors that are responsible

---

3 Poesio et al. (2004) suggested, from their corpus analysis results, “a hybrid view of coherence” that integrates an entity-based account of coherence with rhetorical and temporal coherence, and other factors.
for the differences in perceived degrees of coherence of discourses that convey the same information. One such discourse pair taken from GJW95 is provided in (3.1) and (3.2) below.

(3.1)  a. John went to his favorite music store to buy a piano.
       b. He had frequented the store for many years.
       c. He was excited that he could finally buy a piano.
       d. He arrived just as the store was closing for the day.

(3.2)  a. John went to this favorite music store to buy a piano.
       b. It was a store John had frequented for many years.
       c. He was excited that he could finally buy a piano.
       d. It was closing just as John arrived.

Discourse (3.1) is perceived as noticeably more coherent than discourse (3.2). Centering explains this difference as the outcome emerging from different degrees of continuity in what the discourse is about. Discourse (3.1) centers around a single individual (John), and hence, is clearly about ‘John.’ Discourse (3.2), in contrast, seems to focus in and out on two different entities (John, store, John, store). Centering is intended to capture these variations in (dis)continuity in focus.

3.2.1 Main claims

Centering theory is an entity-oriented theory of discourse coherence (see 3.1.2 above). It intends to model the local mechanisms that create local coherence by operating on the discourse entities in each utterance within a discourse segment. The fundamental assumption of centering is that humans continuously update the local attentional state or local focus as they incrementally process a discourse.

The local focus contains a set of FORWARD-LOOKING CENTERS (CFs), along with the information about the relative salience or RANKING of these CFs. The local focus gets renewed after every UTTERANCE within a DISCOURSE SEGMENT. In this renewal, the current CFs are updated into new ones. The set of CFs introduced in the local focus by utterance $U_i$ is presented as $CF (U_i)$. The members of $CF (U_i)$ are defined as discourse entities that are REALIZED in $U_i$. One unique entity of the $CF (U_i)$ is called the BACKWARD-LOOKING CENTER or $CB (U_i)$. The $CB (U_i)$ links the current utterance to the previous discourse. The intuition that some discourses are perceived to be more coherent than others, as illustrated in (3.1) and (3.2) above, is stipulated such that one way of updating this $CB$ is preferred over another. Different ways of updating are formulated as the TRANSITION types that each utterance is labeled with.

The primary claims of the centering theory in GJW are given in two proposed
centering rules: **Rule 1** establishes constraints on the realization of entities mentioned in an utterance; and **Rule 2** claims a difference in inference load for different centering **TRANSITION** states between utterances.

The precise definitions of these theory-unique terms (indicated in **SMALL CAPS** above) are largely left unspecified, allowing for “a large number of possible instantiations of the theory” (Poesio et al., 2004, page 310). Before we give the definition of centering terms that we employ for this study (in 3.2.3 below), we will briefly overview in the next section previous areas of the application of centering and present our objectives, which affect our own definitions.

### 3.2.2 Applications of the theory

Centering Theory is one of the most influential frameworks in the study of discourse. Since the early development of the theory (GJW83, 86), it has been adopted as the basis for numerous works mainly in computational linguistics. The claims about pronominalization made in Rule 1 have been applied to develop algorithms for both **anaphora resolution** (e.g., Brennan, Friedman, and Pollard, 1987) and **anaphora generation** (e.g., Dale, 1992; Henschel, Cheng, and Poesio, 2000; Yuksel and Bozsahin, 2002). Ideas deriving from Rule 2 about preference order for **TRANSITIONS** have been increasingly found useful in **text structuring/planning** (e.g., Karamanis, 2003; Kibble and Power, 2004). Some predictions of the theory have also been tested with (and verified by) **psychological experiments** (e.g., Gordon, Grosz, and Gillion, 1993; Hudson-D’Zmura and Tanenhous, 1998).

One promising but rather inactive application area of centering is to **language learning/teaching**. One near-track application is found in work by Suri and McCoy (1993a) in which they utilize Sidner’s (1983) local focusing in their algorithm for identifying illegal NP omissions and inappropriate pronominalization in the CALL system designed for native signers of American Sign Language learning English as a second language. In a similar vein, our earlier work attempts to assist learners of Japanese in their interpretation and production of **ZEROS** with centering-based algorithms (Yamura-Takei, Fujiwara, and Aizawa, 2001a, 2001b; Fujiwara and Yamura-Takei, 2002a, 2002b, 2002c).

