Entity Coherence in Comparable Learner Corpora: Seeking Pedagogical Insights∗

Mitsuko Yamura-Takei[a], Miho Fujiwara[b], and Etsuko Yoshida[c]

[a] Faculty of Law, Hiroshima Shudo University,
1-1-1 Ozukahigashi, Asaminami-ku, Hiroshima 731-3195, JAPAN
takeim@shudo-u.ac.jp

[b] Department of Japanese & Chinese, Willamette University,
900 State St. Salem, Oregon 97301 USA
mfujiwar@willamette.edu

[c] Faculty of Humanities and Social Sciences, Mie University,
1577 Kurima-machiya-cho, Tsu, Mie, 514-8507, JAPAN
tantan@human.mie-u.ac.jp

Abstract. This paper describes an ongoing collaborative project, between Japanese and U.S. universities, that aims to build, analyze and use comparable learner corpora in an attempt to promote discourse-level proficiency in foreign language learning contexts. The focus is placed on discourse coherence created by reference to nominal and clausal entities. The corpus analysis results, within the framework of Centering Theory (Grosz *et al.*, 1995), will be presented, along with some pedagogical insights that teachers can utilize, back in the EFL/JFL classrooms where the data was initially collected.

Keywords: learner corpora, entity coherence, event reference, centering theory

1. Introduction

Discourse competence, one of the four principal components of communicative competence put forward by Canale (1983), is defined as the ability to connect sentences in stretches of discourse and to form a meaningful whole out of a series of utterances. Unlike "hard rules" at morphological and syntactic levels, many discourse phenomena are usually governed by "principles" or "preferences." Therefore, it is hard to provide explicit and systematic instructions. While discourse-level research and practice have been attracting attention (cf. McCarthy, 1991), it is undeniable that a focus on sentence-level form tends to be emphasized in many foreign language teaching contexts. Therefore, it is crucial to develop effective ways for learners to better understand the concept of discourse coherence and to utilize relevant cohesive devices.

Corpus-based research in the past two decades has unquestionably had a considerable impact on various linguistic sub-disciplines. Foreign language pedagogy is not an exception. Corpora come in various shapes. For instance, learner corpora are structured collections of language produced by language learners. Comparable corpora are two (or more) sets of corpora in different languages designed along the same lines. Learner corpora can provide information about how learner production differs from a target model and thus can inform the field of second language acquisition (SLA), for which a comparable corpus of native speaker texts is required.

In an attempt to integrate these two pedagogical challenges and promises, we have designed the collaborative corpus project between Japanese and U.S. universities. In this paper, we will

---

first describe the project and then present our corpus analysis results, with a focus on entity-coherence that is characterized in the centering framework (Grosz *et al.*, 1995). Special attention will be paid to event reference found in the data.

2. Corpus Project

Our ongoing collaborative corpus project, between Japanese and U.S. universities, aims to (1) build, (2) analyze and (3) use comparable learner corpora in English/Japanese as a foreign language (EFL/JFL) teaching contexts. The project outline is presented in Figure 1 below.
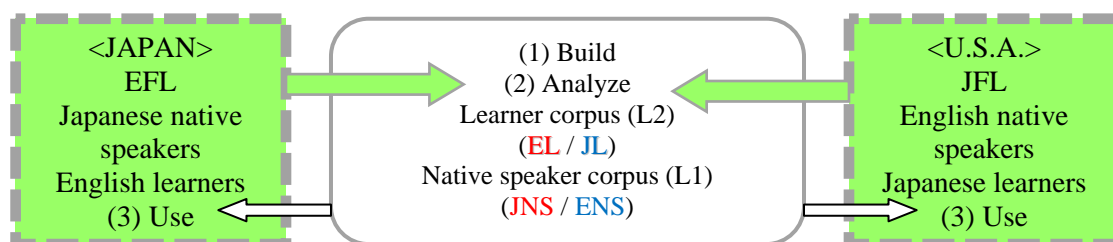

Figure 1: Corpus project outline.

