博士論文英文要旨

Title：Lexical Analysis on Learners' Oral Production of French

Name: Kaori SUGIYAMA

The purpose of this study is to investigate lexical use in oral production by Japanese learners of French and to describe their features. Previous studies having concluded that there are no systematically constructed Japanese learners' corpus of French, two types of learner corpora were built in order to cover a large proficiency level: "task-based corpus" for beginners and "free conversation corpus" for learners just above near-beginner level. The former consists of monologic data that come from tasks performed by A1-A2 level learners as described in the Common European Framework of Reference for Languages, and the latter consists of dialogic data based on conversations carried out in pairs. As referential corpora, we built "task-based corpus" of native speakers of French and we used native corpus elaborated by Tokyo University of Foreign Studies and University of Aix-Marseille. We applied "Contrastive Interlanguage Analysis (CIA)" proposed by Granger (1996), where two comparative approaches are used: one is the comparison of the production by native speakers with that of learners and the other is the comparison within the learners' group differing in competence level. With this CIA method, quantitative and qualitative analyses on the following four lexical points of view are conducted: "lexical richness", "key-words", "collocation" and "N-grams".

This thesis is composed of 10 chapters.

In the introductory chapter, the purpose and the structure of this study are indicated.

In chapter 1, "Corpus linguistics", after defining the word "corpus" linguistically, the development of corpus linguistics will be reviewed. In particular, we focus on measurements of interlanguage and lexical analysis based on learner corpus.

In chapter 2, "Word level Research –Research on Word Units", arguments about the "unit of a word" will first be reviewed. Then, the expansion of lexical studies with emphasis on word lists development, which has laid the foundation of lexical researches on word level, will be examined. After confirming that lexis plays a central roll in second language acquisition, the way to measure lexical competence by focusing on lexical richness and over-/under-use will be discussed. Then, some previous researches on learners' lexical use will be reviewed.

Chapter 3, entitled "From Word Units to Multi-Word Units", includes explanations of Multi-Word Units (MWUs), which is composed of at least two words. After reviewing the historical background in MWUs studies, several measurements of MWUs, particularly for collocation and N-grams analysis, will be examined. In the last part of this chapter, previous studies of MWUs of foreign language learners will be overviewed.

Chapter 4 gives an explanation of the research methods, including research objectives, corpus data, structure of this study and research questions.

In chapter 5, "Analysis on Lexical Richness", the first part of my analysis, lexical richness is measured by means of *VocabProfil*. It enables us to measure lexical sophistication (*VocabProfil* provides frequency information of words by classifying them into frequency zones), lexical density (*VocabProfil* counts content words and function words separately), and lexical diversity (*VocabProfil* automatically counts numbers of tokens and types). Results of analysis on lexical richness show that lexical measurement based on lexical sophistication, lexical density and lexical diversity is highly reliable. Moreover, it is revealed that this measurement can estimate learners' linguistic competence. Also, learners with high level of French have more similar lexical richness to French native speakers than lower level learners.

Chapter 6, "Analysis on Key Words", gives descriptions about over-/underused words, which are calculated by means of log-likelihood. For this study, overused words are indicated by $G^2 \geqq 50$ and underused words, by $G^2 \leqq$ -50. Results of qualitative keyword analysis show that almost all keywords are composed of highly frequent words. Besides, overused words are characterized by nouns and pronouns, and under-used words, by adverbs and polysemy.

In chapter 7, entitled "Analysis on Collocation", how keywords are employed in contexts will be analyzed. The results show several differences in lexical use between

learners and native speakers as follows: overuse of personal pronouns, formal language use, underuse of discourse markers and that of vagueness tags. However, advanced learners' language use is more similar to that of the native speakers than that of the lower level learners: they use less formal language use, which is characterized by omission of negative clitic *ne* and interrogations by intonation. Actually, native speakers also prefer to use this sort of informal language use in a colloquial conversation. The differences between advanced learners and lower level learners can also be seen in communicative strategy: lower level learners use more foreign origin words, while advanced learners employ *comment dire* in order not to rupture the flow of communication.

In chapter 8, "Analysis on N-grams", the features of N-grams used by learners are described. Firstly, we clarify that high frequency N-grams are dominant in learners' language use. By analyzing over-/under-used N-grams, we verify that errors, underuse of complex phrases and that of discourse markers characterize learners' language use. Comparison between advanced learners and lower level learners based on free conversation corpus shows that lower level learners use more repetitions, *ne* as negative clitic and interrogative form "EST-CE QUE", while advanced learners speak in the same manner—as native speakers; they omit *ne* for negations and use the rising-intonation questions.

The concluding chapter overviews the analyses, findings and discussions of this study. Some perspectives for future studies in the field of French language pedagogy will also be suggested.