

A Corpus-based Study on the Use of Descriptive Language by Japanese EFL Learners in Spoken Picture Description Tasks

RISA TERADA

1. Introduction

The proficiency expectations regarding English education in Japan has been evolving rapidly in recent years making the target English-skill levels necessary for Japanese students more explicit than before. For example, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) recommends that Japanese students should get Grades Pre-2 to 2 in EIKEN before they graduate from high school. It is desirable for learners to have production skills that are more logical, coherent, and cohesive speeches and compositions. This present study investigates the nature of language features necessary for logical, coherent, and cohesive utterances by comparing native speakers with L2 learners and comparing learners with different proficiency levels. This study aims to reveal language features and developmental processes of spoken monologues using picture description tasks produced by Japanese learners of English. The author hopes that the results of this research will contribute to set clearer goals that will offer guidance to improve Japanese learners' communicative skills more efficiently in the future.

This study consists of six chapters with Chapter 2 reviewing previous major studies in relevant fields and Chapter 3 explaining the corpus used in the study focusing especially on picture description tasks. This chapter also covers the research method and procedures in detail. Chapter 4 shows the results of the data analysis, and Chapter 5 discusses the implications of the study results. Finally, Chapter 6

summarizes the major findings and suggests some limitations, and issues to be investigated for future studies.

2. Review of related literature

2.1. Previous studies focused on cohesion and coherence

Many researchers have studied learner language in terms of discourse organization and this section introduces previous related studies and summarizes learner features of discourse organization. Khalil (1989) evaluated English essays written by Arab college students from aspects of cohesion and coherence. He counted the number of cohesive elements included in every t-unit and classified them into five categories: references, substitutions, ellipses, conjunctions, and lexical cohesion. Table 1 indicates that the college students used repetitions most frequently in their essays. Repetitions and collocations are the components of lexical cohesion. In addition, references and conjunctions were also used frequently; however, the proportions of substitutions and ellipses were quite low. In coherence evaluations, the composition which elaborated information with specific examples got a high score. The correlation between the number of cohesive features and coherence evaluations indicated that the correlation was very low. Based on Khalil's research, compositions which include many cohesive features are not necessarily coherent.

Miyasako (2000), Sawaya and Suzuki (2016) had similar results to

Table 1. Types of Cohesive Ties (Khalil,1989)

Type of tie	Total	Percentage
Reiteration	117	61.9
Conjunction	34	18.0
Reference	29	15.3
Collocation	7	3.7
Substitution	2	1.1
Ellipsis	0	0.0
Total	189	

Khalil (1989) with Japanese participants. In Miyasako (2000), a class was divided into two groups based on their proficiency levels and students wrote essays in a regular class and were evaluated holistically. Then, cohesive features were counted and classified into five categories. The results showed significant differences in frequencies between the two proficiency groups. His students frequently used personal pronouns (references) and same word repetitions (lexical cohesion) as the cohesive features. Determining factors of holistic composition scores were also investigated using a multiple regression analysis. Only in the low proficiency group, the number of cohesive features was indicated as one of the determining factors of composition scores. From these results, the number of cohesive features does not necessarily relate to holistic evaluations.

Sawaya and Suzuki (2016) had their students write compositions after instructional focus on logical coherence and cohesion. As with the prior mentioned studies, their compositions were rated holistically and classified in terms of cohesive features. Their results revealed that their students frequently used personal pronouns and definite articles of references, conjunctions, and same word repetitions of lexical cohesion. As the correlation between the number of cohesive features and holistic evaluations, only personal pronouns indicated a positive correlation. They stated that this positive correlation related to the topic of compositions rather than the frequent use of personal pronouns.

2.2. Learner features of discourse organization focused on metadiscourse markers

Kobayashi (2009) conducted discourse analysis comparing native speakers of English and Japanese learners of English. This study used the Japanese Component of International Corpus of Learner English (the ICLE-JP) as the data of Japanese learners. The native speakers of English were extracted from the Louvain Corpus of Native English Essays (the LOCNESS). He used the classification of metadiscourse markers (Hyland, 2005) because metadiscourse markers help receivers to interpret and evaluate information appropriately. Hyland's (2005)

list is composed of ten categories with their functions and examples of each category are summarized in Table 2. He conducted a discriminant analysis and raised four metadiscourse features that contribute to the discrimination between native speakers and Japanese learners. Japanese learners overused frame markers, boosters, and self-mentions, while they underused hedges when compared to native speakers. Frame markers have a role in indicating discourse order, stages, and purposes. Japanese learners especially overused sequential expressions like: last, lastly, next, second, then, and third. Boosters can emphasize

Table 2. The categories of Metadiscourse Markers (Kobayashi, 2009)

