A STUDY OF L2 KANJI LEARNING PROCESS

Analysis of Reading and Writing Errors of Swedish Learners in Comparison with Level-matched Japanese Schoolchildren

Fusae Ivarsson

Department of Languages and Literatures

UNIVERSITY OF GOTHENBURG

Doctoral dissertation in Japanese, University of Gothenburg,
18 March, 2016

Cover: Fusae Ivarsson, Thomas Ekholm

Print: Reprocentralen, Campusservice Lorensberg, Göteborgs universitet, 2016

Distribution: Institutionen för språk och litteraturer, Göteborgs universitet,
Box 200, SE-405 30 Göteborg

ISBN: 978-91-979921-7-6

http://hdl.handle.net/2077/41585

ABSTRACT

Ph.D. dissertation at the University of Gothenburg, Sweden, 18 March, 2016

Title: A Study of L2 Kanji Learning Process: Analysis of reading and writing errors of Swedish learners in comparison with level-matched Japanese schoolchildren.

Author: Fusae Ivarsson

Language: English, with a summary in Swedish

Department: Department of Languages and Literatures, University of Gothenburg, Box 200, SE-405 30 Gothenburg, Sweden

ISBN: 978-91-979921-7-6
http://hdl.handle.net/2077/41585

The present study investigated the characteristics of the kanji learning process of second language (L2) learners of Japanese with an alphabetic background in comparison with level-matched first language (L1) learners. Unprecedentedly rigorous large-scale experiments were conducted under strictly controlled conditions with a substantial number of participants. Comparisons were made between novice and advanced levels of Swedish learners and the respective level-matched L1 learners (Japanese second and fifth graders). The experiments consisted of kanji reading and writing tests with parallel tasks in a practical setting, and identical sets of target characters for the level-matched groups. Error classification was based on the cognitive aspects of kanji. Reading errors were classified into phonological, circumstantial, orthographic and semantic types, and writing errors into the same four types and an additional pseudokanji type. The error type occurrence patterns were analysed according to skill (reading/writing), level (novice/advanced) and the learner groups' L1 (Swedish/Japanese), with a focus on the kanji processing unit, preferred methods of character/pronunciation retrieval from the mental kanji lexicon and reading and writing difficulties.

This study made a number of new findings and verified various observations made in previous studies. Some of the findings that are unique to this study are: (i) L1 phonological transfer for Swedish novice learners and its decrease at the advanced level; (ii) L2 learners' less developed configurational awareness and lesser degree of inter-level development than L1 learners; and (iii) a shift in inter-level characteristics for L2 learners, while these remained consistent for L1 learners. The hypotheses confirmed include the following characteristics of L1 alphabetic learners: phonological approaches to retrieval, component-based units of processing, predominance of pseudokanji type writing errors, and greater inter-writing system differences in reading but greater inter-level differences in writing. This study demonstrated that the challenges experienced by L1 alphabetic learners stem from the shift from phoneme-based to component-based processing of graphemes, taking the less familiar lexical route in the decoding and encoding of grapheme-sound correspondences, and the use of less efficient strategies in reading and writing.

Keywords: Kanji, L2, L1 transfer, alphabetic writing system, error analysis, Swedish, reading and writing, cognitive aspects, level matching, schoolchildren.

ACKNOWLEDGEMENTS

During my years of teaching Japanese at the University of Gothenburg, I came across various types of recurring errors in kanji reading and writing made by my students. Some types were familiar to me as a native user of kanji, and others seemed peculiar to L2 learners of kanji, or even unique to Swedish students. There were level-specific errors, and certain learner groups always made the same types of errors. After a while, I started detecting correlations between these error types and different cognitive aspects of kanji. This experience interested me in comparing level-matched L1 and L2 learners and analysing their kanji errors from a cognitive point of view in order to investigate the Swedish learners' error occurrence tendencies and to explore the implications for L2 kanji teaching.

It has taken a long time to complete this dissertation, partly because of the amount of the data collected in this study, and partly due to health issues. During these years I have received cooperation, support and encouragement from countless people, to all of whom I am deeply grateful.

First of all, I would like to thank my supervisor Yasuko Nagano-Madsen, who provided me with academic guidance and moral support. Her advice, encouragement and quick replies even outside of normal working hours have been greatly appreciated.

I also owe sincere thanks to those who were involved in the experiments: Hisashi Fusegi, who brought me in contact and helped me make arrangements with the Japanese primary schools; the pupils at these schools and at my alma mater, who spared me a few precious class hours for taking the tests, as well as their teachers, who allowed me to give the tests in their classes and gave me kind assistance; Sawako Murao and Yukiko Shimizu, who assisted me greatly in carrying out the tests at the respective Swedish universities; and of course, all the students who took the tests there; and the Japanese children who lived in Gothenburg at the time and participated in the pilot test. Without the support of all of you, the experiments could not have been conducted.

I owe very much and am deeply grateful to Shinichiro Ishihara, who acted as the opponent at my final seminar and thereafter as the assistant supervisor. Without his overall advice and instructions on statistical processing, I would not have been able to finalise this work to a satisfactory level. I am also very thankful to Laura Downing for reading my draft and giving valuable advice at the final seminar.

I thank all the personnel at the Department of Languages and Literatures, as well as at my previous Department of African and Oriental Languages, and more particularly my present and former colleagues in the Japanese section of the Department: Noriko Thunman, Kazuyo Lundström, Satoko Berger, Martin Nordeborg, Pia Moberg, Yuka Okamoto-Gustafsson, Thomas Ekholm, Yoko Takau-Drobin, Yoko Kumagai, Sachiyo Larsson, Naoya Irikura, Lisa Pääjärvi and Máté Sall Vesselényi, for all their reading and discussions with me, as well as their comments, advice, help, support and encouragement.

Many international colleagues have commented on parts of my work and given feedback at conferences and by e-mail. I thank them all, and most particularly Toshiko Ishida, Yoko Matsumoto-Sturt and David Hebert for their input and advice.

For contributing to the funding of field trips and other expenses I gratefully acknowledge Kungliga och Hvitfeldtska stiftelsen, Adlerbertska Stipendiestiftelsen, Stiftelsen Paul och Marie Berghaus donationsfond and Stiftelsen Stipendiefonden Viktor Rydbergs minne.

Finally, I express my sincere gratitude to my friends and family for all the support, help and encouragement. Special thanks to Halbe for the initial set up of the database, to Wolmet for helping me with the statistical processing, and to Adrian for reading and correcting my earliest draft. Last but certainly not least, I am deeply grateful to my family in Japan and Sweden: my parents Yoshihisa and Kimiyo, my brothers Harutomo and Tetsuji, my parents- and sister-in-law Lisbeth, Jan and Petra, and my husband Patrik and daughter Erika. Thank you all for everything!