Learner corpora research, when compared against comparable native-speaker corpora, offers important quantitative and qualitative insights to language pedagogy (Mukherjee, 2006). With the same premise, the project focuses on discourse-level features, rather than on well-studied lexico-grammatical patterns. Also, the project is designed along the lines of what Seidlhofer (2002) terms "local learner corpora." In other words, the texts that the project participants have themselves produced will be used as a resource for teachers and learners in their own classrooms. Thus, the compiled corpus is not intended for public distribution.

2.1. Corpus Design

The authors, in EFL and JFL teaching environments in Japan and the U.S., collect student writing in their first and second languages (L1 and L2). In collecting data, a video episode of *Pingu*, a Swiss clay animation, is presented to students to prompt production of a written narrative (i.e., synopsis writing). A 5-minute *Pingu* episode is ideal for this purpose in that it involves a limited number of characters (i.e., discourse entities) whose dialogue is in a make believe "penguin language," making the story easy to follow regardless of the viewers' language background.

Students in Japanese and American universities are instructed to produce a synopsis of an episode of *Pingu* in both L1 and L2. To minimize L1 transfer (i.e., linguistic interference from a native language norm), the subjects are instructed to write first in their L2 and then in L1 and not to translate from one writing to the other.

This provides us with four subsets of corpora: a Japanese native speaker (JNS) corpus, an English learner (EL) corpus, an English native speaker (ENS) corpus, and a Japanese learner (JL) corpus (see Figure 1 above). These four subsets enable several different dimensions of comparability, including comparison of native/non-native speakers (NS/NNS) discourse in English and Japanese, and comparison of L1 and L2 between individual learners.

2.2. Data

The data used for this paper has been collected from 32 Japanese and 27 U.S. university students, each of whom wrote a synopsis of either one or both of two different *Pingu* episodes. Quantitative information for the data is summarized in Table 1.

Table 1: Basic counts of the data.

	ENS	EL	JNS	JL	TOTAL
Number of texts	48	36	34	48	166
Number of sentences	755	492	473	799	2,519
Average number of sentences per text	15.73	13.67	13.91	16.65	15.17

For this study, the corpus needs to be local and small enough to be compiled in a specific environment to address questions specific to a particular group of learners (Seidlhofer, 2002).

3. Key Concepts and Theory

3.1. Entity Coherence

Our major concern is entity coherence, as opposed to relational coherence that is ensured by rhetorical relations (Mann and Thompson, 1987), of a discourse. Let us first present some key concepts and their definitions that we employ for the discussion here.

Entity coherence concerns repeated reference to the same entity in a discourse. By the standard definition, discourse entities can be individuals, objects, sets, events, facts, propositions, etc., and may be evoked or inferred in a discourse model. In many cases, reference is made to concrete individuals or objects that are typically introduced into a discourse in nominal forms, via explicit linguistic mention (e.g., pronouns, definite NPs), as in the example (from our corpus data) below.

(1)   Pingu is playing with wooden blocks. **He** has trouble stacking **the blocks** so **he** becomes frustrated.

In this example, reference is made to an individual entity *Pingu* by a personal pronoun *he*, and to an object entity *wooden blocks* by a definite NP *the blocks*.

Reference can also be made to so-called "abstract entities" (Hegarty, 2003), also known as "higher order entities" (Gundel *et al*, 1999). This type of entity is introduced to a discourse by a clause, a sequence of clauses or even a larger unit such as a whole discourse segment. Event reference is one such instance. An example from Gundel *et al*. (1999) is presented in (2).

(2) There was a snake on my desk. **That** scared me; …

In this particular example, the demonstrative pronoun *that* refers to an "event entity" introduced by a whole previous clause.

Discourse entities can be concrete or abstract, as contrasted in examples (1) and (2). We will call the former type "nominal entities" and the latter "clausal entities" in this paper.