Category	Function	Examples
Interactive resources	Help to guide reader through the text	
Transitions (TRA)	Express semantic relation between main clauses	in addition/ but/ thus/ and
Frame markers (FRM)	Refer to discourse acts, sequences, or text stages	finally/ to conclude/ my purpose here is to
Endophoric markers (END)	Refer to information in other parts of the text	notes above/ see Fig/ in section 2
Evidentials (EVI)	Refer to source of information from other texts	according to X/ (Y, 1990)/ Z states
Code glosses (COD)	Help readers grasp functions of ideational material	namely/ e.g./ such as/ in other words
Interactional resources	Involve the reader in the argument	
Hedges (HED)	Without writer's full commitment to proposition	might/perhaps/ possible/ about
Boosters (BOO)	Emphasize force or writer's certainty in proposition	in fact/ definitely/ it is clear that
Attitude markers (ATM)	Express writer's attitude to proposition	unfortunately/ I agree/ surprisingly
Engagement markers (ENG)	Explicitly refer to or build relationship with reader	consider/ note that/ you can see that
Self-mentions (SEM)	Explicit reference to author(s)	I/ we/ my/ our

speakers or writers' certainty and the expression *I think* was heavily used in this category. Self-mentions indicate explicit references to authors within their discourse. The Japanese learners of English heavily used the first personal pronoun, I and we.

2.3. Summary of learner features of discourse organization

This section summarizes learner features of discourse organization based on the results of previous studies. Learner features are summarized in two ways in this study. The first way is that learners use logical expressions like anaphoric and endophoric references, and conjunctions frequently in their discourse. The second way is that learners seem to have difficulties expressing their opinions and feelings efficiently by using hedges, boosters, self-mentions, and the like.

Logical expressions are really important factors to make discourse more transparent and their importance has been emphasized repeatedly in language teaching. However, overuse does not lead to coherent discourse. In the field of cohesion and coherence, many studies concluded that there is no, or weak, correlation between the number of cohesive features and coherence or holistic scores of language production. Cohesive features are generally considered mere factors contributing to coherence (Khalil, 1989; Miyasako, 2000; Crossley & McNamara, 2010; Sawaya & Suzuki, 2016). In other words, the increase in the number of cohesive features used in discourse does not necessarily lead to better discourse organization, which is also true for expressions related to speakers or writers' opinions, attitudes, and feelings. Kobayashi (2009) reported heavy overuse of the first personal pronouns by Japanese learners of English. As he stated, self-mentions can display its rhetorical effects by using them unexpectedly within an objective context. Nevertheless, in the case of Japanese learners, rhetorical effects of self-mentions are lost because subjective expressions account for most of the discourse. Japanese learners also extensively used the phrase, *I think* and it was interfered because of Japanese language influence and was not used appropriately. The overuse of one element repeatedly makes receivers feel uncomfortable.

Kobayashi (2009, 2010) mentioned that one of the causes of this tendency is from classroom instructions about paragraph writing. Examples of logical and sequence expressions are often introduced to instruct writing organizations in Japan (Nakagawa, 2016; Mita & Shimoda, 2021). These sequence expressions are helpful for learners to some extent however, focusing on only the surface features and rhetorical techniques causes learners to overuse one element and narrow their expressive variations. Examining only the quantitative data about surface features does not seem to offer any interesting suggestions for future instructions.

Researchers have stated that elaborateness of information and content also greatly influences transparent discourse (Khalil, 1989; Sawaya & Suzuki, 2016). Therefore, this study focuses on the more concrete aspects of discourse to obtain more beneficial findings. This study attempts to understand learner features of discourse, focusing on specific language expressions and what they intended to express in discourse.

2.4. Research question

Many researchers have investigated learner language in terms of discourse organization and previous studies have shown that there are differences of discourse organization between Japanese learners of English and native speakers of English. When instruction only focuses on surface features of discourse, it makes learners overuse these features, leading to non-target-like discourse structures. It is also important that learners need to decide what content should be included in their discourse.

With these perspectives in mind, the author explores some features of the spoken monologue by Japanese learners of English, with a special focus on the information selection and organization in picture description tasks. This study works on one research question “What information do speakers include as they describe the picture?” In order to reveal the content choices by speakers, the range of this research is specific and narrow, however, this research can provide

useful information. The following chapters, explain the methodologies used, the results of the research, and a detailed discussion of the results.

3. Method

3.1. The corpus used in this study

Learners' spoken data was extracted from the National Institute of Information and Communication Technology Japanese Learner English Corpus (NICT JLE Corpus, henceforth). The NICT JLE Corpus is one of the largest spoken learner corpora. The corpus includes 1,281 files by Japanese learners and 167 of them are tagged for grammatical and lexical errors. It also provides 20 native speakers' speech performance as a sub-corpus. This corpus was constructed based on the results of the speaking test called the Standard Speaking Test (SST, henceforth), which was designed and developed, following the ACTFL Oral Proficiency Interview (OPI). NICT worked with ALC PRESS to transcribe all the 1,200 audio files and annotated each transcript for spoken characteristics. Unfortunately, it does not include audio data (Izumi et al, 2004).

It takes each examinee approximately 15 minutes to have an interview test in SST. The SST is comprised of five stages: i) warm-up conversation, ii) a picture description task, iii) a role play, iv) a storytelling task and v) wind-down conversation. Each stage is divided into two parts: a task part and a follow-up part. In the task part, interviewees are asked to do specified tasks and in the follow-up part, interviewers have a question-and-answer sessions about the topics covered in the task with interviewees. The spoken data were evaluated by trained raters and scored based on their performance of all stages. Test scores are then converted to the SST level 1 (the least proficient) to level 9 (the most proficient).