Gothenburg, February 2016
Fusae Ivarsson

TABLE OF CONTENTS

TABLES

FIGURES

ABBREVIATIONS AND SYMBOLS

A	approach to retrieval
ACC	accusative case particle
CFA	configural frequency analysis
CFL	Chinese as a foreign language
D	displayed knowledge
GEN	genitive case particle
GSC	grapheme-sound correspondence
INS	instrumental case particle
L1	first language, native language
L1WS	first language writing system
L2	second language
L2WS	second language writing system
LV240	level at which approx. 240 kanji have been learned
LV800	level at which approx. 800 kanji have been learned
LOC	locative case particle
NJ	native Japanese learners
NJ240	native Japanese learners who have learned approx. 240 kanji
NJ800	native Japanese learners who have learned approx. 800 kanji
NOM	nominative case particle
NON-PAST	non-past tense suffix
PAST	past tense suffix
PSC	phono-semantic composite
Q	question

SC	semantic composite
SJ	Swedish learners of Japanese
SJ240	Swedish learners of Japanese who have learned approx. 240 kanji
SJ800	Swedish learners of Japanese who have learned approx. 800 kanji
TOP	tip-of-the-pen
TOPIC	topic marking particle
U	unit of processing
WS	writing system
−	lacking knowledge
?	uncertain aspects, disregarded discrepancy or underdeveloped ability
[ ]	phonetic representation
/ /	phonological representation

GLOSSARY OF JAPANESE WRITING SYSTEM AND PHONOTACTICS

geminate consonant	doubled (non-nasal) consonants
hanzi	Chinese characters developed to write the Chinese language
hiragana	a set of *kana* that derived from a whole kanji character and are used mainly to write function words
Jōyō Kanji	the officially announced guide to kanji characters for regular use
kana	phonetic script in which each symbol represents a mora
kanji	Chinese characters adopted to write the Japanese language
katakana	a set of *kana* that derived from a partial kanji character and are used mainly to write loan words and onomatopoeia
kun-reading	a pronunciation of a kanji character which is an interpretation of its meaning into native Japanese
mora (pl. morae)	a syllable-like phonetic unit conceived as being temporally constant
moraic nasal	a nasal consonant that constitutes a mora alone, transcribed as /N/
moraic obstruent	a fricative or plosive consonant that constitutes a mora alone, coincides with geminate consonants and is transcribed as /Q/
okurigana	an inflectional ending written in *hiragana*

on-reading	a pronunciation of a kanji character based on its original Chinese pronunciation
palatalised consonant	a consonant which is pronounced by moving the point of contact between the tongue and the palate forward in the mouth
phono-semantic composite	a Chinese character created by combining a phonetic radical with a semantic radical
radical	an intra-character component of kanji (and *hanzi*) under positional constraint with semantic and/or phonetic cueing functions
rendaku	sequential voicing
romaji	Roman letters adopted to represent the pronunciation of the Japanese language
semantic composite	a Chinese character created by combining two semantic radicals
sequential voicing	conditional voicing of the initial voiceless consonant of the non-initial morpheme of a compound
special morae	long vowels, geminate consonants and moraic nasals
yomi	"reading" or the pronunciation of kanji

1. INTRODUCTION

1.1. The Present Study

The Japanese writing system is a systematic intermixture of logographic, syllabic and alphabetic writing systems. Learners of Japanese must learn the orthographies of logographic **kanji** (Chinese characters adopted to writing the Japanese language), moraic ***kana*** (two sets of Japanese phonetic scripts in which each symbol represents an approximation of a syllable) and alphabetic ***romaji*** (Roman letters adopted to represent the pronunciation of the Japanese language) to gain working knowledge of written Japanese. Among the three orthographies, meaning-based kanji, which forms the basis of the written Japanese vocabulary, is of the utmost importance and requires the greatest effort to learn because of the copious number and configurational complexity of the characters and the opaqueness of the grapheme-sound (character-pronunciation) correspondence.

The present study aims to explore the characteristics of the kanji learning process of Swedish learners of Japanese, focusing on the unit of processing kanji, how they attempt to retrieve from memory the pronunciation and form of the characters they have learned, and their difficulties in reading and writing kanji.

In order to achieve this objective, this study compared Swedish learner groups with level-matched groups of first language (L1) learners (Japanese schoolchildren) to illuminate the second language (L2) characteristics of Swedish learners. It examined kanji reading and writing errors of novice and advanced groups of Swedish learners and those of Japanese learner groups at the corresponding levels through carefully designed experiments under well-controlled conditions, and analysed each group's error occurrence tendencies in order to identify the Swedish characteristics of kanji learning process.

1.2. Background

1.2.1. Learning of L2 writing system

The learning of second language writing systems (L2WS) is not an easy task (Cook & Bassetti 2005). It is difficult enough even if the same scripts are used in L1 and L2, as in the case of native English speakers learning Swedish. Although both languages are written in the Roman alphabet, English and Swedish are spelled and pronounced differently. A grapheme (a written symbol, or the smallest unit in a writing system) corresponds to a

phonological unit and vice versa according to certain rules, and different languages have different grapheme-sound correspondence (GSC) rules. L2WS learning becomes rather challenging if learners have to learn not only different GSC rules but also a new script (e.g. native English speakers having to learn the Cyrillic alphabet to read and write Russian).

The learning is quite a struggle when the L1 and L2 employ different writing systems, as is the case when the L1 is English, using the Roman alphabet (a sound-based writing system), and the L2 is Chinese, using Chinese characters (a meaning-based writing system), or vice versa. English learners of Chinese are required to learn a new writing system in which a grapheme bears not only sound but also meaning. In addition, the graphemes to be learned for a meaning-based system are highly complex and extremely numerous: learners must acquire several thousand configurationally complex characters instead of a couple of dozen visually simple letters.

The above comparisons should aid realisation of the challenges experienced by L2 learners of the Japanese writing system, especially those whose first language writing system (L1WS) is alphabetic. The learning of the Japanese writing system presents the difficulty of learning *romaji*, an alphabetic writing system with different GSC rules from those of learners' L1WS; the challenge of learning *kana* script (in which each symbol represents a quasi-syllabic sound unit); and the struggle of learning kanji, which is an orthography involving a few thousand visually and phonologically complex characters in an unfamiliar meaning-based writing system. Furthermore, mastery of the Japanese writing system requires the extra effort of learning the proper way of combining these three orthographies.