One significant work in an attempt to use centering for practical application in an educational context is (Mitsakaki and Kukich, 2000a, 2000b, and 2004), in which **ROUGH-SHIFT TRANSITION** was used as an indicator of incoherence in students’ essays. More recently, Tanimura (2004) employs Centering Theory as an explanatory tool to show the contrast between native speakers and learners of English with regard to coherence establishment and choice of referring expressions. This work is one of only a few that utilize centering in second language acquisition research.

Our present study also aims at an educational application of the theory, but in a
slightly different way from our earlier work. Kameyama (1985) views centering mechanisms as part of “linguistic competence” operational in human discourse production and comprehension (page 91), and also describes the mechanisms as a “hypothetical cognitive process involved in discourse processing of any human language” (page 94). We subscribe to this view and utilize a centering analysis as a measure in our attempt to explicate human perception of coherence and demand on inference in processing ZERO-containing discourses, in the belief that the findings provide significant pedagogical implications.

3.2.3 Concepts and definitions

Centering Theory is a conceptual framework for theorizing about local coherence; some notions and definitions are left unspecified, and rules are provided as preferences, rather than as hard rules. This has motivated much subsequent work that attempts to make further specifications, reformulations and extensions of the theory (inter alia, Walker, Iida, and Cote, 1994; Strube and Hahn, 1999; Kibble, 2001). Many such attempts have been made to develop efficient algorithms for anaphora resolution and generation, and to attempt cross-linguistic applications of the theory. As Poesio et al. (2004) describes Centering as a “parametric theory;” it allows for language-by-language “parameter” setting (cf. Walker, Iida, and Cote, 1994). That being the case, it should even allow for analysis-by-analysis setting so that it can best suit the objectives of particular applications of the theory.

In this study, as well, we set parameters so that they may best fit the purpose of our analysis, by either choosing from among a variety of previous parameter settings reviewed comprehensively by Poesio et al. (2002, 2004), or making necessary revision and further elaboration.

3.2.3.1 Utterance and discourse segment

The definition of utterance, as a basic CENTER update unit, is a crucial one. There has been a debate concerning how and the previous utterance should be regarded (inter alia, Suri and McCoy, 1993b; Kameyama, 1998; Miltasakaki, 2003). We will follow the suggestions of Kameyama (1998) and consider the basic utterance unit of centering to be the tensed clause. Using this approach, we analyze the CFSs and CBSs of each clause in a linear manner such that the centering output of one clause is the input to the analysis of the next adjacent clause.

Centering is meant to capture within-segment coherence. Therefore, segment boundaries are as important a concept as utterance boundaries. As we mentioned earlier in 3.1.1, however, reliably identifying segment boundaries is extremely difficult. Therefore, some heuristics have been employed in the centering literature. One is to treat a whole text as one discourse segment, ignoring any other possible segmentation. Some researchers use surface linguistic structure, such as paragraphing (Miltsakaki, 2003) and subsection (Poesio et al., 2004), as a conventional indicator of segments. We regard, in this study, a paragraph as a discourse segment, mainly because it is clearly indicated (by indenting and/or line spacing) in our corpus.

3.2.3.2 CENTERS and realization

The term CENTERS is used to represent “semantic objects, not words, phrases, or syntactic forms” (GJW95, page 208). CENTERS are entities constructed in a discourse in which they occur, thus a sentence in isolation does not have CENTERS. GJW used the notion REALIZE to define the relation between utterance (U) and CENTERS (c), and to relate CENTERS to linguistic expressions, as given in (3.3).

\[(3.3) \quad \text{U directly realizes c if U is an utterance of some phrase for which c is the semantic interpretation.}\]

Two linguistic options for English that GJW provide for an “NP that directly realizes c” are a definite description and a pronoun. GJW also discuss another possibility for the realization relation: c is “realized but not directly realized” (GJW95, page 217) in case of utterances containing NPs that express functional relations (e.g., ‘the door’) whose arguments have been directly realized in previous utterances (e.g., ‘a house’).

In sum, GJW95 consider two possible ways in which a discourse entity may be “realized” in an utterance: DIRECT realization and INDIRECT realization (cf. Poesio et al., 2004, page 9).

As for Japanese, Kameyma (1985) proposed that “pronouns” in English, as a direct realization of c, correspond to “zero pronominals [in her terminology]” in Japanese with respect to the interactions with centering. Walker, Iida and Cote (1990, 1994) followed this proposal, and claimed that “zero pronouns [in their terminology]” are “realized from information specified in the subcategorization frame of the verb” (1994, page 199). We subscribe to this view and treat ZEROS as a type of “direct” realization. We also regard ZEROS as “implicitly” realized entities, in order to make a contrast to linguistically “explicit” realizations, such as NPs and pronouns that are visible.