3.2. Main focus of the present study

This research mainly used the spoken data of the second stage of the SST, the picture description task. Recall that examinees were

required to describe the situation of a given picture so that the examiners could imagine the content of the picture. The SST prepared seven versions of pictures: classroom, electric shop, map, neighborhood, restaurant, room, and skiing. The picture given to each examinee was randomly chosen among these seven pictures.

In order to analyze the descriptions from the detailed content, this study selected data from only one of the pictures because the content created from the seven pictures were quite different from each other and as a result, hard to control in terms of vocabulary. I selected the picture of a restaurant scene for this study and in this picture the interviewee describes a scene from a luxurious restaurant where different groups of people are having dinner. In order to communicate the atmosphere of the restaurant, the interviewee needs to describe the furniture, equipment, or decorations seen in the picture. They also need to mention about the people's behavior in the restaurant. The selection of what information to describe about the scene is really important in order for others to understand clearly what is the situation in the picture.

3.3. File selection

The section of the picture description task was extracted from each interview file before analyzing and Table 3 shows the number of files used in the analysis.

Table 3. The Number of Files Used in The Content Analysis

SST	Native	9	8	7	6	5	4	3	2
File	5	3	11	15	15	45	45	30	15

Depending on the SST level, the total number of files available to each SST level was quite different. For the lower-intermediate level, the SST levels 3 to 6, include the majority of learners while the beginner and the intermediate levels had a smaller number of files. This present analysis included all the files when the picture description task was about the picture of a restaurant, which was for the SST levels 2, 7,

8, and 9, as well as for native speakers. For the file extraction of the SST levels 3 to 6, 15 to 45 samples from each level were randomly selected. Regarding the SST level 1, it included only one file that matches the selection criteria for this study. Therefore, the SST level 1 was excluded in this analysis.

3.4. Content analysis

Content analysis of descriptive data was conducted to answer the research question: “What information do speakers include as they describe the picture?” To this end, it is essential to develop annotation schemes for coding picture information. In developing annotation schemes, since the NICT JLE Corpus does not provide actual pictures used in the SST, it was necessary to guess what was in the picture by looking at the content from the examinees’ answers. The coding category was made based on actual descriptive data mainly of native speakers and the more proficient learners.

In this study, two kinds of coding schemes were developed with the first being focused on informational content in the picture and it has two main categories: PLACE and PERSON. The PLACE category refers to all the information in their responses other than people in the picture, while the PERSON category refers to the information about people in the picture. The second coding scheme focused on utterance functions: DESCRIPTION and INTERPRETATION. The functional classification coded as DESCRIPTION is related to the extent of verifiability and the information which was objectively identified from in the picture. On the other hand, the information adding their own interpretations based on the picture was coded as INTERPRETATION. Below are a few examples of the coding used.

(1) PLACE:

- a. And ur urr there’s a number of wine glasses on the table.
(file11-native)
- b. And the time is seven o’clock. (file00654-SST level 7)

(2) PERSON:

- a. Mmm a man and a woman are in a restaurant. (file01211-SST level 9)
- b. Ur sommelier's in a tuxedo. (file01243-SST level 9)
- (3) DESCRIPTION:
 - a. And further in the background, there is some trees. (file11-Native)
 - b. And er one woman is playing the piano. (file00654-SST level 7)
- (4) INTERPRETATION:
 - a. It looks like a French restaurant. (file01236-SST level 8)
 - b. And he said, "It's good". (file00059-SST level 6)

The coding unit was a sentence and identification of sentences was made based on the transcriptions of the NICT JLE Corpus, where each sentence was delimited by their <s> tags. Each sentence was coded from two aspects, whether the sentence referred to PLACE or PERSON and whether the sentence belongs to DESCRIPTION or INTERPRETATION. There were some cases where more than two types of content were included within a single sentence. Also, in some sentences connected a clause of DESCRIPTION with a clause of INTERPRETATION and in these cases, the sentence was coded separately. Complete coding examples are presented here. In (5c), the former half of the sentence, "So um he has his own drink", was judged as the DESCRIPTION while the latter half, "which is his usual but I assume he's gonna probably try the new one in a minute" was judged as the INTERPETATION. The coding process was done manually by the author.

- (5) Coding examples:
 - a. script: And er one woman is playing the piano. (file00654-SST level 7)
annotation: PERSON-DESCRIPTION
 - b. script: It looks like a French restaurant. (file01236-SST level 8)
annotation: PLACE-INTERPRETATION
 - c. script: So um he has his own drink, which is his usual but I

assume he's gonna probably try the new one in a minute.
(file01234-SST level 8)

annotation: PERSON-DESCRIPTION

annotation: PERSON-INTERPRETATION

3.5. Statistics used in this study: HCFA (Hierarchical Configurational Frequency Analysis)

The frequency data of the content analysis was further analyzed by a hierarchical configurational frequency analysis (HCFA) to determine what information is preferred or not preferred by the different proficient speakers as they describe the same picture. HCFA is an extended version of a chi-square test. HCFA can test the relationships of a multi-dimensional table while the chi-square test is usually applied to a two-dimensional table. This test can examine whether the observed frequencies of variable level combinations are significantly different from the expected frequencies that are expected by chance. The variable level combinations are called *configurations* and “If a configuration is more frequent than expected, it is referred to as a *type*; if it is less frequent than expected, it is referred to as an *antitype*” (Gries, 2009: 244). In this study, configurations were based on three variables, PLACE/PERSON, DESCRIPTION/INTERPRETATION and SST levels.