1.2.2. Research on second language writing system

Cook & Bassetti (2005) summarise how L2 literacy has become one of the prominent topics of exploration in the areas of psychology, education, linguistics and L2 acquisition research over the past twenty years: research on L2WS covers reading, writing, learning and awareness of L2WSs by learners who are already literate in their L1, and it extends over a variety of disciplines, such as applied linguistics, psycholinguistics and neurolinguistics. An important issue of research enquiry has been the influence of the L1WS on the L2WS, and a particular focus of attention has been on the transfer between sound-based and meaning-based writing systems; research on L2WSs has repeatedly shown that L2WS users behave differently from L1WS users, and also from L2WS users with other L1WS backgrounds. Such differences are ascribable to transfer from the

learners' L1WS as well as from other writing systems they have learned earlier.

There are two major types of methods in L2WS research, namely, experimental methods and descriptive methods. Experimental methods for reading and writing involve tasks such as word naming (e.g. Akamatsu, 2005), silent reading with eye tracking (e.g. Bernhardt & Everson, 1988), lexical judgment (e.g. Chikamatsu, 1996) and spelling tests (e.g. Okada, 2002). Descriptive methods are based on the collection and description of L2 learners' writing. The main technique in this collect-and-describe approach is error analysis, a starting point of which is Corder (1974), in the area of L2 acquisition research.

Regardless of the types of methodology, there are areas of L2WS studies that have been scarcely explored: most L2WS studies use a single group of L2 learners with the same L1 background as subjects, and there is a dearth of studies in which comparisons are made between L2WS learners and L1WS users, or between groups of L2 learners with different L1WS backgrounds; and the focus of research has been on the reading process, while the writing process has not been well investigated (Cook and Bassetti, 2005).

This imbalance of subject groups' L1WS and skill type is even more noticeable in L2WS research on Japanese kanji. There have been only a small number of such studies to begin with, and the majority of those are on L2 kanji word reading/recognition, such as Matsumoto-Sturt (2004) with learners with the same L1WS background, and Matsumoto (2013), Mori (1998) and Tamaoka (1997) with L2 learner groups of different L1WS backgrounds. Studies on L2 kanji writing, especially those involving error analysis of handwritten characters, are quite limited in number. Among those few, Okita (2001) compared L2 learner groups of different L1WS backgrounds, and Chikamatsu (2005) and Hatta et al. (1998, 2002) are the only ones comparing L2WS learners and L1WS learners/users.

These previous kanji error analysis studies comparing L1 and L2 groups had a number of limitations: they used experimental (L2) and control (L1) groups at different levels, which led to the use of different test materials with non-identical sets of target kanji, or of free writing without a particular setting of the target kanji. The use of non-identical sets of target kanji or the lack of target kanji leads to lack of control in terms of the characters' phonological or semantic transparency, frequency, complexity, position within the word, etc. In some cases, data collection methods were also different between the two groups. Under such conditions, it is unclear if the observed differences in patterns are based on the level difference,

dissimilar data sources or L1WS transfer. Furthermore, there have been no studies comparing both reading and writing errors of L1 and L2 learners under the same conditions. Another issue in previous studies is the choice of the experimental group's L1 background. The great majority of such groups are native English speakers, and there has been no L2WS research on kanji reading and writing involving Swedish native speakers as the experimental group, except for Ivarsson (2011) and Ivarsson (2016 forthcoming), which are research plan notes and preliminary reports on the present research.

As mentioned earlier in this subsection, transfer between sound-based and meaning-based writing systems has been a focus of research interest, and therefore a study that explores transfer from the sound-based alphabetic L1WS to the meaning-based logographic L2WS is of undeniable significance. Although a multitude of research has confirmed that L2WS users, due to influence of their L1WS, behave differently from L1WS users, neither the difference between L2 learners with alphabetic background and L1 learners/users nor the L2 learners' L1WS transfer has been well investigated where kanji as the L2WS is concerned. The few previous studies have pointed out that the observed differences may be based on the level difference between the L1 and L2 learner groups (Hatta et al., 1998, 2002) and that the level difference leads to testing the two groups on different sets of target characters with unbalanced features (Chikamatsu, 2005), which could be a factor affecting the results apart from L1 transfer.

It is therefore crucial to investigate the differences between level-matched L1 and L2 learner groups of kanji. The nature and extent of L1WS transfer can be effectively examined by comparing L2 learners' results with those of L1 learners. Level-matching between L2 and L1 learner groups is essential, and comparison of two different levels of both learner groups and examination of both reading and writing errors would provide a more precise and comprehensive description of the differences. A study with Swedish learners as the L2 learner group will provide valuable pedagogical data for this rarely investigated learner group of Japanese. Furthermore, it will present new research data from a group of learners with a single alphabetic L1WS, on the basis of which to explore similarities and differences between L1WS alphabetic learners with different L1 backgrounds.

1.2.3. The problem

Acquisition of kanji is often regarded as an arduous task for L2 learners of Japanese, especially for those with a phonographic background (sound-based L1WS), the majority of whom have an alphabetic writing system as

their L1WS (Gamage, 2003; Mori, 1999; Mori & Shimizu, 2007; Okita, 1997, 1998; Toyoda, 1995, 1998; Yamashita & Maru, 2000). Many L2 learners with alphabetic backgrounds find the task daunting because of the copious number and configurational complexity of the graphemes (kanji characters). To make matters worse, an abundance of homophonous characters and characters with multiple readings makes the GSC very opaque (i.e. many graphemes to one sound and vice versa), thereby requiring additional effort for mastery of the writing system.

However, these conditions are no different for L1 Japanese speakers/readers. Both L1 and L2 learners have to learn approximately 2,000 characters of the Jōyō Kanji (常用漢字), the official guide to kanji characters for regular use,[1] with the same visual complexity and phonological opaqueness, in order to acquire working knowledge of written Japanese. L1 learners are required to learn them by the end of their secondary education, and L2 learners to pass the highest level of the Japanese Language Proficiency Test. What, then, are the particular challenges experienced by the L2 learners with alphabetic backgrounds? What are their difficulties in recognising and producing kanji, and how do they process kanji and retrieve them from the mental lexicon when tackling reading and writing tasks? What are the general tendencies and how do these learners differ, according to skill and level, from L1 learners?