3.2. Centering Framework

One important work that is concerned with entity-oriented coherence is Centering Theory (Grosz *et al*., 1995).   This theory proposes 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 people continuously update their local attentional focus (called CENTER) as they incrementally process a discourse. Different ways of updating CENTER are formulated as the types of TRANSITION from one utterance to the next. The types are called continuation (CON), retaining (RET) and shifting (SHIFT), in the order of preference.   The combination of these three TRANSITION types makes a total of nine TRANSITION sequence patterns: CON-CON, RET--CON, SHIFT-CON, CON--RET, RET-RET, SHIFT-RET, CON-SHIFT, RET-SHIFT, and SHIFT-SHIFT, which characterize CENTER transition environments

(Yamura-Takei, 2005).   A sentence which does not share any discourse entities with an immediately previous one is labeled NULL (or elsewhere called "No CB" condition).

It is important to note how centering defines a set of discourse entities called CENTERs. Kameyama (1986) assumes that CENTERs are "(sets of) individuals, objects, states, actions and events" (p. 200). Grosz *at al.* (1995) also suggests that "events and other entities that are more often directly realized by VPs can also be centers" (p. 209, footnote 6), but leaves this beyond their scope, and so does the majority of succeeding centering work.

For the purpose of applying centering analysis to the data, we will follow this line and consider only nominal entities for CENTERs. We will then examine in this analysis result how clausal entities interact.

4. Corpus Analysis Results

The purpose of the analysis is to explicate learner-specific and language-specific tendencies that could serve as a pedagogical base to help raise students' and teachers' awareness of local referential cohesion back in the EFL/JFL teaching/learning environments where the data was initially collected.

4.1. Centering Analysis

In order to characterize (nominal) entity coherence of the data, let us first look at our centering analysis results.[1] Figures 2 and 3 present the distribution of centering TRANSITION types in English and Japanese data respectively.
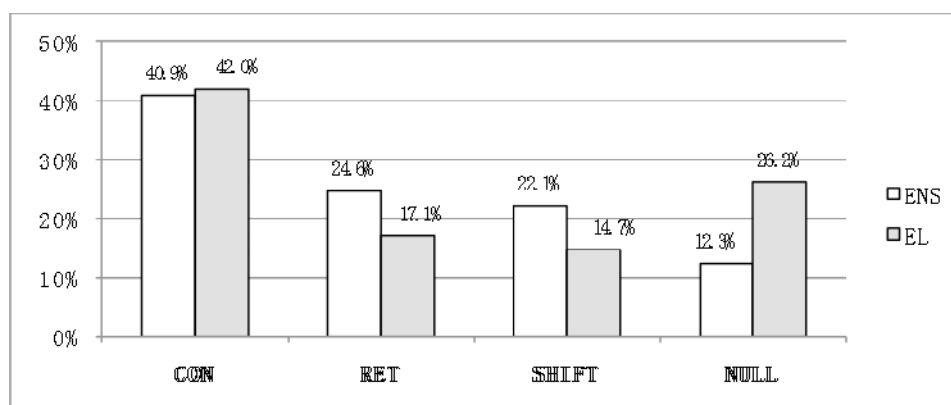

Figure 2: Use of four centering TRANSITION types in the ENS and EL writing samples.
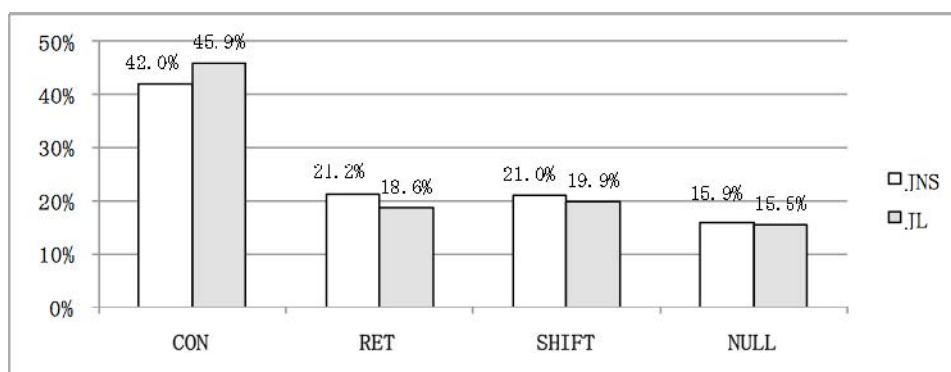