HCFA processing was done with R based on the R script and a manual introduced in Gries (2009). In processing, the author is aware that a chi-square test should not be performed on proportional data, however this study normalized the mean frequency data of one file against 30 files since the dataset for native speakers were too small. In addition, learner files were grouped together based on their SST levels and corresponding CEFR-J levels according to Tono (2013) in order to highlight features of each level more explicitly. SST levels were aggregated into four classes: native, B, A2 and A1. A breakdown was the following: native (native), B (SST levels from 6 to 9), A2 (SST levels 4 and 5), A1 (SST levels 2 and 3).

4. Results

4.1. Results of the content analysis

Table 4 presents the mean frequencies of the two main content categories: PLACE/PERSON. Mean frequencies were applied in this table because the sample size was quite different among the levels. It becomes easy to compare the values using mean frequencies. This table shows that the mean frequencies increased as speakers' proficiency levels improved with native speakers producing the highest value. Looking at each main category, in general, speakers referred to the PERSON category more than the PLACE. The proportional use of each category within a description is different when comparing learners and native speakers. In the learner group the proportion of the PERSON is almost twice as large as the PLACE. However, the native speakers referred to the PLACE more frequently than learners, and used PERSON and PLACE in almost equal amounts.

Table 4. The Mean Frequencies of Each Content

SST	PLACE	PERSON	ALL
Native	7	7.2	14.2
9	5.3	7.7	13
8	3.6	9.1	12.7
7	3.7	6.4	10.1
6	3.3	6.4	9.7
5	3.5	7	10.5
4	2.4	6.3	8.7
3	2	4.3	6.3
2	1.5	2.7	4.2
ALL	32.3	57.1	89.4

Table 5 summarizes the average frequencies of two main functional categories: DESCRIPTION and INTERPRETATION. From this table, the mean frequencies of the DESCRIPTION are higher than the INTERPRETATION and this is not surprising that speakers included verifiable information more often than their own interpretations. They were asked to describe the given picture during the SST.

As an interesting point, the intermediate level learners added their own interpretations more frequently than the other groups.

Table 5. The Mean Frequencies of Each Utterance Function

SST	DESCRIPTION	INTERPRETATION	ALL
native	11.2	3	14.2
9	10	3	13
8	8	4.7	12.7
7	6.7	3.5	10.2
6	5.7	4	9.7
5	6.5	4	10.5
4	6	2.8	8.8
3	4.8	1.5	6.3
2	3.2	1	4.2
ALL	62.1	27.5	89.6

4.2. Results of HCFA

In this section, the results of HCFA are presented and explanation of how to interpret tables precede the results. The “content” means the variable PLACE/PERSON, while the “function” refers to the variable DESCRIPTION/INTERPRETATION, and the “sst” means the SST levels, respectively. The column named “Freq” shows the observed frequencies of each variable or variable interaction, whereas the column named “Exp” shows the expected frequencies. “Cont. chisq” stands for a contribution to chi-square since the HCFA provides the results of a chi-square test for each subtable. The chi-square value of a whole table is calculated by adding up differences between observed frequencies and expected frequencies in all cells. From the contribution to chi-square, the breakdown of the chi-square value in a whole table can be represented. The column named “Obs-exp” indicates types or antitypes for each cell. If an observed frequency of a cell is larger than an expected frequency, the cell is regarded as a type and is expressed by using >. In contrast, if an observed frequency is smaller than an expected frequency, the cell is regarded as an antitype

and is expressed by using $<$. In this study, type cells were marked in dark gray and antitype cells were marked in light gray. The “P.adj. Holm” is the adjusted p-value, because in processing HCFA with R, we can choose two adjusted p-values for multiple post hoc tests: the Bonferroni correction, and the Holm correction. Gries (2009) recommended using the Holm correction because it guarantees that an overall probability of error does not exceed 0.05. These corrections are quite important so that the research does not reject null hypotheses by mistake. The “Dec” indicates the degree of significance of each adjusted p-value and the degree of significance is expressed at three stages with asterisks. If a cell is highly significant, the column provides three asterisks (***)). However, if the significance of a cell is low, the column provides one asterisk (*). Finally, in the case where a cell is not significant, the column provides “ns”. The rightmost column named “Q” provides the size of the effect and the larger the value of Q, the stronger the configuration contributes to the overall results.