As mentioned in the previous subsection, there has been a dearth of L1/L2 comparative descriptive research on kanji, especially on the writing of kanji. The few previous studies had a number of limitations such as the unmatched levels of the participant groups, which led to the use of different test materials with non-identical sets of target kanji (therefore not equally controlled in terms of the characters' intrinsic features and extrinsic factors); or unmatched data collection methods, making the bases of the observed differences (the unmatched levels, the unmatched data sources or L1 influence) unclear. Furthermore, there have been no studies comparing both kanji reading and writing errors of L1 and L2 learners under the same conditions. There has also been no research on kanji reading and writing with L1 Swedish participants, except for research plan notes and preliminary reports on the present research (Ivarsson, 2011; Ivarsson, 2016 forthcoming).

In view of the current situation in the research field, it is necessary to conduct new research to analyse kanji reading and writing errors of Swedish learners as L2 learners with an alphabetic background in comparison with those of level-matched L1 learners, using the same data collection methods and identical test materials.

1.3. Aim

The aim of the present study is to find out the characteristics of Swedish learners of Japanese concerning their kanji interlanguage (a new "in-between" system created in learners' mental lexicon, which deviates from the target writing system under the influence of their own L1WS) in the developmental process of kanji learning, especially their units of processing, approaches to retrieval and difficulties in recognition and production. An effective way to achieve this objective is to examine their error occurrence patterns in kanji tests as compared to those of L1 learners. Using tests rather than essays and other styles of text writing as data collection method prevents the participants from avoiding the use of kanji they are uncertain of, and comparing the results of the L2 learners with those of the L1 learners will enable us to identify the L2 learners' characteristics. Examination of both reading and writing errors at different levels in comparison with level-matched Japanese learner groups should provide a more comprehensive picture of their characteristics and a better understanding of their developmental process of kanji learning.

1.4. Research questions

This study analyses the kanji reading and writing errors of Swedish learners of Japanese (as L2 learners with an alphabetic background) in comparison with Japanese schoolchildren at the equivalent level (as level-matched L1 learners). By examining the error occurrence patterns of both groups and identifying the factors involved, the study aims to explore the Swedish learners' approaches to retrieval and difficulties in reading and writing, and the possible influence of their L1 on processing, retrieval, recognition and production of kanji.

In short, comparisons will be made between the following three parameters:

Skill: **reading** vs. **writing**

Level: "**LV240**"
(the novice level at which 240 characters have been learned)
vs.
"**LV800**"
(the advanced level at which 800 characters have been learned)

L1: **SJ** (Swedish learners of Japanese)
vs.
NJ (native Japanese learners)

The data source will be errors collected from (i) a set of kanji reading and writing tests for LV240 taken by Swedish and Japanese novice learner groups (**SJ240** and **NJ240**) and (ii) a set of reading and writing tests for LV800 taken by Swedish and Japanese advanced learner groups (**SJ800** and **NJ800**).

In order to achieve the aim stated in section 1.3, the research questions for the present study are set as follows:

A. What are the kanji error type occurrence patterns and their similarities and differences according to skill (reading/writing), level (LV240/LV800) and L1 (SJ/NJ)?

B. What are the units of processing, approaches to retrieval and difficulties in reading and writing of kanji of SJ and NJ in their respective developmental stages?

C. What are the characteristics of the kanji learning process of SJ?

The above research questions are of particular interest and significance because no previous studies have carried out such multifaceted comparisons under strictly controlled conditions in order to explore the characteristics of the kanji learning process of L2 learners, especially of Swedish learners, a learner group that remains unexplored in the field of L2 Japanese research. Another point of interest is that the errors will be collected from a test setting using material that aims to reproduce everyday usage of kanji. Except for Hatta et al. (1998, 2002) and BAASC (2007), previous studies were carried out in an experimental setting disconnected from everyday kanji usage, and therefore did not provide data to show learners' usual natural behaviour.

1.5. Outline

This dissertation consists of eight chapters. Subsequent to this introductory chapter, Chapter 2 will explain the Japanese writing system and the concepts related to kanji error classification, presenting theories, facts, data and examples.

Previous studies will be reviewed in Chapter 3 in order to explore findings and discussions in the fields relevant to this study, namely, theories of writing system studies, kanji recognition and processing, and kanji production. Each of the latter two fields will be divided into L1 studies, L2

studies and studies with L1/L2 comparison. A brief review of phonological studies relating to Swedish learners of Japanese will also be included.

Chapter 4 will describe the experiments for the present study. After a quick overview, the general design of the experiments will be described, followed by accounts of the participants, materials and procedures according to the level. These are summarised as follows:

- There were four different tests in total (a reading test and a writing test on each of the novice and advanced level);
- The levels were set using the number of kanji learned by the participant groups by the time of the experiments as an indicator: 240 characters for the novice level ("LV240"), and 800 for the advanced level ("LV800");
- The participants were 49 Swedish university students and 191 Japanese second graders for LV240, and 20 Swedish university students and 135 Japanese fifth graders for LV800;
- The materials were written tests with fill-in-the-blanks questions. The blanks were to be filled in to complete the sentences, with the reading of the target kanji words for the reading tests, and with handwritten kanji for the writing tests;
- The target kanji characters were selected from the characters recently learned by both Swedish and Japanese groups, and thus two sets of target characters (one each for LV240 and LV800) were prepared.

Chapter 5 will clarify how errors collected from the experiments were classified. Comprehensive descriptions of reading and writing error classification criteria will be given, together with summaries and examples.

The results will be reported and analysed in Chapter 6. The tests were marked and the results were entered into the database. Reading errors were classified into four error types (phonological, circumstantial, orthographic and semantic), and writing errors into five types (the four aforementioned types and an additional pseudokanji type). The results and statistical analyses thereof will be presented according to skill and level: firstly LV240 reading, LV800 reading and LV240/LV800 comparison of reading; secondly LV240 writing, LV800 writing and LV240/LV800 comparison of writing; and finally a comparison of reading and writing.

Chapter 7 will present the research hypotheses formulated from the findings of the previous studies and will verify them based on the analyses

of error generating tendencies in the previous chapter. It will further discuss the differences in the participant groups' difficulties in kanji recognition and production, perception patterns and approaches to retrieval, as well as possible factors therein. An overall discussion will clarify the characteristics of the kanji learning process of the Swedish learners of Japanese. This will be summarised with the conclusion, insight and implications in Chapter 8.

Notes

[1] The current list of the Jōyō Kanji, issued in 2010 by the Japanese Ministry of Education, Culture, Sports, Science and Technology, consists of 2,136 characters. In this dissertation, however, the term refers to the list of the Jōyō Kanji issued in 1981 by the Japanese Ministry of Education unless otherwise stated, since the participants of the experiments for the present research had been educated before the 2010 list came into effect and in conformity with the 1981 list. The 1981 list of the Jōyō Kanji consists of 1,945 characters, and the current and former lists overlap to a great degree; in fact 90.8% of the characters included in the 2010 list (1,940 out of 2,136 characters) are identical with those in the 1981 list.