Figure 3: Use of four centering TRANSITION types in the JNS and JL writing samples.

---

[1] Centering annotation was done by a single author.

The majority of utterances in each subset (87.7% for ENS, 73.8% for EL, 84.1% for JNS, and 84.5% for JL) are more or less "entity-coherent" in a centering sense, (i.e., CON, RET, SHIFT). The distribution is roughly analogous to previous corpus analyses: CON is most frequently used (i.e., most preferred) followed by RET and SHIFT. Research shows that NULL-labeled utterances are contained in any natural-occurring data and their proportion is widely varied according to corpus types: 75.7% in Japanese email corpus (Fais, 2004), 37.8% in Japanese written texts (Yamura-Takei, 2005), and 21% in Brown corpus with a variety of written texts (Hurewitz, 1998) among others.[2] The NULL proportion of our data seems to be within a reasonable range. It is interesting to note, however, that English learners' writing is less "entity-coherent" with a relatively high proportion of NULL-labeled utterances (26.2%) when compared to the other three sub-corpora (12.3%, 15.9% and 15.5%). Considering that the four subsets present the same content in the same discourse mode (narrative), this difference is worth noting.

Now let us turn to the TRANSITION sequences. Among the nine TRANSITION sequence patterns, CON-CON, CON-RET, RET-SHIFT, and SHIFT-CON are the four preferred "smooth" CENTER flow patterns, as theoretically formalized in Grosz *et al*. (1995) and statistically verified in Yamura-Takei (2005). In contrast, the other five patterns (CON-SHIFT, RET-CON, RET-RET, SHIFT-RET, SHIFT-SHIFT) are considered to be "rough" and unexpected patterns.   The smooth patterns are grouped in (A) and the rough patterns are grouped in (B) in Table 2.

Table 2: Ratio of TRANSITION sequence pattern in ENS/EL and JNS/JL data.

	type	ENS	EL	JNS	JL
(A)	CON-CON	17.6%	16.8%	21.0%	19.3%
	CON-RET	14.6%	9.3%	12.6%	10.9%
	RET-SHIFT	8.6%	6.6%	6.9%	8.8%
	SHIFT-CON	5.4%	4.9%	6.4%	5.1%
	(A) total	46.2%	37.6%	46.9%	44.2%
(B)	CON-SHIFT	5.6%	5.1%	6.6%	5.7%
	RET-CON	10.9%	6.4%	9.3%	8.6%
	RET-RET	3.3%	0.5%	0.6%	0.9%
	SHIFT-RET	3.3%	2.7%	1.9%	5.5%
	SHIFT-SHIFT	7.9%	3.0%	6.4%	6.6%
	(B) total	31.0%	17.8%	24.8%	27.2%
(C)	NULL-x, x- NULL	22.8%	44.7%	28.3%	28.6%

Comparison of groups (A) and (B) in Table 2 suggests that native speakers of English use more "rough" patterns (31.0%) than English learners do (17.8%) in (B) (z = 5.03, p < .01). However, it is not the case in the Japanese data. One possible explanation of English native speakers' abundant use of "rough" TRANSITION patterns is that they might be using other coherence-creating devices such as connectives and lexical cohesion effectively to maintain coherence of the discourse. It will require more careful examination of the data to verify this possibility.