Table 6 indicates the results of the HCFA for the interaction between PLACE/PERSON x SST levels and there are three significant configurations in this table. The configuration of PLACE x native is reported as a type with high significance. The configurations of PERSON x native and PLACE x A2 are considered antitypes. All of the other configurations are not considered as significant. This table is

Table 6. The results of HCFA for the (PLACE/PERSON x SST levels)

content	sst	Freq	Exp	Cont.chisq	Obs-exp	P.adj.Holm	Dec	Q
person	native	324	392.6598	12.0057	<	0.0002506	***	0.049
place	native	315	246.3402	19.1368	>	2.49E-05	***	0.044
person	A2	297	265.4602	3.7473	>	0.1037473	ns	0.02
place	A2	135	166.5398	5.9731	<	0.0288289	*	0.019
person	B	319	298.0282	1.4758	>	0.0980049	ms	0.014
place	B	166	186.9718	2.3523	<	0.1652333	ns	0.013
person	A1	171	154.8518	1.684	>	0.1911403	ns	0.01
place	A1	81	97.1482	2.6842	<	0.1936447	ns	0.009

in agreement with the data in Table 4. Native speakers doing the picture description tended to include PLACE elements more frequently than did the learners. Moreover, they did not describe PERSON elements more frequently than did the learners.

Table 7 summarizes the results of the HCFA for the interaction between DESCRIPTION/INTERPRETATION x SST levels. The configurations of DESCRIPTION x native, INTERPRETATION x B, and INTERPRETATION x A2 are regarded as types with low significance. On the contrary, the configuration of INTERPRETATION x native is regarded as an antitype with high significance. This table can explain the findings presented in Table 5 because native speakers and beginners were more likely to describe verifiable information, while the intermediate learners were more likely to include their own interpretations in picture description.

Table 7. The results of HCFA for the (DESCRIPTION/INTERPETATION x SST levels)

function.	sst	Freq	Exp	Cont.chisq	Obs-exp	P.adj. Holm	Dec	Q
description	native	504	453.0962	5.7189	>	0.0203187	*	0.038
interpretation	native	135	185.9038	13.9384	<	0.0001883	***	0.031
description	B	310	343.8993	3.3416	<	0.0865532	ms	0.023
interpretation	B	175	141.1007	8.1443	>	0.0155996	*	0.02
description	A2	279	306.3186	2.4364	<	0.1352168	ns	0.018
interpretation	A2	153	125.6814	5.9381	>	0.0390396	*	0.016
description	A1	189	178.6858	0.5954	>	0.2183819	ns	0.006
interpretation	A1	63	73.3142	1.4511	<	0.238826	ns	0.006

5. Discussion

5.1. Summary of major findings

The research question is “What information do speakers include as they describe the picture?” With this question in mind, content analysis was carried out using picture description data in the NICT JLE Corpus. The data was annotated according to the two coding schemes:

one focusing on picture content, and the other examining utterance functions. The annotation schemes were developed in a bottom-up manner by considering test examinees' actual descriptive data. The coding scheme focusing on the picture content was divided into the PLACE and the PERSON, while the other coding scheme focusing on the utterance functions was divided into the DESCRIPTION and the INTERPRETATION. Each sentence was classified manually by the author and after coding the data, the frequencies were counted. In addition, the author conducted hierarchical configural frequency analysis (HCFA) to test the contribution of certain configurations to the overall frequency changes across the speakers' proficiency.

The results show the findings are related to the general features and the features specific to each group. First, the amount of information generated increased with speakers' proficiency levels. Second, speakers preferred describing the information related to the people in the image rather than the places, and they also tended to include verifiable information more than supplying their own interpretations.

A comparison between learners and native speakers revealed that the proportion of information on the PLACE and the PERSON was different. While native speakers described both categories almost equally, learners described the PERSON category much more often. In addition, native speakers and novice-beginners included more verifiable information, while low-intermediate and intermediate learners preferred adding their own interpretations in their descriptions.

5.2. Implications for differences between native speakers and Japanese learners

The major findings of the present study revealed that there are some interesting differences in terms of the information that is included in the picture description. One of the differences is related to the distributions of the PLACE and the PERSON categories. Native speakers described a wide range of information on both the PLACE and the PERSON and in particular, they described the information about the place very differently from learners. For example, in (6a), this native

speaker described the distances between each table and the appearance of the entrance of the restaurant in a very specific way in order to describe the atmosphere of the restaurant. In another example(6b), a different native speaker also mentioned the rope in the entrance perhaps showing the restaurant was exclusive. In addition to these specifics, native speakers frequently mentioned the people in the picture who were not located at the center of the picture. Overall, native speakers' descriptions covered more information, and each description seemed to serve a particular discourse function.

(6) Examples of the PLACE by native speakers:

- a. And there are only three table in this picture, and they look like they're really far apart. So maybe it helps people to have their own space and not have to listen to other people's conversation. There's also a velvet rope that er seems to be blocking one of the entrances. And it almost looks like the entrance there's there's no door, like it goes straight outside. So maybe this is a very warm place to to have a restaurant. (file1-native)
- b. Ur there's a doorway with er some kind of rope strung across it. (file11-native)

On the other hand, the learners paid more attention to the PERSON category only. They focused on detailed information, and in particular, about the people at the center of the picture. Examples describing the clothes of the guests are shown in (7). While native speakers use phrases just like *dressed up* and *formal dresses and suits*, the learners even mentioned the colors and accessories.

(7) Examples of the guest-clothes:

- a. Well it looks like a very fancy restaurant where the waiter is dressed up in a tuxedo and he's serving wine to the customers who are also very dressed up in er formal dresses and suits. (file1-native)

- b. And there are some other nicely dressed couples in the background. (file14-native)
- c. And man is wearing yellow suit. And the woman is wearing a pink dress with pearl necklace and earrings. (file01281-SST level 7)

When learners described the information about the place, they mainly referred to concrete furniture: *the clock*, *the table*, and *the piano*, among others. The restaurant was also frequently explained and most explanations were related to a type of restaurant, and seemed to be related to topic introduction. (8) shows examples which described the restaurant.