2. FACTS AND CONCEPTS RELATED TO KANJI ERROR CLASSIFICATION

In order to facilitate understanding of the previous studies, experimental designs and error classifications stated in subsequent chapters, this chapter will explain how the Japanese writing system works and clarify concepts relating to kanji error classification by presenting facts, theories, data and examples. Section 2.1 will give an overview of the Japanese writing system and section 2.2 a general description of kanji, which should together provide better understanding of how each orthography works, the complexity of the writing system and the difficulty of mastering it. Section 2.3 will present a cognitive model of kanji retrieval from the mental lexicon, section 2.4 will explain the error generating mechanisms, section 2.5 will discuss the problems with multiple readings, section 2.6 will review the visual aspects of intra-character components, and section 2.7 will deal with component-based analogy.

2.1. The Japanese writing system

The Japanese did not have written language until they started borrowing Chinese characters to record their oral language around the 5th century. The borrowed characters were called kanji (literally "Han character," based on an association with the Han dynasty, 206 BC – AD 220). The borrowing was not a straightforward process, because Chinese characters were created to write Chinese, an isolating language in which word forms do not change, and were therefore not quite suitable for transcribing Japanese. Japanese is an agglutinating language in which affixes are added to the stem of a word to indicate grammatical functions or changes in meaning, and Chinese characters were unfit for transcribing affixes. This linguistic difference between Japanese and Chinese facilitated the invention of a sound-based writing system to complement kanji.

In the 9th century, two sets of *kana* scripts (*hiragana* and *katakana*) were made from kanji by using the characters purely for their phonetic values and simplifying the configurations. ***Hiragana*** ("plain *kana*") was devised from the cursive calligraphic style of a whole character and used primarily by court ladies for the production of literary works. ***Katakana*** ("partial *kana*") was made from abbreviated parts of a regular style character and used by Buddhist monks and scholars to study sutras and the Chinese classics. For instance, the semantic value of the character 加 ("to add") was abandoned and used purely for the pronunciation *ka*, and the cursive style of the whole character 加 was simplified into *hiragana* か *ka*, while the left half of the character was used as *katakana* カ *ka*. Each symbol of *hiragana*

and *katakana* represents a mora, a temporal unit that resembles a syllable.[2] In modern Japanese, each set of *hiragana* and *katakana* contains 46 basic symbols, which, together with diacritics and sequential configurations for palatalised syllables, can represent all of the Japanese syllables/morae.

Romaji, which has never become part of the main stream of the Japanese writing system, was initially developed in the 16th century based on the Portuguese orthography for Catholic missionaries learning Japanese. Modern *romaji* was created in the 19th century and is generally based on the English orthography. There are two commonly used modern varieties of *romaji*: Hepburn system (ヘボン式), which largely follows the English phonology, and *Kunrei-shiki* (訓令式), which is more consistent with the moraic structures of *kana*. In this study, the Hepburn system *romaji* is used, as it provides a more accurate transcription of the actual pronunciation.

Table 1 is a total inventory of *hiragana* and *katakana* with corresponding *romaji*:

Table 1 Total inventory of hiragana, katakana and romaji

H: *hiragana*; K: *katakana*; and R: *romaji*

H	K	R	H	K	R	H	K	R	H	K	R	H	K	R
Basic symbols														
あ	ア	a	い	イ	i	う	ウ	u	え	エ	e	お	オ	o
か	カ	ka	き	キ	ki	く	ク	ku	け	ケ	ke	こ	コ	ko
さ	サ	sa	し	シ	shi	す	ス	su	せ	セ	se	そ	ソ	so
た	タ	ta	ち	チ	chi	つ	ツ	tsu	て	テ	te	と	ト	to
な	ナ	na	に	ニ	ni	ぬ	ヌ	nu	ね	ネ	ne	の	ノ	no
は	ハ	ha	ひ	ヒ	hi	ふ	フ	fu	へ	ヘ	he	ほ	ホ	ho
ま	マ	ma	み	ミ	mi	む	ム	mu	め	メ	me	も	モ	mo
や	ヤ	ya				ゆ	ユ	yu				よ	ヨ	yo
ら	ラ	ra	り	リ	ri	る	ル	ru	れ	レ	re	ろ	ロ	ro
わ	ワ	wa										を	ヲ	o
ん	ン	n												
Symbols with diacritics														
が	ガ	ga	ぎ	ギ	gi	ぐ	グ	gu	げ	ゲ	ge	ご	ゴ	go
ざ	ザ	za	じ	ジ	ji	ず	ズ	zu	ぜ	ゼ	ze	ぞ	ゾ	zo
だ	ダ	da	ぢ	ヂ	ji	づ	ヅ	zu	で	デ	de	ど	ド	do
ば	バ	ba	び	ビ	bi	ぶ	ブ	bu	べ	ベ	be	ぼ	ボ	bo
ぱ	パ	pa	ぴ	ピ	pi	ぷ	プ	pu	ぺ	ペ	pe	ぽ	ポ	po
Palatalised syllables														
きゃ	キャ	kya				きゅ	キュ	kyu				きょ	キョ	kyo
しゃ	シャ	sha				しゅ	シュ	shu				しょ	ショ	sho
ちゃ	チャ	cha				ちゅ	チュ	chu				ちょ	チョ	cho
にゃ	ニャ	nya				にゅ	ニュ	nyu				にょ	ニョ	nyo
ひゃ	ヒャ	hya				ひゅ	ヒュ	hyu				ひょ	ヒョ	hyo
みゃ	ミャ	mya				みゅ	ミュ	myu				みょ	ミョ	myo
りゃ	リャ	rya				りゅ	リュ	ryu				りょ	リョ	ryo
Palatalised syllables with diacritics														
ぎゃ	ギャ	gya				ぎゅ	ギュ	gyu				ぎょ	ギョ	gyo
じゃ	ジャ	ja				じゅ	ジュ	ju				じょ	ジョ	jo
びゃ	ビャ	bya				びゅ	ビュ	byu				びょ	ビョ	byo
ぴゃ	ピャ	pya				ぴゅ	ピュ	pyu				ぴょ	ピョ	pyo