As for indirectly realized CENTERS that are claimed to play a crucial role in maintaining CENTERS, and hence in creating coherence (e.g., Hahn, Markert and Strube, 1996; Strube and Hahn, 1999), no work in Japanese explicitly includes them in the
centering analysis, to the best of our knowledge. However, Japanese, as well, does exhibit functionally dependent anaphoric relations, as in the example below.

(3.4)  

a. 家が ある。

`ie-ga aru.`

‘There is a house.’

b. 屋根は 赤い。

`yane-wa akai.`

‘The roof is red.’

Recall our discussion of zero nominal arguments in Chapter 2. We give a different treatment to this kind of indirect realization, in which two entities are indirectly related via functional dependency, or elsewhere called “bridging” and “association.” In most previous work, mainly for English, an entity ‘roof’ in (b) is considered to be an indirect realization from the previously mentioned entity ‘house.’ We see this instead as a direct realization of an entity `ie` by way of an implicit argument of the entity, `yame`, as illustrated in (3.5).

(3.5)  

Ø 屋根は 赤い。

`(Ø-no) yane-wa akai.`

`(Ø-GEN) roof-TOP is-red.`

‘The roof (of Ø) is red.’

Indirect realization, either explicitly or implicitly, in our definition, can be illustrated in (3.6) and (3.7) respectively.

(3.6)  

a. 太郎が コンビニを 探していると、

`Taro-ga konbini-o sagasite-iru-to`

‘Taro was looking for a convenience store, and’

b. 角に セブンイレブンが あった。

‘There was a 7-11 corner.’
corner-at Seven Eleven-NOM be-PAST

‘Seven Eleven was at the corner.’

In this example, an entity ‘convenience store’ in (a) is realized as “inferable” in (b) in the form of another head noun ‘Seven Eleven’ that naturally evokes the entity in the discourse model updated after the utterance (a) is heard. Thus, ‘Seven Eleven’ is an indirect and explicit realization of a ‘convenience store.’

(3.7)  a. 太郎が  セブンイレブンに  入ると、

Taro-ga sebunirebun-ni hairu-to

Taro-NOM Seven Eleven-in enter-when

‘When Taro entered Seven Eleven,’

b. Ø いきなり Ø あいさつされた。

(Ø-ga) ikinari (Ø-ni) aisatu-sare-ta

(Ø-NOM) abruptly (Ø-by) greet-CAUS-PAST

‘(he ‘Taro’) was abruptly greeted by (them ‘shop clerks’).’

In (3.7), on the other hand, the entity ‘Seven Eleven’ in (a) evokes an entity ‘sales clerk’ in (b) which is realized as a ZERO. Here, a ZERO that denotes ‘sales clerk’ is an indirect and implicit realization.

Possible linguistic options for CENTER realization in Japanese can be summarized in a two-by-two classification, according to their “directness” and “explicitness,” in Table 3.1.

<table>
<thead>
<tr>
<th>Explicit (non-ZERO)</th>
<th>Implicit (ZERO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td></td>
</tr>
<tr>
<td>[A] same head anaphoric</td>
<td>[C-1] zero verbal argument,</td>
</tr>
<tr>
<td></td>
<td>[C-2] zero nominal argument</td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
</tr>
<tr>
<td>[B-1] same entity/different head associative</td>
<td>[D] zero argument with situational reference (inferable, situationally evoked),</td>
</tr>
<tr>
<td></td>
<td>[B-2] situationally evoked, inferable</td>
</tr>
</tbody>
</table>

Table 3.1: Linguistic options for CENTER realization in Japanese

An example entity pair for each of the six sub-groups is given below, out of which [B-2]
[D] (highlighted) are excluded from our definition of realization.

(A) 豪華客船 客船
gooka-kyakusen kyakusen
luxury passenger boat passenger boat

(B-1) タイタニック 船
taitanikku hune
Titanic boat

(B-2) 客船 港
kyakusen minato
passenger boat harbor

(C-1) 客船（客船）
kyakusen (Ø kyakusen)
passenger boat (Ø ‘passenger boat’)

(C-2) タイタニック（タイタニックの）乗客
taitanikku (Ø taitanikku-no) zyookyaku
Titanic ‘passenger (Ø ‘of the Titanic’)

(D) タイタニック（乗客）
taitanikku (Ø zyookyaku)
Titanic (Ø ‘passenger’)

[B-1] and [C-2] are normally combined as one type under the name of what is elsewhere called “bridging” and “associative;” the distinction depends on whether or not the relation between the two entities can be expressed in A no B form (see Chapter 2 for detailed discussion on [C-2] type). Inclusion of zero nominal arguments [C-2] in our definition of realization is one novel aspect of the centering analysis that follows.  