(8) Examples of the restaurant:

- a. It looks like a French restaurant. (file01236-SST level 8)
- b. Eh I think eh Western eh European type. (file00116-SST level 5)

As a possible reason for this difference, the way the space is recognized may be different between learners and native speakers. Native speakers seem to regard the information about the place just as important as the information about the person. Abstract spatial aspects like the entrance, distances of tables were described as well as specific pieces of furniture and this was a unique feature for native speakers. However, learners constructed the picture description mainly focusing on the information about the people. Therefore, they seem to consider the information about the place as merely the background and did not emphasize them. The feature that learners focused on the people more often than the places would be related to some findings in Izumi (2013). She stated that learners rarely used inanimate subject sentences, while native speakers often used them during the picture description task. In her discussion, Izumi postulated that the different restrictions on both English and Japanese verbs may have an influence on this tendency. Differences of language restrictions might also lead to the different

recognition of space.

Another interesting difference is the degree of speakers' interpretations within the descriptions. In this task, interviewees were asked to describe the content of the given picture. It is natural that more verifiable information in the picture was expressed compared to speakers' own interpretations. However, the findings revealed that the lower-intermediate and intermediate learners included their own interpretations more often than the native speakers. Rich interpretations were added especially to the focused area of the picture. For example, many sentences that were guessing the relationships between guests were found from these groups of learners. On the contrary, these explanations were rare in native speakers' descriptions. From examples in (9), we can see that while native speakers referred to the relationship of guests simply, learners tried to describe them in detail.

(9) Examples of the relationship of the guests:

- a. Erm in this picture, it looks like there's a man and a woman out on a date. (file14-native)
- b. So in the night, the couple, maybe they are not so familiar to each other. Because that if they are they already had they already develop the strong relationship each other, that they don't maybe they don't need to be in this kind of formal expensive restaurant. So I think that they are er very in a beginning of their relationship. (file00980-SST level 8)

It may be that the way speakers thought in their mind may be different between learners and native speakers. The Standard Speaking Test, which the NICT JLE Corpus was based on, does not make available the evaluative criteria of picture description tasks. Thus, each test candidate might work on the picture description task in their own way. Native speakers seemed to transmit the information in the picture description objectively and generally. Therefore, their descriptions for each piece of information were simple and did not include many personal interpretations. On the other hand, learners seemed to

try to express the information in the picture in more creative ways and use their imaginations more freely. In (10), one can see some examples which describe the picture by making original stories, which are unique to a group of Japanese learners of English.

(10) Examples of the original story:

- a. This woman is er Momoko. And this man is Kentaro. And they are a umm they're good friends and decided to er come to have dinner together. And er they often date, but this time, they decided to er come to a good restaurant because er Kentaro tried to er propose her. (file00042-SST level 6)
- b. Um Mr. and Mrs. Yamamoto and er went to the restaurant to have the to have dinner. And uh they when they were engagement, they used to go to the restaurant such kind of restaurant er very often, but now er they have the children, so she ah they can hardly go to the such kind of fine restaurant. And er that night, they could find the baby sitter and er they could come to the er to the restaurant to have dinner. (file00059-SST level 6)

5.3. Implications for differences between different proficiency learners

What are the differences between learners at the different proficiency levels? Possible reasons for distinguishing their scores are discussed and relevant factors may be different for novice-beginners and lower-intermediate learners.

First, novice-beginners are likely to miss the basic information in their descriptions. For instance, information about the restaurant, the clock, and the guests are the most characteristic information in the picture. If speakers can mention at least these items, then the listeners can have an image of the picture more easily. However, beginners cannot even explain the most basic information very well. It may be that the beginners lack the vocabulary to express even the most basic elements of the picture. Although some of them tried to mention some of

these items, the information was too limited for the listeners to understand the picture appropriately. From (11a), the learner probably tried to describe the guest and the waiter, however, this description lacked other specific information. Moreover, they might also lack even the basic language skills needed to complete the task. This tendency matches with results obtained from Izumi (2013). Her analysis also revealed that beginners have difficulties making accurate sentences and ordering them cohesively. (11b) shows the characteristic information about the picture, but each sentence included many grammatical errors and fillers. The sentences are also not connected to each other very well. These aspects may make their descriptions less intelligible to listeners. Therefore, they need to improve vocabulary, grammar, and fluency to express the basic information in a structured way, rather than broadening the coverage of additional information.

(11) Examples of novice-beginners:

- a. He put up glass in uh uh in wine wine in glass. And he talked with eeto nanchuuno eh with waitress waiter. So eeto maybe he asked waiter wine's wine's uh what kind wine. (file01139-SST level 2)
- b. Er there comes party. And she plays piano. Mmm. She ah he have a glass. Waiter have wine. This time is seven o'clock. They have dinner. (file01133-SST level 3)

The descriptions by lower-intermediate learners had fewer grammatical errors, fillers, and repetitions when compared to the novice-beginners. The amount of information also increased gradually. However, they tended to heavily focus on a limited area of the picture, as shown in (12). Although this learner can explain the center part of the image really specifically, the information about the other people and places were not mentioned. From this explanation, the listeners can only get a limited amount of information about the picture. Higher level learners were able to mention the information about the place as a whole, and then the people at the center table, as well as the other

parts. In order to improve their description, lower-intermediate learners would need to broaden the range of information that they provide.