In the current Japanese writing system, kanji is used to write the majority of content words (e.g. nouns and verb/adjective stems). *Hiragana* is used for nearly the all function words (grammatical morphemes), including particles (postpositional words for marking the functions of the preceding words within the phrase/clause), auxiliary verbs, verbal suffixes and adjectival suffixes, as well as a small number of content words. Loan words of non-Chinese origin and the majority of onomatopoeic words are written in *katakana*. *Romaji* are used either as a pronunciation guide for non-native

speakers, to give the word visual prominence and a sense of novelty, or in abbreviations such as TV and DVD. The following sentence is an example of the standard way of mixing kanji, *hiragana*, *katakana* and *romaji*:

私は TSUTAYA で古いミュージカル映画の DVD を買 った。
"I bought a DVD of an old musical film at TSUTAYA*."
(* name of a video shop chain)

Although no space is inserted between words in a sentence, phrasal boundaries are visually marked to a certain degree with content words in kanji/*katakana*/*romaji* followed by function words in *hiragana*, as indicated in the orthographic and grammatical breakdown of the same sentence below. The words transliterated in **BOLD CAPITALS** are written in kanji (KJ). Likewise, lowercase letters correspond to *hiragana* (HR), CAPITALS to *romaji* (RM), and **bold lowercase letters** to *katakana* (KT):

私は$_a$	TSUTAYA で$_b$	古い$_c$	ミュージカル$_d$	映画の$_e$	DVD を$_f$	買った$_g$
WATASHI-wa	TSUTAYA-de	**FURU**i	**myūjikaru**	**EIGA**-no	DVD-o	**KA**tta
(KJ) (HR)	(RM) (HR)	(KJ) (HR)	(KT)	(KJ) (HR)	(RM) (HR)	(KJ)(HR)
I-TOPIC	Tsutaya-LOC	old-NON PAST	musical	film-GEN	DVD-ACC	buy- PAST

The above breakdown is tabulated in Table 2 to clarify how orthographies are distributed. The grey-marked transliterations follow the same principle as above and the orthographies employed are indicated in parentheses. The content words are written in kanji/*romaji*/*katakana*: kanji is used for the pronoun *I* and the noun *film* as well as for the stems of the adjective *old* and the verb *buy*; TSUTAYA, a chain store that values visual prominence and sense of novelty, and the abbreviation *DVD* are written in *romaji*; and **myūjikaru**, which is borrowed from the English word *musical*, is in *katakana*. All the content words are followed by function words in *hiragana*, except for **myūjikaru**, which is used here attributably and unaccompanied by function words.

Table 2 Orthographical and grammatical breakdown of a Japanese sentence

Phrase		a	b	c	d	e	f	g
Content word	**Noun/ Pronoun**	私 **WATASHI** (kanji) I	TSUTAYA TSUTAYA (*romaji*) Tsutaya		ミュージカル **myūjikaru** (*katakana*) musical	映画 **EIGA** (kanji) film	DVD DVD (*romaji*) DVD	
	Verb stem							買 **KA** (kanji) buy
	Adjective stem			古 **FURU** (kanji) old				
Function word	**V/Adj suffix**			い i (*hiragana*) NON-PAST				った tta (*hiragana*) PAST
	Particle	は wa (*hiragana*) TOPIC	で de (*hiragana*) LOC			の no (*hiragana*) GEN	を o (*hiragana*) ACC	

Of the three orthographies, *romaji* is the least difficult to learn for learners with an alphabetic background, since the script is the Roman alphabet and its GSC rules are similar to and simpler than the rules for English or Swedish (the relationships between orthographies and GSC will be explained in subsection 3.1.1). Unfortunately though, *romaji* is used only supplementarily in writing Japanese. *Kana* is more challenging, for two sets of mora-based syllabaries *hiragana* and *katakana* must be learned. *Kana* is relatively easy to learn because of the high regularity of GSC and the moderate number of graphemes (two sets of 46 basic symbols). Furthermore, *kana* plays an active part in the Japanese writing system. Meaning-based kanji, which has numerous and more complex graphemes than *romaji* and *kana*, is of the greatest importance within the Japanese written vocabulary. Section 2.2 will review kanji briefly, focusing on its history and phonological opaqueness.

2.2. Overview of kanji

Chinese characters originated at least 3,000 years ago, the oldest confirmed evidence being the inscriptions on oracle bones from the late Shang dynasty (c. 1200 – 1050 BC) (Kern, 2010; Keightley, 1978). Although

these characters were created and developed to write Chinese, they were also adopted to write several other Asian languages, including Korean and Japanese.

Chinese characters are meaning-based and are often referred to as ideographic (idea-based), logographic (word-based) or morphographic (morpheme-based). The term preferred in this study is logographic because, in principle, each character represents a word.

Chinese characters are not examples of ideographs in the strict sense. An ideograph is a graphic symbol representing a concept without indicating words/phrases of any particular language, such as pictographic signs (e.g. the green and white sign depicting a man running through a door, which denotes the location of the closest emergency exit), whereas a Chinese character bears meaning and is bound to particular speech sounds. For example, *hanzi* (Chinese characters used in China) are bound to Chinese speech sounds, and kanji to Japanese speech sounds.

Morphograph is not an ideal term to refer to Chinese characters, either. A morphograph is a character representing a morpheme (the smallest meaningful unit of a language), but a majority of Chinese characters can be broken down into smaller meaningful units, i.e. a character is often a combination of two or more morphemes. There are characters that represent monomorphemic words, such as 人 (“person”) and 木 (“tree”). These morphemes/words can act as components of multimorphemic characters. The character 休 (“rest”) is one such character; it is a combination of 亻 (the componential form of 人) and 木 and depicts a person resting against a tree. This bimorphemic character 休 can be combined with another character to form a two-character compound, e.g. 休 (“rest”) + 日 (“day”) = 休日 (“holiday” or “day off”). Since a majority of characters can be broken down in this way to intra-character morphemic components, and many such morphemic components can form words on their own, a character is deemed to represent a word, and therefore the kanji is best referred to as logographic.

Although Japan had had limited contact with Chinese characters since the 1st century in the form of inscriptions on objects such as coins and bronze mirrors, the Japanese language had been an oral language until around the 5th century, when the borrowing started. Kanji characters were actively imported from China into Japan over a long span of time, between the 5th and 17th centuries. In the beginning, the characters and their pronunciation, based on the Southern dialect of Chinese, were gradually imported via Korea, concurrently with the introduction of Buddhism. Between the 7th and 9th centuries, Japanese envoys and students in China made extensive and systematic importation of characters based on the pronunciation of the

Central dialect. Subsequent importation was sporadic and the pronunciation was based on the Northern dialect, together with the importation of new concepts and goods by monks and merchants.

