The present research aims at examining Japanese EFL (NNS) learner preferences in compiling and consulting corpora as a collaborative activity. For a presentation task, each group compiled a native speaker (NS) corpus and compared it to a Japanese L2 learner corpus (LC) that included their own writing in a narrative topic type. The result displayed most groups being less conscious of genre-based corpus compilation and their research focusing on limited functional words, i.e., prepositions, pronouns, and infinitives. Their findings are simply based on difference in frequency rate in terms of over/underuse and there is no investigation into the reasons for the results that learners found. It is noted, however, that some groups notably compiled the same learner corpus format (genre) as that of the NS corpus. Others conducted intense analysis of modal verbs, infinitive usage, or lexical research into intra-genre comparison (e.g., NNS vs. spoken NS vs. written NS).

1. Introduction

In previous studies, the following usages of corpus are reported:
(1) for teaching grammatical rules, inductive and deductive approaches (Sun, Y & Wang, L, 2003)
(2) for collocation learning (Kita, K & Ogata, H, 1997)
(3) for fostering writing ability (Todd, R, W, 2001; Nesselhauf, N, 2004)
(4) for focusing on lexico-grammatical features and increasing “contextual and linguistic awareness” of L2 learners (Tribble, 2002, p. 133)
(5) for correcting grammatical errors (Gaskell & Cobb, 2004; O’Sullivan & Chambers, 2006; Hegelheimer, 2006)

Yoon & Hirvela (2004) employed a similar research framework to the present study and focused on the perspective of corpus use in ESL academic writing courses. They concluded that the student’s perception of the corpus approach contributed to development and confidence in regard to L2 writing skills.

O’Sullivan & Chambers (2006) also supported corpus consultation for enhancing learner’s writing skills. They recommended corpus consultation by stating, “corpora can provide students with target language input in the form of authentic texts of naturally occurring examples of real language use,” (p. 52) and can, therefore, foster writing ability. They added the notion that corpus consultation “enhances learner autonomy” (p. 53) by encouraging learners to uncover the forms and patterns of language use in the targeted corpus.

The present study aims at investigating learner ability (autonomy) to discover differences between focused features of EFL learners and those of native speakers. The theoretical framework of this research is founded upon corpus use for developing the writing ability of L2 learners and creating the autonomy required to discover differences in focused forms and patterns between learner corpus and learner-compiled corpus(es). The following research questions arise next in the study:
1. Do L2 learners employ genre-based perspectives to learner corpus in compiling corpus?
2. On what lexical and grammatical features do they focus?

2. Method

In order to build learner corpus, learners wrote narrative essays titled “self-introduction”. The essays of the 48 participants were, then, assembled into one corpus to function as one NNS comparative corpus. When building NS (native speaker) contrastive corpora, LC was deemed as required.
The other corpus(es) for comparison with LC, was/were allowed to be chosen by the learner group. Each group discussed the size, genre/register, age, and source of contrastive corpus to compile. If two corpus sizes are different, frequency rate of occurrence should be estimated according to each corpus. The instructor served as a facilitator to avoid problems and offered advice to encourage each group to reach the required destination.

Each group composed of 3-5 self-selected members and the 48 participants were thus divided into 13 groups. Six continuous lessons were allotted, including a presentation period.

To analyze a corpus compiled in text format, Antconc is adopted. Antconc is free software that enables users to conduct corpus analysis by offering the following functions: concordancer, wordlist, keyword list, and clustering.

3. Result

As in Table 1, with regard to genre/register perspectives in compiling corpora to make contrastive research, almost half the groups (6/13) compiled corpuses of different language types and mixed spoken and written languages. Only two groups attempted to establish a corpus of the same genre in the format of written self-introduction. Moreover, in terms of genres, 3 groups chose written fictions from the 18th and 19th centuries. Considering these results, Japanese EFL learners seem to pay little attention to differences in genre and register. With regard to the first research question, therefore, L2 learners exercise insufficient genre-based perspective when compiling and analyzing comparative corpus.

With regard to the second research question, lexical and grammatical features that learners employed are: pronouns (4 groups), conjunctions (4 groups), verbal features including infinitives (3 groups), prepositions (2 groups), modals (1 group), adverbs including intensifiers (1 group) (Note that three groups addressed multiple research questions). Main findings in learner research are as follows (G + numeral: group number).

3.1 Nouns and pronouns

(1) Fictional prose has more proper nouns and less pronouns than the movie script. (G1)
(2) The current US president offered high frequency of we as a subject in his speech, whilst the first president used many I's. (G10)
(3) LC contains twice as many I's as the NS corpus even though both corpuses belong to the same genre of self-introduction. (G13)

3.2 Conjunctions

(1) And, so, and because have a high frequency rate in LC. (G1)
(2) And/but + pronoun occurring in the initial position of a sentence appears more than 5 times as frequently in LC as it does in NS corpus. (G4)
(3) The relative pronoun that has a lower frequency in LC than it has in NS corpus. (G1)
(4) In NS corpus, and is frequently used for connecting sentence/clause and sentence/clause, but in LC, it is mainly used to connect single-standing words. (G5, G13)

3.3 Prepositions

(1) In LC, at and on are mainly used for expressing time, while in is used with spatial meaning. In NS corpus, by is often used to note means of