---

5 To the best of our knowledge, no work for Japanese has explicitly included zero nominal arguments in a centering analysis and discussion. The inclusion shows that the interaction of zero verbal arguments and zero nominal arguments in the centering mechanism interact in the same way that personal and possessive pronouns (in bold below) interact in the following example discourse in English used in Kameyama (1998, page 104).

1. Her entrance in Scene 2 Act I brought some disconcerting applause
2. even before she had sung a note.
3. Thereafter the audience waxed applause happy
4. but discriminating operagoers reserved judgment
5. as her singing showed signs of strain
6. her musicianship some questionable procedure
Excluded is [D] type, what we define as ZEROS with situational reference (see 2.4.2.6 for the definition and wait for 4.4.1.3 for relevant examples from the corpus). Also excluded is type [B-2], which is totally beyond the scope of our study, but Fais (2004) attempts to include this type in her centering study. She proposes, for a better characterization of coherence, a new TRANSITION state named “cohesive shift” that considers lexical relatedness in determining CBS for TRANSITION states otherwise categorized as “NULL” (see 3.2.3.5 below for the definition of “NULL”).

### 3.2.3.3 CF ranking

The basic elements of Centering Theory are the discourse entities that appear in each utterance, called FORWARD-LOOKING CENTERS or CFs. Because the notion of salience is crucial to Centering Theory, these entities are ranked in the CF list for each utterance according to language specific ranking principles that reflect the relative salience of the entities.

CF ranking is one of the best-researched parameters of the “parametric” theory. The factors in determining CF ranking have traditionally been grammatical relations. The CF ranking initially proposed by GJW for English is as follows.

\[
(3.8) \quad \text{SUBJECT} > \text{OBJECT} > \text{OTHERS}
\]

A slight modification was made by Brennan, Friedman, and Pollard (1987), who made a further distinction between objects and indirect objects. In addition, some subsequent cross-linguistic studies have augmented the ranking with other language-specific features, while others have proposed alternative potential factors for certain languages: lexical conceptual structures for English (Cote, 1998), thematic roles for Turkish (Turan, 1998), and information structure for German (Strube and Hahn, 1999) among others. While maintaining the grammatical role-based ranking, Kameyama (1985, 1986) proposed and Walker et al. (1990, 1994) agreed that, in addition to the role of grammatical function hierarchy, two special discourse devices in Japanese contribute to the salience of an entity: topics (marked by a topic marking particle, \textit{wa}) and empathy or IDENT in Kameyama’s terminology (normally indicated by certain empathy-loaded verbs). The ranking in (3.9) has since been the dominant ranking used for the centering study of Japanese.

---

7. and her acting uncomfortable stylization.
8. As she gained composure during the second act
9. her technical resourcefulness emerged stronger
10. though she had already revealed a trill almost unprecedented in years of performances of Lucia

We will present similar samples in Japanese in Chapter 4 and some data in Chapter 7.
Both TOPIC and EMPATHY are placed higher than the otherwise highest ranked SUBJECT. They allow some entities in syntactically less salient positions to be elevated to higher-ranked positions (than subjects) and, as a consequence, to continue to be CENTERS in subsequent utterances. In another view, these features can be seen as strategies that native speakers subconsciously employ for the purpose of continuing CENTERS and maintaining coherence, and that learners need to consciously or subconsciously acquire. We will look further into these two devices and discuss our position for the ranking used in this study.

**Topic**

One linguistic typology classifies English as a subject-prominent language and Japanese as a topic-prominent language (Li and Thompson, 1976). In the topic-prominent languages, the grammatical units of topic and comment are basic to the sentence structure. Moreover, topics in Japanese are explicitly marked by a so-called topic marker *wa*. Obana (2000) examines the characteristics of topic-prominence from a language learning perspective. She also discusses the discourse function of topic by summarizing the literature including Kuno (1978). Once a topic is introduced in an initial utterance, this *wa*-marked entity may be readily omitted in subsequent utterances until another *wa*-marked NP is introduced to change the topic. It is pedagogically plausible to direct learners’ attention to this topic chain phenomenon and topic-hood (rather than subject-hood) as a strong indicator of salience in Japanese discourse.