The gradual and systematic adoption of kanji over this long period of time made kanji words essential constituents of the written Japanese vocabulary. The prolonged adoption process has also affected the pronunciation of kanji (traditionally called *yomi* 読み or "**reading,**" in the sense of speaking aloud written symbols) in many ways. Each *hanzi* corresponds, in principle, to a single Chinese syllable, although there are many homophonous characters. During the adaptation to Japanese, however, a majority of kanji characters have developed multiple readings, i.e. one character may correspond to more than one reading.

Chinese-based reading is called *on-yomi* (音読み) or "***on*-reading.**" It literally means "sound reading", indicating that it is based on the original *hanzi* pronunciation. Since the reading of kanji was imported from different Chinese dialects over a period of a thousand years, the geographical and historical variations of *hanzi* pronunciations have given many kanji characters more *on*-readings than one. For example, the character 行, which means "to go", has three *on*-readings: the Southern dialect based *gyō* as in *gyō-retsu* 行列 ("procession"), the Central dialect based *kō* as in *kō-shin* 行進 ("parade"), and the Northern dialect based *an* as in *an-gya* 行脚 ("walking tour as pilgrimage"). To complicate matters, the phonological Japanisation in the adaptation process has made the already plentiful homophonous characters in the original *hanzi* pronunciation extremely abundant in kanji reading, even for characters that had been differentiated in *hanzi*. For instance, *hanzi* 三 *sān* ("three") and 山 *shān* ("mountain") have become homophonous in kanji (both 三 and 山 have the same *on*-reading *san*), due to the relative simplicity of the Japanese phonotactics.

Furthermore, native Japanese words have been assigned to a majority of kanji according to their meanings, as additional readings of the character. These Japanese-based readings are called *kun-yomi* (訓読み) or ***kun*-reading**s. This literally means "interpretation reading", because it is the interpretation of the character's meaning into native Japanese. For example, the character 山 ("mountain"), which was first adopted with the *on*-reading *san*, can also be read as *yama* ("mountain" in native Japanese) in *kun*-reading. Since a kanji character is a logograph, it has both semantic and phonetic values and can be represented with an equation <character> = {*meaning* : sound} (Iwasaki, 2013). The following is a representation of these values for the *on*- and *kun*-readings of the character 山:

on-reading 山 = { *mountain* : /san/ }
kun-reading 山 = { *mountain* : /yama/ }

As seen above, the *kun*-reading merely replaces the phonetic value of the *on*-reading, leaving the semantic value unchanged. In general, *on*-reading is applied to compounds and *kun*-reading to non-compounds, although there are quite a few exceptions (i.e. compounds of *kun*-reading and single-character words of *on*-reading).

The complication of multiple readings can be exemplified with the characters 三 and 脚. The character 三, meaning "three", has *san* as its *on*-reading and *mi* and *mitsu* as its *kun*-readings, whereas the character 脚 ("leg") has a variety of *on*-readings *kyaku*, *kaku*, *kya* and *gya*, as well as a *kun*-reading *ashi*. These characters can be combined to form the compound 三脚 ("tripod"), which is read as *san-kyaku*, applying the *on*-reading of each constituent character, while the same combination of characters can be read as *mitsu-ashi*, combining the *kun*-reading of each character, in which case the compound means "three legs" or "three-legged." [3]

Words with *on*-readings such as *sankyaku* are called *kango* (漢語) or Sino-Japanese words, while *kun*-reading words are called *wago* (和語) or native Japanese words. The relationship between Sino-Japanese words and native Japanese words in the Japanese vocabulary can be compared to words of Latin/Greek origins and Anglo-Saxon words in the English vocabulary; the Sino-Japanese word *sankyaku* corresponds to the English word "tripod" (*tri* "three" + *pod* "foot" via Latin *tripodis* from Greek *tripous*), and the native Japanese *mitsuashi* to "three-legged" of Anglo-Saxon origin. Sino-Japanese words tend to be used in technical terms and formal expressions, while native Japanese words are often found among basic words and everyday language, analogously to the contrast between words of Latin/Greek origin and those of Anglo-Saxon origin in English. This analogy, however, cannot be extended to the aspect of phonological transparency. The English pair "tripod" and "three-legged", sharing a combination of the concepts "three" and "foot/leg", are not spelled alike and are pronounced as they are spelled (i.e. differently from each other), whereas the corresponding Japanese pair 三脚 *sankyaku* and 三脚 *mitsuashi* are visually identical but are read (pronounced) differently, making the written form 三脚 phonologically opaque.

Although a majority of kanji characters have multiple readings, compounds with multiple readings such as 三脚 (*sankyaku*/*mitsuashi*) are rather uncommon. A great majority of words written in kanji, either compounds or single-character words, have only one correct way of reading them, and application of another reading of the component character is deemed incorrect. For example, 行列 ("procession") is always read as *gyō-retsu* and reading it as *kō-retsu*, misapplying another reading of the character 行, is incorrect.

The problems discussed above pertain to the reading of written characters and compounds, but phonological opaqueness can constitute a problem in writing as well. When only a pronunciation is presented as *sansai* (さんさい in *hiragana*), it is not clear if it should be written as 三歳 *sansai* ("three years of age") combining characters 三 *san* ("three") and 歳 *sai* ("age"), or 山菜 *sansai* ("edible wild plant") combining 山 *san* ("mountain") and 菜 *sai* ("vegetable"), unless either a context or a properly accented speech sound is provided as a determiner of the meaning.

As reviewed above, phonological opaqueness is one of the causes of difficulty in learning kanji. Subsequent sections will discuss this aspect more closely, as well as the numerousness and configurational complexity of characters that make learning kanji a challenge.

2.3. Cognitive model of kanji retrieval

This section will present a cognitive model of kanji to explain the three basic elements of kanji, the concept of the mental kanji lexicon, and the mental activities involved in the retrieval of kanji.

It is a widely acknowledged concept that cognition of a kanji character is supported by knowledge of other related kanji characters. In this regard, Saito (2006) explains that, in an encounter with a character that is yet to be mastered, the dynamics of the kanji cognition system summon knowledge and awareness to process the character in question, and that the cognition system utilises this insufficient (partial) information to reconstruct the sufficient (whole) information to speculate the whole image of the character.

Partial information can be a visual representation of a character, in which case the phonological and semantic representations must be reconstructed. In a writing task, visual representation must be reconstructed based on phonological and/or semantic representation(s). A kanji reading or writing error occurs when the reconstruction attempt fails and a wrong guess is given.

In conformity with many of the previous studies in the field, this study shares this cognitive perspective of error making and uses error classification in accordance with cognitive portrayal of the mental kanji lexicon, which is suitable for descriptions of learners' retrieval methodologies and difficulties.