EL writing, on the other hand, contains a fairly high proportion of NULL-involving sequences (44.7%) as indicated in (C), which causes frequent gaps in entity-coherence of a discourse. This is also the area that needs thorough examination of the data to determine whether or not these gaps are supplemented by other coherence relations.

A seemingly minor but interesting finding is the higher proportion of RET-CON sequence of native speaker samples in English (10.9%), than of learner writing (6.4%) (z = 2.62, p < .01).

---

[2] These figures include the utterances with hard-to-define "inferable centers."

This RET-CON often creates RET-CON-RET-CON sequence in which its CENTER moves back and forth in various syntactic positions while maintaining a reasonable degree of coherence with the same CENTER, as exemplified in a segment below from ENS data (with CENTER in bold).

(3)
  a. [CON] When his mother offers him a potato, **Pingu** throws it back at her. (subject)

  b. [RET] While Mama cleans up, Papa scolds **Pingu** for the way he's acting. (object)

  c. [CON] Angry at his parents, **Pingu** starts rocking back and forth in his chair, until he rocks too far forward and falls over, bumps into the table, and spills all the food onto the floor. (subject)

  d. [RET] **Pingu's** parents are very angry, and spank **him** before starting to clean up the mess. (possessive/object)

  e. [CON] Angrier than ever, **Pingu** walks outside to run away from home. (subject)

A possible explanation for this phenomenon might be found in Table 3 below. The table shows the ratio of syntactic position that CENTERs occupy in each sub-corpus. The data contains CENTERs realized by either full NPs or pronouns (including zero arguments in Japanese).

Table 3: Ratio of CENTER positions in ENS/EL and JNS/JL data.

CENTER position	ENS	EL	JNS	JL
subject	69.3%	74.3%	77.9%	71.1%
possessive	18.6%	4.4%	4.7%	9.1%
object	8.9%	17.2%	15.7%	15.0%
other	3.2%	4.1%	1.6%	4.8%

English native speakers realize more CENTERs in possessor positions (18.6%) than learners do (4.4%). It potentially implies that English learners tend to maintain CENTER in subject and object positions due to the lack of skills of "retaining" CENTER in possessor positions, or due to L1 transfer (J->E). That is, English learners place CENTERs in possessor positions as infrequently in their native language (4.7% in JNS data). Conversely, JL data exhibits a higher percentage of possessive CENTERs than JNS, which could also suggest L1 transfer (E->J).

In contrast, EL data exhibits an abundant use of CENTERs in object positions (17.2%), most of which, however, are realized in the form of full NPs rather than pronominal forms. This explains the infrequent occurrences of accusative pronouns (i.e., *him*) in EL data, as indicated in Table 4 below. The table shows the occurrence ratio per sentence of the three personal pronouns (i.e., *he*, *his*, *him*), which, in most cases, refer to *Pingu*, the major discourse entity.

Table 4: Ratio of three pronouns per sentence in English data.

	ENS	EL
his	0.624	0.137
he	0.404	0.120
him	0.285	0.037

We have attempted to characterize each sub-corpus in terms of its (nominal) entity coherence within the framework of centering. In the next section, we will focus on (clausal) entity coherence established by event reference within and beyond the centering account.

4.2. Event Reference

As is defined in section 3.1, "event reference" is a type of reference made to clausal entities evoked in a discourse. This phenomenon is not very frequent but is cross-linguistic. Event reference found in our data comprises a total of 157 cases (47 cases in ENS, 9 in EL, 53 in JNS, and 48 in JL). In English, events are usually referred to by demonstrative pronouns (*this*, *that*) or demonstrative NPs (e.g., *this behavior*), and less commonly by a personal pronoun (*it*). In case of Japanese, typical referring expressions for event entities are demonstrative pronouns (*kore*, *sore*) and demonstrative NPs (e.g., *sono taido*, 'the attitude').

As is previously discussed in section 3.2, event entities are not counted as CENTERs in our centering analysis; hence reference to such entities do not contribute to coherence that centering account can describe. Our interest is in how these 157 cases interact with the standard centering analysis. Figures 4 and 5 below present the distribution of event reference in each TRANSITION type.