However, this argument is made less persuasive by the fact that in Japanese, topics and subjects often coincide. We examined our corpus (see Chapter 4) for *wa*-marked topic NPs and their grammatical functions in the utterances in which they occur. The result is shown in Table 3.2 below (next page).

---

6 Here, we tentatively limit our definition of topics to *wa*-marked NPs, but our intuition calls for further investigation on other possible topic constructions, such as NP-*to-ieba* ‘speaking of NP.’ This has not been discussed in either theoretical or centering literature.
As is apparent, the vast majority of topicalized NPs are canonically nominative, occupying the subject position of utterances. Topicalized accusative NPs, often defined as a theoretically possible construct, are extremely rare in our corpus, a situation which is also pointed out by Kameyama (1985, page 114, ft.). The cases of topicalized non-obligatory arguments are also very infrequent.

Moreover, Kameyama (1985) limits her definition of topic in the CF ranking to topicalized subjects, objects and possessors, excluding topicalized obliques and adjuncts.\(^8\) In fact, it is questionable whether all topicalized NPs, regardless of their canonical case, are equally salient or are always more salient than subjects. These observations taken together lead us to question the validity of specially ranking TOPIC at the most salient position in the CF.

On a related issue, Walker et al. (1990, 1994) demonstrated that topics sometimes affect the determination of preferred interpretation even when they are not overtly wa-marked (i.e., zero topics) and proposed the Zero Topic Assignment (hereafter ZTA) rule in (3.10).

\[(3.10) \quad \text{When a zero in } U_{i+1} \text{ represents an entity that was the CB (} U_i \text{), and when no other CONTINUE transition is available, that zero may be interpreted as the ZERO TOPIC of } U_{i+1}.\]

This rule allows some ZEROS in syntactically less salient positions to continue to be CENTERS. A typical ZTA example is that a zero object (or a zero genitive) is realized as

---

7 This is a compound particle that indicates role or function; it can be translated as ‘as.’

8 She states that adjuncts “appear to be associated with global coherence rather than local coherence” (page 116) and leaves further discussion for future studies.
Zero (in a subject position) in the immediately following utterance. We also disregard this rule in our analysis.

**Empathy**
The notion of empathy was proposed by Kuno and Kaburaki (1997) and Kuno (1978). Empathy expresses the perspective or position that a speaker takes in describing an event. In Japanese, the speaker’s empathy is encoded by using empathy-loaded verbs. These include verbs of giving and receiving. Empathy locus is defined as the argument position whose referent the speaker identifies with. A list of such verbs and their empathy loci are summarized, following Kuno and Kaburaki, Tsujimura (1996) and Obana (2000), in Table 3.3.

<table>
<thead>
<tr>
<th>Empathy Loci</th>
<th>Verb</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJECT</td>
<td>yaru, ageru</td>
<td>giving</td>
</tr>
<tr>
<td></td>
<td>morau</td>
<td>receiving</td>
</tr>
<tr>
<td>OBJECT</td>
<td>kureru</td>
<td>giving</td>
</tr>
<tr>
<td></td>
<td></td>
<td>receiving</td>
</tr>
</tbody>
</table>

Table 3.3: Empathy loaded verbs and their empathy loci

These verbs can also be used as auxiliary verbs, attaching to the main verbs in complex predicates in quite a productive way.

Yanagimachi (2000) reports from his observation of spoken narrative discourse that native speakers of Japanese effectively use these empathy-loaded verbs to fix their viewpoint. This results in infrequent topic shifts and continuous use of ZEROS. Learners, on the contrary, tend to switch from one topic to another due to the lack of mastery of this viewpoint fixation technique. This causes frequent topic shifts and requires the use of overt anaphoric forms each time.

Tanaka (2001, 2004) also points out in her study of cross-linguistic influence on the acquisition of viewpoint fixation (encoded by passives and empathy-loaded expressions) that English speaking learners of Japanese tend to show a delay in the early-stage development, compared to other speakers (Korean and Chinese), and to show gradual progress.

The reports of Yanagimachi and Tanaka suggest that empathy is a critical pragmatic device that learners of Japanese need to “learn” for natural discourse creation and interpretation. This further implies that empathy-involving discourse may not be perceived to be equally coherent between native and non-native speakers of Japanese, and among learners with different proficiency levels.

In conclusion, we decided to eliminate “empathy,” in addition to “topic,” from
listing in our CF ranking for the purpose of elucidating perceived degrees of coherence by Japanese language learners. Thus, we leave the ranking as simple and syntactic as possible, as formulated in the original centering account, which is given in (3.11).  