Table 1. Corpus Features of 13 Groups

<table>
<thead>
<tr>
<th>Group No</th>
<th>LC(NNS)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoken/ Written</td>
<td>W</td>
<td>W</td>
<td>S</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>S</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>W</td>
</tr>
<tr>
<td>Period</td>
<td>21st</td>
<td>21st</td>
<td>90s</td>
<td>21st</td>
<td>21st</td>
<td>90s/50s</td>
<td>21st</td>
<td>19th</td>
<td>21st</td>
<td>19th</td>
<td>21st/18th</td>
<td>21st</td>
<td>21st</td>
<td>21st</td>
</tr>
<tr>
<td>Corpus size</td>
<td>11457</td>
<td>S/S</td>
<td>S/S</td>
<td>M</td>
<td>S/S</td>
<td>M</td>
<td>L</td>
<td>M</td>
<td>L</td>
<td>S</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td></td>
</tr>
</tbody>
</table>
action and as is used following an adjective or as a conjunction. (G8)

(2) In NS corpus, like shows high frequency of prepositional usage in exemplary description with a similar meaning to as if and for example, but in LC such usage is found less frequently. (G6)

3.4 Verbs and modals

(1) In LC, 70% of infinitives are used in a collocation of want to, while in NS corpus 50% of infinitives employed adverbial usage to show aims and results. Movie scripts include infinitives to modify preceding nouns in 28% of cases, which is the highest of all usages. (G2)

(2) With regard to the usage of causative verbs such as make, get, have, and let, let in NS corpus shows more than 5 times the frequency rate of that in LC. Verb forms accompanying bare infinitives (make, have, let) and to-infinitives (get) are less common than those found in NS corpus, and in particular, make is often used in containing a single object form. (G3)

(3) NS corpus includes many examples of shifting from verb to noun, whilst in LC the lexical variation of verbs is limited in number and usage. Play, for example, is found to be used exclusively as in “having fun” without any wider meanings such as in play a role, play on words, or play truant. (G6)

(4) NS corpus contains high usage of modal verbs such as could (11 times as frequent as in LC), should (3 times as frequent), would (27 times as frequent), and must (3 times as frequent). Might appears 14 times and shall appears 15 times in NS corpus but frequency cannot be compared with that in LC, because LC contains neither of them. It should be noted that this group concludes that NS corpus includes a wide range of affirmative degree and NNSs have limited knowledge regarding modals and that difference of genre of corpus (narrative LC vs. fictional NS) might influence the features of each corpus. (G7)

3.5 Others

(1) Very much in LC appears 23 times more frequently than in NS corpus. Only one group pays attention to the usage of negation and notes that NNSs often use “negation + many/much” form. (G4)

(2) The NS corpus has a wide variation of adverbs such as just, never, and no. All of them show more than 100 times the frequency than that found in LC. In contrast, in the case of so, Japanese learners use twice as many so’s as NSs. (G11)

4. Discussion

In order to guide learners to proper goals, a plausible research design should be established. Focusing on similar genre, size, and ages of production along with authenticity of the corpus compiled might aid in this.

Regrettably, most of their research discusses over/under use of several usages of a functional word such as a popular preposition (of, in, on, at) or a prevalent verb(s) (have, make, play). Deeper investigation into the reasons for the discoveries is required.

In short, research stays at the surface level, reflecting a lack in collocational research and semantic contrast in rhetorical influence.

However, analysis in learner needs will shed some light on the effective contribution of native speaker corpus on language learning. For example, G1 successfully notices what modal means in text construction by studying functional features of auxiliary verbs. This is one of the areas in which Japanese EFL learners in my research have been less conscious (Kashiwagi, 2006). Another group (G2) conducts comparative research into infinitive usage in three phases: spoken, written, and LC. They found clear contrast in functional difference in those three phases.

Several groups propose that Japanese learners show excessive preference for: in referring to time sequence rather than spatiality, on referring to spatiality, want to referring to writer expectation, and and as a word connector instead of a sentence connector.

Most of the groups agree that wider variation of
the researched features occurs in NS corpus and limited number of usages is found among Japanese learners. Japanese learners also maintain the importance of analysis of authentic usage of native speakers. Sharing research results with other learners also contributes to developments in distinguishing quasi-English from authentic English.

The contrast found in the study of pronoun as subject used in two president’s speeches provides some suggestions in the socio-linguistic field and offers historical perspectives on a leader of a country and its people (I vs. We: G10).

The corpus of a poetic genre seems unusual to the author, but most L2 learners are interested in popular song lyrics to a great extent. Research into lyrics of The Beatles (British, 1960s, Males) and Mariah Carey (American, 1990s, Females) leads to the selection of words such as no, never, let, know to contrast with LC and between the artists themselves in key-wordlists. To the author, gender should be the main reason for contrast rather than nationality or age.

A summary of the contributions to EFL learners is:
1. Noticing differences in lexical, grammatical features, function and meaning, overuse and underuse in targeted corpuses
2. Confirming unclear matters in the learning process with regard to cognitive stimulus to EFL learners through discussing and determining research items
3. Activating consciousness to monitor differences between learners’ L2 output and authentic English

In other words, finding grammatical and lexical gaps between rhetorically influenced corpuses and authentic corpuses will be useful as an activator of second language learning. Hopefully, most learners find the activity to be helpful in learning authentic English and discover that they need more knowledge regarding L2 grammar and lexis.

References