Introspectively speaking, event reference functions to fill (nominal) entity-coherence gaps and appears in NULL-labeled environments. However, it does not seem to be the case; the distribution is noticeably varied among the four types of TRANSITION types. Over 70 % of cases for each subset appear in nominal-entity-cohesive situations, i.e., CON, RET and SHIFT. One RET example from ENS data is illustrated in (4).
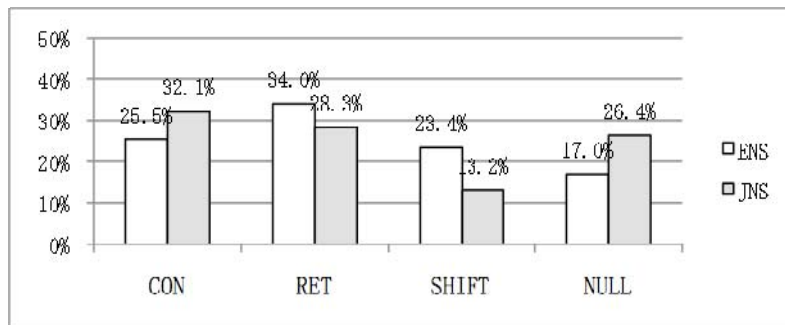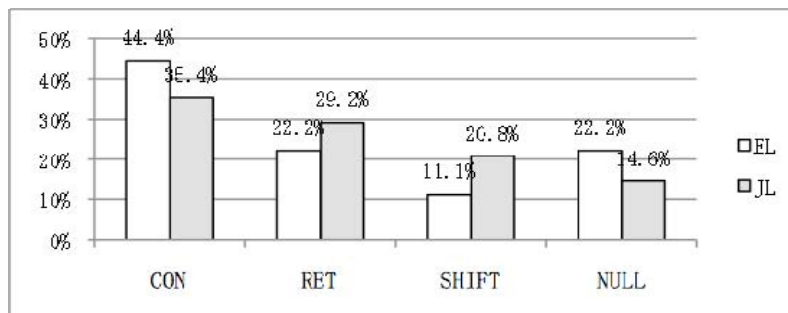

Figure 4: Distribution of event reference in NS data.

Figure 5: Distribution of event reference in NNS (learner) data.

(4) [RET]   **Pingu's** father got very angry, and scolded Pingu.
     [RET]   <u>This</u> made **Pingu** even angrier, and **he** began rocking in his seat.

This example exhibits both nominal entity reference made by a full NP (*Pingu*) and event reference made by a demonstrative pronoun (*this*) that together enhance coherence between the utterances.

On the other hand, there is also the case in which event reference alone creates a cohesive link between adjacent pairs of utterances, as exemplified in (5).

(5) [CON] **Pingu** walked over to Pinga's bed and she woke and started crying.
[SHIFT] Mama hugged **Pinga** and began playing with her.
[NULL] Pingu saw <u>this</u> and became jealous.

In this particular case, the CENTER "roughly" shifts from *Pingu* to *Pinga* and introduces a new entity *Mama* in a salient subject position in the 2nd utterance (SHIFT), which results in an entity-coherence gap when the next utterance exclusively mentions the comeback entity *Pingu* (NULL). This CON-SHIFT-NULL sequence is a very abrupt pattern in a centering sense. The presence of the event reference realized by *this* in the object position, however, successfully establishes the connection to the previous utterance and makes this three-utterance discourse unit reasonably smooth and natural.