(12) Examples of lower-intermediate learners:

I think mmm this two people urr is mm was was married from, ur mm I think, ten years ago. So urr ten ten years mm anniversary, ur to they come to the restaurant come to this restaurant. And urr I think they they're they they like drinking. So, now, they selected wine ur select wine. Mm. Which is a ur he asked him which is a best wine for me for us. Mm. So ur the mm shop's owner bar bartender ur ur recommend this wine to him.
(file00610-SST level 4)

5.4. Pedagogical implications

There are pedagogical implications that learners can understand to construct more intelligible descriptions based on the major findings and possible reasons for higher scores. There are three implications on future instructions. The first implication is that teachers should prepare and present task purposes explicitly when they carry out picture descriptions. This study revealed that learners and native speakers approached picture description tasks quite differently. One of the reasons for this difference may be related to how they perceived the purpose of the task. While native speakers seemed to describe the overall information in the picture with simple and objective expressions, Japanese learners seemed to describe about a focused area of the picture with more subjective feelings, impressions, and with objective expressions in detail. Appropriate expressions and organizations are different depending on task purposes and we cannot judge which style is more appropriate. Therefore, to define the task purposes in advance is an essential factor. If the purpose is to describe more objectively, then learners can participate in a brainstorming activity together to check the overall information in the picture. In contrast, if the purpose is making a story, then it would be good to discuss the settings of the people and the situation with their full imagination.

The second implication is that it is important for beginners to start practicing from small areas rather than using the entire picture. The findings show that beginners may lack the basic language ability like vocabulary and grammar to successfully accomplish the task. To focus on the various aspects of the task at the same time would be a difficult for them. In order to improve gradually, they would need to practice describing with only one or two sentences and about a limited area. Gradually, this would lead to further steps for constructing sustained descriptions and later adding their own interpretations.

The third implication is that it is necessary to introduce language activities that more proficient learners can do to improve the quality of their picture description. They will become more able to make sustained speech, fluently and cohesively, to some extent. Their problems are based in the amount of information and a limited repertoire of language expressions. Teachers should introduce further language activities so that they can broaden the learners' viewpoints and language variations. For example, learners can check their picture description with each other after they complete it. Through listening to others' descriptions, they can learn new perspectives on the information and different language expressions from each other.

6. Conclusion

6.1. Statements of the present study

This study examined the spoken monologue from various aspects in order to get beneficial suggestions for improving Japanese learners' speaking ability. It is necessary for them to speak their opinions more logically, coherently, and cohesively. To reveal language features among different proficient speakers, a research question, "What information do speakers include as they describe the pictures?" was posed in this study. Two analyses were carried out to answer the research question: content analysis and HCFA (hierarchical configural frequency analysis).

This study could provide interesting findings in terms of the spoken monologue and there were different tendencies between Japanese

learners of English and English native speakers, as well as differences between learners with different levels of proficiency in English. As a difference between learners and native speakers, while native speakers tended to cover a wide range of information with verifiable, simple, and objective expressions, learners tended to describe a limited range of information. Moreover, learners' description included many subjective interpretations. In general, more proficient learners could provide more information during the task. According to the major findings, in order to construct logical, coherent, and cohesive discourse, learners need to care about the informational choice and presentation of their information, using basic language ability like vocabulary and grammar.

6.2. Limitations

This section discusses limitations of the study and the first limitation relates to the coding process because coding the information was carried out only by the author. In fact, reliability and validity are really important factors for discourse analysis and content analysis. This present study should have found another annotator to keep inter-annotator reliability, and to ensure the validity of coding. The second limitation is about task purposes of the picture description in the NICT JLE Corpus. As stated in Chapter 5, the specific task purposes of the picture description task were not indicated explicitly. Although learners and native speakers included the information in the picture differently, we cannot judge which style is more appropriate. If explicit information about the task purposes were available, implications of differences could have been discussed from different perspectives. The last limitation is related to the scoring systems of the SST since the SST is composed of five stages and examinees are asked to finish the interview test in 15 minutes. The spoken data based on the SST was accompanied with the SST score 1–9, which was the total performances of all stages. Thus, the final SST level might not necessarily indicate the level of performance in the picture description task. This might skew the analysis of the corpus data classified by the overall SST levels.

6.3. Future perspectives

This section mentions future perspectives as conclusions to this study. This present study analyzed picture description data in terms of concrete aspects including the information used in the description. As a result, general features and interesting differences were found between L2 learners and native speakers as well as between learners with different proficiency levels. However, there are further opportunities to analyze the picture description tasks. In this study, the overall discourse structures could not be dealt with. Based on the results obtained from this study, general structural patterns of picture description could be extracted. Through applying techniques of move analysis, further interesting findings and implications may be revealed. Moreover, this study chose to focus on only one picture as a research target however the NICT JLE Corpus has an additional six pictures. Although the picture description tasks at Stage 2 uses a single picture, the storytelling tasks have interviewees tell stories using several of them. It would be possible to compare the information in the picture, language expressions and structural patterns depending on the themes as well as the numbers of the pictures. The present author hopes that this study has shed some light on the possibility of corpus approaches toward the study of discourse organization processes for descriptive tasks by L2 learners of English.