(3.11) SUBJECT > OBJECT (S) > OTHERS

We regard other semantic and pragmatic factors, including topicalization and empathy, as potential resources for additional information required in inference processes in interpreting CENTER (dis)continuation.

Complex NPs

Standard accounts of centering do not include provision for the ranking of the entities that make up complex nominal phrases. A typical complex nominal construction in Japanese, the A no B phrase, is of great interest in this study, as discussed earlier in Chapter 2. Thus, the ranking within the construction is an important issue to discuss. Let us first review how complex nominals have so far been treated in the literature, mainly for English.

Walker and Prince (1996) proposed the Complex NP Assumption as a hypothesis about how to handle the multiple discourse entities evoked in complex phrases in English. This assumption states that such entities are ordered on the CF ranking as they appear from left to right within the complex NP. Other researchers (Gordon and Hendrick, 1997; Gordon, Hendrick, Ledoux, and Yang, 1999; Turan, 1998) reject the notion that surface word order alone can characterize salience (or “prominence”). The work of Gordon and colleagues with the processing of name and pronoun references in complex phrases in English sheds light on the effect that embedding has on the prominence of referential expressions. Based on experiments in which subjects’ reading times for short discourses containing possessive structures with both names and pronouns are measured, Gordon and colleagues concluded that the more deeply embedded element, namely, the possessed element, was more prominent.

Tetreault (2001) evaluated the performance of an algorithm based on Walker and Price’s Complex NP Assumption and an algorithm based on Gordon et al.’s claim that the possessed entity was the more salient. He based his evaluation on how accurately the two algorithms could resolve pronominal reference to elements of possessive phrases. His conclusion was that the Complex NP Assumption yielded slightly better results. Poesio and Nissim (2001) also compared these two approaches. Their results showed that the Complex NP Assumption not only led to fewer violations of major principles of Centering Theory, but also predicted subsequent reference to the possessor better than Gordon et al.’s hypothesis. Di Eugenio (1998) uses as a “working

---

9 The role of grammar in the control of inferences was the original motivation of the centering model (Joshi and Kuhn, 1979; Joshi and Weinstein, 1981).
hypothesis” that a possessed entity that is animate precedes a possessor entity (whether animate or not); otherwise (i.e., if the possessed entity is inanimate), the possessor precedes the possessed entity on the CF ranking list. Poesio and Nissim revise this account with an amendment that the possessor is more highly ranked if it is pronominalized.

These studies were conducted on the English phenomenon of the possessive construction, which is only approximately analogous to the Japanese A no B construction, so it is not possible to make very specific arguments by comparing the two. However, we can say that there seems to be no single factor, such as word order or animacy, that determines the ranking. This is also true of the Japanese case, as is suggested by the results reported by Yamura-Takei and Fais (ms.) in which the salience ranking of entities in the A no B phrase was examined on the assumption that an entity of the phrase that provides the antecedent for a subsequent anaphor must be the more salient of the two entities. This examination yielded an interesting result, which is far from being as simple as is suggested by the studies for English mentioned earlier. Defining criteria for characterizing salience in this complex nominal construction seem to stem from both syntactic and semantic factors.

Now then, how do we rank entities within complex NPs? As we have stated several times, the purpose of our centering analysis is not to evaluate a centering algorithm for pronoun resolution; efficiency of parameter setting is not an issue. We rather attempt to examine how syntactic constraints affect coherence establishment. So we simply place B nouns, which are syntactic heads, in a higher position in the ranking.

### 3.2.3.4. Pronominalization: Rule 1

As previously discussed (in 3.2.3.2), we regard ZEROS as equivalent to pronouns in English (and some other languages) that centering is concerned with. Therefore, we directly apply a hypothesis about the relation between centering and pronominalization, which is called Rule 1, to ZEROS. The formulation of Rule 1 defined in GJW is as follows.

\[(3.12) \text{If any element of } \text{CF} (U_n) \text{ is realized by a pronoun in } U_{n+1}, \text{ then the CB} (U_{n+1}) \text{ must be realized by a pronoun also.}\]

### 3.2.3.5 TRANSITION: Rule 2

Adjacent utterance pairs are characterized in terms of TRANSITION types. In GJW, three types of TRANSITION relations are defined according to two criteria: (a) whether or not the CB (U_{i-1}) is maintained in U_i, and (b) whether or not CB (U_i) is also the most
highly ranked entity (CP) of \( U_i \). This can be summarized as in (3.13).

\[
(3.13) \quad \text{CENTER continuation (CON): } CB (U_i) = CB (U_{i-1}), \text{ and } CB (U_i) = CP (U_i) \\
\text{CENTER retaining (RET): } CB (U_i) = CB (U_{i-1}), \text{ but } CB (U_i) \neq CP (U_i) \\
\text{CENTER shifting (SHIFT): } CB (U_i) \neq CB (U_{i-1})
\]

Later, Brenann et al. (1987) introduced a further distinction between two types of SHIFT according to whether or not \( CB (U_i) = CP (U_i) \), and Walker et al. (1990, 1994) named the two distinct states SMOOTH-SHIFT and ROUGH-SHIFT, respectively. A widely-used classification, as a result, is as in the following table.