Table 5: Ratio of event reference occurrences in NULL

subset	ratio
ENS	8.6%
EL	1.5%
JNS	19.2%
JL	5.6%

Table 5 above shows the ratio of event reference occurrences per all NULL situations in each subset. Native speakers, both English and Japanese, tend to use more event reference in NULL than learners do. Out of 93 NULL cases found in ENS data, eight cases of event reference are found (8.6%), including example (5) above, while EL data exhibits only two instances out of 128 NULL cases (1.5%). Out of 73 NULL cases found in JNS data, 14 utterances exhibit event reference where there is no common nominal entities shared between two adjacent utterances (19.2%), as exemplified in (6) below. On the other hand, JL data includes only seven instances of event reference in 124 NULL cases (5.6%).

(6) ピングーは野菜をママに投げつけてしまいました。
Pingu-wa   yasai-o   mama-ni   nage-tukete-simai-masita.
Pingu-TOP   vegetable-ACC   mom-at   throw-PAST
'Pingu threw vegetables at Mom.'

このことでパパはさらに怒りました。
kono-koto-de papa-wa   sarani   okori-masita.
this-thing-by papa-TOP   more   get-angry-PAST
'Because of this, papa got even angrier.'

As is apparent, native speakers, consciously or subconsciously, fill the gap in nominal entity-coherence, more often than their non-native counterparts, with the use of event reference. The high ratio for JNS data (19.2%) is particularly outstanding.

Now let us further examine the behaviors and characteristics of event reference in our comparable data, apart from the centering framework. Table 6 and 7 show the occurrence rates of event reference per 100 sentences, which are classified by the linguistic forms and syntactic

positions of the referring expressions. The two figures separated by a slash are for a native speaker and a learner data respectively.

Table 6: Types of event reference found in ENS / EL data.

Form	subject	object	others	Form TOTAL
this	**2.1** / 0.0	1.2 / 0.0	0.5 / 0.0	**3.8** / 0.0
this + N	-	0.1 / 0.2	0.8 / 0.0	0.9 / 0.2
that	0.5 / 0.0	-	0.8 / 0.2	1.3 / 0.2
that + N	-	-	0.0 / 0.8	0.0 / 0.8
it	0.1 / 0.6	-	-	0.1 / 0.6
Position TOTAL	**2.8** / 0.6	1.3 / 0.2	2.1 / 1.0	**6.2** / 1.8

Table 7: Types of event reference found in JNS / JL data.

Form	subject	object	others	Form TOTAL
kore	-	0.0 / 0.3	0.0 / 0.1	0.0 / 0.4
kono + N	-	-	0.2 / 0.3	0.2 / 0.3
sore	0.4 / 0.1	**4.4** / 0.5	0.6 / 0.1	**5.5** / 0.8
sono + N	0.0 / 0.1	0.4 / 0.1	4.0 / 4.4	**4.4** / 4.6
sonna + N	-	0.2 / 0.0	0.6 / 0.0	0.8 / 0.0
sou	-	-	0.2 / 0.0	0.2 / 0.0
Position TOTAL	0.4 / 0.3	**5.1** / 0.9	5.7 / 4.9	**11.2** / 6.0

Most noticeable is a higher occurrence rate in NS data for both English (6.2 versus 1.8) and Japanese (11.2 versus 6.0). This may imply that, in terms of the event reference, learners have not reached the level of native speaker norms.

To characterize native speaker norms, let us closely look at the referring expressions used in the NS data. A demonstrative pronoun *this* plays a major role in English (3.8), while *so*-demonstratives (e.g., *sore*, *sono*) are dominant in Japanese (11.0 altogether). As for the syntactic positions in which they appear, there seems to be a clear contrast: subjects for English (2.8) and objects for Japanese (5.1), as represented in examples (7) and (8) below.

(7) **This** frustrates Pingu, and he begins to cry.

(8) ピングーは　　**それを**　　　　見て、やきもちを　　　焼いてしまいました。
　　　Pingu-wa　　**sore-o**　　　　*mi-te,　yakimoti-o*　　*yaite-simai-masita.*
　　　Pingu-NOM　**this-ACC**　　　see　　jealousy-ACC　feel-PAST
　　　'Pingu saw **this**, and became jealous.'