NOTE

This is a shortened version of my MA thesis submitted to the Graduate School of Global Studies Tokyo University of Foreign Studies in partial fulfillment of the requirements for the degree of MASTER OF ARTS in Linguistics.

REFERENCES

- Crossley, Scott A and McNamara Danielle S. "Cohesion, coherence, and expert evaluations of writing proficiency." *Proceedings of the Annual Meeting of the Cognitive Science Society* 32 (2010): 984–989. <https://www.researchgate.net/deref/https%3A%2F%2Fescholarship.org%2Fuc%2Fitem%2F6n5908qx>
- Gries, Stefan Th. *Statistics for linguistics with R*. Walter de Gruyter, 2009. <https://doi.org/10.1515/9783110216042>
- Hyland, Ken. *Metadiscourse: exploring interaction in writing*. Continuum, 2005.

- Izumi, Emi, Isahara Hhitoshi, and Uchimoto Kiyotaka. *Nippon-jin 1200-nin no eigo speaking corpus [A spoken corpus of 1200 Japanese EFL learners]*. ALC Press, 2004.
- Izumi, Emi. “Nippon-jin gakusyusya no eigo hatsuwa ni okeru ronritenkairyoku [Logical construction of English speaking by Japanese learners of English].” *Eigo gakusyusya corpus katsuyo handobukku [A handbook of learner corpus research]*. Ed. Tono, Yukio, Kaneko Asako, Sugiura Masatoshi and Izumi Emi. TaishukanShoten, 2013. 178–186.
- Khalil, Aziz. “A study of cohesion and coherence in Arab EFL college students’ Writing.” *System* 17, 3 (1989): 359–371. [https://doi.org/10.1016/0346-251X\(89\)90008-0](https://doi.org/10.1016/0346-251X(89)90008-0)
- Kobayashi, Yuichiro. “NS/NNS tekisutobunruimoderu ni motozuku nippon-jin eisakubun no tokucyocysyutsu [Extracting characteristics of English essays written by Japanese EFL learners on the NS/NNS text categorization model].” *Jinmonkon 2009 Ronbunshu [Jinmonkon collections of academic papers 2009]* 16 (2009): 261–268. <http://id.nii.ac.jp/1001/00090086/>
- Kobayashi, Yuichiro. “Tahenryokaiseki ni yoru sekai no eigo gakusyusya no danwabunseki [A multivariate discourse analysis of non-native Englishes].” *Jinmonkon2010ronbunshu [Jinmonkon collections of academic papers 2010]* 15 (2010): 41–48. <http://id.nii.ac.jp/1001/00090106/>
- Ministry of Education, Culture, Sports, Science and Technology. “Chugakko gakusyushidoyoryo [The guidelines for the course of study of the junior high school].” 2017. https://www.mext.go.jp/content/1413522_002.pdf
- Ministry of Education, Culture, Sports, Science and Technology. “Kotogakko gakusyushidoyoryo [The Guidelines for the course of study of the high school].” 2018. https://www.mext.go.jp/content/1384661_6_1_3.pdf
- Mita, Kaoru and Shimoda Atsuko. “Gakusei no eibun writing ryoku kozo no bunseki sono2: bunpo/kozo/ronri no zyutenteki shido ni yoru writing no syuzyukudobetsuhenka [Analysis of students’ improvement in English writing. part 2: changes in writing proficiency with focused instruction in grammar, organization, and logic].” *Jissen English communication* 51 (2021): 12–46. <http://doi.org/10.34388/1157.00002251>
- Miyasako, Nobuyoshi. “An analysis of high school student compositions in the sentence-based approach.” *Annual Review of English Language Education in Japan*, 11 (2000): 111–120. https://doi.org/10.20581/arele.11.0_111
- Nakagawa, Hiroyo. “Tankidaigakusei wo taisyo to shita paragraph writing no shido zissen no sono koka [The effectiveness of paragraph writing activities for college level writers of English].” *Kenkyuronsyu [Journal of Inquiry and Research]* 104 (2016): 117–127. <http://doi.org/10.18956/00007706>
- Sawaya, Yusuke and Suzuki Tomoki. “Eigo writing ni okeru kessokusei to hyoka no kannkeiseimatomari no aru writing wo mezasita zissentekikenkyu- [The relationship between evaluation and cohesion in English writing: Practical research aims for coherent writing].” *Hokkaido eigo kyoikugakkai kiyo [Journal of English Education in Hokkaido]* 15, 0 (2016): 35–54. https://doi.org/10.24675/helesje.15.0_35
- Tono, Yukio. *Eigo totatatsudo shihyo CEFR-J gaidobukku: CAN-DO risutosakusei katsuyo [The CEFR-J handbook: a resource book for using CAN-DO descriptors for English language teaching]*. Taishukan Shoten, 2013.