<table>
<thead>
<tr>
<th>( CB (U_i) = CB (U_{i-1}) )</th>
<th>( CB (U_i) \neq CB (U_{i-1}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTINUE</td>
<td>SMOOTH-SHIFT</td>
</tr>
<tr>
<td>RETAIN</td>
<td>ROUGH-SHIFT</td>
</tr>
</tbody>
</table>

Table 3.4: TRANSITION definitions

Given these definitions, Rule 2 claims that differences in inference cost in discourse interpretation, and thus in coherence, result from using different sequences of TRANSITIONS. Rule 2 is defined, in GJW, as in (3.14).

\[
(3.14) \quad \text{Sequences of continuation (dubbed CON in this thesis) are preferred over sequences of retaining (RET); and sequences of retaining are to be preferred over sequences of shifting (SHIFT).}
\]

Although Rule 2 was originally formulated in terms of sequences of utterances, many applications of this rule to discourse processing algorithms (after the work of Brennan et al., 1987) have restricted the rule to pairs of utterances, as formulated in (3.15).

\[
(3.15) \quad \text{Transition states are ordered. The CONTINUE transition is preferred to the RETAIN transition, which is preferred to the SMOOTH-SHIFT transition, which is preferred to the ROUGH-SHIFT transition.}
\]

These uses of Rule 2 fail to capture the intuition that what matters to coherence are centering TRANSITIONS throughout a segment, not only between pairs of utterances. It is, however, easier to evaluate coherence between a pair of utterances than over a whole segment (Grosz and Sidner, 1998, page 48). A somewhat intermediate approach was taken by Di Eugenio (1998) and Turan (1995), who attend to certain pairs of TRANSITIONS (e.g., CON-CON, RET-CON, SHIFT-CON). Strube and Hahn (1999) took a slightly different position. In their formulation, pairs of TRANSITIONS \( \langle U_i, U_j \rangle \),
<U_j, U_k>> that are “cheap,” i.e., CP (U_j) = CB (U_k) are preferred over those that are “expensive,” i.e., CP (U_j) ≠ CB (U_k).

Rule 2 “reflects our intuition that continuation of the CENTER and the use of retentions when possible to produce smooth transitions to a new CENTER provide a basis for local coherence (GWJ95, page 215).” This implies that the CON-RET-SHIFT sequence is a valid way for CENTER movement, or “topic change” to take place. The rule also predicts that certain sequences produce a higher inference load upon the reader than others. The CON-CON sequence is predicted to require a lower inference cost than, for example, the RET-RET or the SHIFT-SHIFT sequence. The CON-SHIFT sequence is hypothesized to be more costly than the CON-RET sequence. We follow these claims as postulated in the original version of centering, and take the TRANSITION-sequence approach to coherence-driven preferences, rather than the single-TRANSITION approach as proposed by Brenann et al. (1987).

In addition to the canonical TRANSITION types (CON, RET, and SHIFT), corpus studies revealed that natural-occurring discourses contain quite a few utterances without a CB, that is, there is no common entity between U_i-1 and U_i (cf. Passonneau, 1998; Poesio et al., 2002, 2004). Such utterances are labeled “NULL” or elsewhere called “No CB” (e.g., Di Eugenio, 1998).

**Single TRANSITION versus TRANSITION sequence**

Let us first clarify what we mean, in this thesis as well as in the previous studies, by a TRANSITION and a TRANSITION sequence. Look at the next sample discourse, which consists of utterances (U_1) through (U_4), in (3.16).

(3.16)

(U_1) よこはまの おばあさんは  りょうりを
yokohama-no obaasan-wa ryoori-o
Yokohama-GEN grandmother-TOP recipe-ACC

たくさん 知っています。
takusan sitte-i-masu.
many know-POL

‘Grandma in Yokohama knows a lot of recipes.’

(U_2) Ø いいりょうりの本を 持っています。
(Ø-ga) ii ryoori-no hon-o motte-i-masu.

This sequence may follow another stretch of continuation to talk about a newly established center.