In addition, the demonstrative pronoun *this* in a subject position often co-occurs with emotion verbs, such as *upset*, *worry* and *cheer*. However, no such example is observed in EL data. On the other hand, *so*-demonstratives in Japanese usually appear in object positions of perception verbs, such as *miru* 'see' and *kizuku* 'notice.' This is extremely rare in JL data.

Also as frequent are "others" in both English (2.1) and Japanese (5.7). They include event reference in adjunct positions, typically realized as either temporal or locative phrases. Examples are: *after that*, *at this time* for English, and *sono yoko* 'next to that', *sono koro* 'during that' among others. A compatibly high occurrence rate of "others" type in JL (4.9) is striking. This is probably because Japanese learners acquire these (temporal and locative) phrases at early stages of learning, rather as set phrases.

5. Pedagogical Insights and Future Work

The goal of our comparable corpus project is to utilize pedagogical insights sought and found in our centering-based analysis results back in our own EFL/JFL teaching contexts. The analysis has provided us with some language-specific features and learner-specific tendencies, and some potential difficulties for learners have been uncovered. For example, the examination of CENTER positions reveals one possible L1 transfer case: JL's frequent use and EL's underuse of CENTERs in possessor positions. As for event reference, NS data shows a clear contrast between the languages in terms of syntactic positions in which the event referent appears, i.e., subjects for English and objects for Japanese. Both EL and JL fail to make this native-like performance.

In addition, the centering analysis has provided us with some key issues for future work. The four sub-corpora are reasonably entity-coherent in a centering sense; however, EL contains relatively higher ratio of NULL-labeled utterances than the other subsets. We need to determine how problematic these utterances are from a viewpoint of perceived degree of coherence. The analysis result shows that some NULL-labeled utterances contain event reference that serves to fill the gap in nominal entity coherence. We would like to further investigate what else is happening in NULL environments. Likewise, the fact that ENS use "rough" TRANSITION sequence patterns more frequently than EL is another interesting point to explore. More writing samples and more detailed analysis of the samples might answer these questions.

Some of these findings are currently being utilized, while using NS data as target norm models, in our own foreign language teaching contexts as a pilot study. We will need to evaluate how effective it is and to continue seeking more pedagogical implications.

References

Canale, M. 1983. On some dimensions of language proficiency. In J. Oller (ed.), *Issues in Language Testing Research*, pp. 333-342. Oxford: Oxford University Press.

Fais, L. 2004. Inferable centers, centering transitions, and the notion of coherence. *Computational Linguistics*, 30(2), 119-150.

Grosz, B. J., Joshi, A. K. & Weinstein, S. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 12(3), 175-204.

Gundel, J., Borthen, K. & Fretheim, T. 1999. The role of context in pronominal reference to higher order entities in English and Norwegian. In: P. Bouquet et al. (eds.): *CONTEXT'99, LNAI 1688*, pp. 475–478.

Halliday, M. A. K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.

Hegarty, M. 2003. Semantic types of abstract entities. *Lingua* 113, 891-827.

Hurewitz, F. 1998. A quantitative look at discourse coherence. In M. Walker, A. Joshi, and E. Prince (eds.), *Centering Theory in Discourse*, pp. 273-291. Oxford: Clarendon Press.

McCarthy, M. 1991. *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: the state of the art - and beyond. In S. Braun, K. Kohn & J. Mukherjee (eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, pp. 5-24. Frankfurt am Main: Peter Lang.

Seidlhofer, B. 2002. Pedagogy and local learner corpora: Working with learning-driven data. In S. Granger, J. Hung & S. Petch-Tyson (eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, pp. 213-234. Amsterdam: Johns Benjamins.

Yamura-Takei, M. 2005. *Theoretical, Technological and Pedagogical Approaches to Zero Arguments in Japanese Discourse: Making the Invisible Visible*. Unpublished doctoral thesis. Graduate School of Information Sciences, Hiroshima City University.