Poesio et al. (2002, 2004) distinguished the NULL TRANSITION between utterances neither of which has a CB and the ZERO TRANSITION from an utterance with a CB to one without. We collapse these two TRANSITIONS into the NULL TRANSITION.
‘(She) owns a good recipe book.’

(U₃) ひろこさんの お母さんは ときどき おばあさんに
hiroko-san-no okaasan-wa tokidoki baasan-ni
Hiroko-GEN mother-TOP sometimes grandmother-DAT

電話を かけます。
denwa-o kake-masu.
telephone-ACC ring-POL

‘Hiroko’s mother sometimes telephones (her ‘Hiroko’s’) grandmother.’

(U₄) そして Ø Ø いろいろなりょうりを 開きます。
sosite (Ø-ga) (Ø-ni) iroirona ryooiri-o kiki-masu.
and (Ø-NOM) (Ø-DAT) various recipe-ACC ask-POL

‘And (she ‘mother’) asks (her ‘grandmother’) for various recipes.’

Table 3.5 (next page) schematically represents the relationship between CENTER structure and TRANSITION state of each of these utterances, and between TRANSITION and TRANSITION sequence.
Table 3.5: Schematic view of TRANSITION and TRANSITION sequence

“TRANSITION” characterizes the relation, in terms of CENTER movement, between the two adjacent utterances (e.g., U₂ and U₃); it labels the latter utterance (underlined) (e.g., U₃). “TRANSITION sequence,” on the other hand, characterizes the relation between the two adjacent TRANSITION states (e.g., CON and RET) that involve three successive utterances (e.g., U₁, U₂, U₃). The label (CON-RET) is assigned to the last utterance in the sequence (e.g., U₃ underlined). Thus, when we say the use of ZERO in a certain sequence, it concerns a ZERO in the last utterance in that sequence.

**TRANSITION sequence and inference cost**

There are a total of eleven possible sequence patterns out of combinations of CON, RET, SHIFT, and NULL TRANSITION types.¹² We tentatively divide the eleven types into three groups in accordance with the predictions outlined by GJW’s Rule 2: “low-cost” sequence types and “high-cost” sequence types, placing in between “medium-cost” sequences, as presented in Table 3.6. Although Rule 2 explicitly mentions only three sequence types, CON-CON, RET-RET and SHIFT-SHIFT (indicated in bold in the table), we take into account the claims made by GJW concerning the rule (see 3.2.3.5 above) and list all possible sequence types accordingly.

¹² The NULL-SHIFT sequence is theoretically impossible.
Table 3.6: Inference cost-based classifications of sequence patterns (tentative)

<table>
<thead>
<tr>
<th>“low-cost” sequence types</th>
<th>“medium-cost” sequence types</th>
<th>“high-cost” sequence types</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON-CON, CON-RET, RET-SHIFT, SHIFT-CON</td>
<td>RET-RET, SHIFT-SHIFT, NULL-CON, NULL-RET</td>
<td>CON-SHIFT, RET-CON, SHIFT-RET</td>
</tr>
</tbody>
</table>

This grouping is made because the distinction is not, of course, binary, but rather is graded, as claimed by GJW95: “to the extent a discourse adheres to centering constraints, its coherence will increase and the inference load placed upon the hearer will decrease (page 210).” We then compromise, for the sake of simplicity, with this three-scale hierarchy.

We use this distinction as a starting point to analyze the effect of ZERO use in each sequence type on perceived degree of coherence in a discourse, and hence inference cost required for interpreting the ZERO(s) contained, and ultimately for understanding the discourse. We also attempt not only to estimate the amount of inference that centering concerns, but also to elucidate the types of information resources required for such inference. This will be done by analyzing both textual and contextual environments that enable the use of ZEROs in “high-cost” sequences. In Chapter 4, particularly in 4.4.3, we will make an empirical assessment of the centering-predicted coherence/inference measure, as depicted in the table above, and make adjustments, if necessary, according to the analysis result. The analysis will be accompanied by statistical data and numerous relevant discourse samples from our corpus.

3.3 Summary

In this chapter, we first discussed some fundamental concepts in understanding discourse coherence and overviewed approaches to coherence proposed in the literature. We then introduced Centering Theory, a model that we chose as an explanatory tool for the relationship between ZEROS and coherence/inference in Japanese discourse. We emphasized the original intention of Centering as discussed in GJW, and fully described the “parameter” settings that we adopted for the purpose of characterizing ZERO-involving coherence. We finally proposed a tentative version of inference cost-based classifications of TRANSITION sequence patterns, on which our later corpus analysis is based.