Kanji is a writing system represented by characters consisting of the three elements 形 *kei* "form," 音 *on* "sound" and 義 *gi* "meaning", and processing of these three elements is interlocked and coordinated (e.g. Saito et al. 1998, 1999; Chikamatsu, 2005; Saito 2006). For example, a character depicted as 花 (the "form" element) is pronounced *hana* in *kun*-reading and *ka* in *on*-reading (the "sound" element), and stands for "flower" (the "meaning" element). When learned, a kanji is stored in these three element-based categories in the learner's mental lexicon. As the number of acquired characters increases, the three elements of each character are stored, sorted and intertwined with the elements of other characters in each category, forming a kind of association network. When a learner/user reads or writes kanji, knowledge of the characters in question is retrieved from the mental lexicon. After acquiring multiple characters, retrieval of knowledge of a particular character from the mental lexicon often entails conscious/unconscious recollection of other characters that are orthographically similar, phonologically identical and/or semantically related. Such an association network is illustrated in Figure 1 as a simplified cognitive model for the character 花 after acquiring approximately 500 characters, partially based on the diagrammatic concept of Hatta et al. (1998):

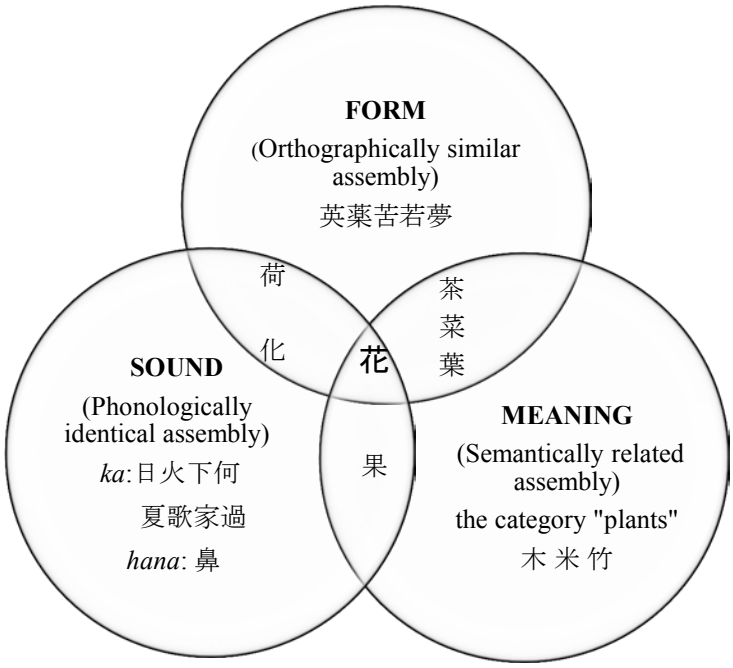

Figure 1 Simplified cognitive model of the character 花 in the mental kanji lexicon

The "Form" circle in the above model is an assembly of orthographically similar characters that typically share a radical (an orthographic component

of a character used for classification of kanji/*hanzi*) of the character 花, namely, ⺾ or 化. Although a shared radical is not a necessary and sufficient condition for orthographic similarity, and a stricter criterion of similarity judgement was used for classification in this study, as described later in subsection 5.3.2, it is used here in order to simplify the model. When a total of approximately 500 characters have been acquired in a normal order of learning (simple, basic and frequently used characters first), this circle is most likely to be filled with characters such as 英, 薬, 苦, 若, 夢, 茶, 菜, 葉, 荷 and 化.

Activation of the Form-Sound-Meaning interlink within a single character has ripple effects on other characters' interlinks (Saito et al., 2000; Saito, 2006). When a learner/user comes across and recognises or tries to recall the form 花, it is often the case that the memory of forms of these orthographically similar characters is also activated.

- Likewise, at around the 500 character-level, the "Sound" assembly would typically be filled with phonologically identical (= homophonous) characters sharing the *on*-reading *ka* (日, 火, 下, 何, 夏, 歌, 家, 過, 荷, 化, and 果) and the *kun*-reading *hana* (鼻). The "Meaning" circle is a semantically related assembly which would include characters denoting plants, such as 木 "tree," 米 "rice," 竹 "bamboo," 茶 "tea," 菜 "vegetable," 葉 "leaf" and 果 "fruit" (again, for the purpose of simplification, metaphoric meanings of the character 花 ("flower") such as "beauty" and "one's prime" are disregarded here). In addition to this classificatory grouping, synonymous and antonymous characters, if any, would also be placed in this assembly.

  Although there can be numerous homophonous characters, orthographically identical (= homographic) characters can rarely be found. Each and every character takes, in principle, a different form, and there are an infinitesimal number of exceptions (e.g. 芸 "art"/"cutting grass", 灯 "lamplight"/"intense fire" and 浜 "beach"/"creek" are the few example characters with homographic variants, which were created by simplifying orthographically complicated characters 藝 "art", 燈 "lamplight" and 濱 "beach" into the same configuration as existing characters with simpler configuration and other meanings 芸 "cutting grass", 灯 "intense fire" and 浜 "creek"). Since kanji is meaning-based and one distinguishable graphic representation is assigned to each meaning, it is not an overstatement to say that there are no semantically identical characters, except for geographical and historical style variations for the same character (e.g. "ten thousand" can be depicted as 萬 in the traditional style and 万 in the simplified style, both of which are pronounced *man*).

There are a few characters in the overlapping areas of the circles: 荷 and 化 are orthographically similar to and phonologically identical with the central character 花; 茶 and 葉 are orthographically similar and semantically related; and 果 is phonologically identical and semantically related.

As the total number of acquired characters increases, more characters will be included in each assembly. A connection model similar to this can be formed for each character in the mental lexicon and each model has multiple connections with the other models, forming an extensive and intricate cognitive network of kanji knowledge.

2.4. Error generating mechanisms

This section will provide explanations for mechanisms of kanji writing errors.

Hatta et al. (1998) hypothesised that different types of writing errors are the results of the malfunctioning of different mechanisms of kanji recognition (see subsection 3.3.3). When this hypothesis is applied, a writing error such as writing 化 in an attempt to write the character 花 can be explained as follows: the attempt to retrieve 花 itself from the mental lexicon fails for some reason, but the orthographic and phonological assemblies of its cognitive model are activated, and the orthographically similar and phonologically identical 化 is erroneously retrieved. Due to the inactivation of the semantic assembly, however, the absence of 化 from the semantic assembly (化 covers the basic concept of "change" and has no detectable semantic connection with flowers or plants) is not registered. In other words, semantic screening malfunctions, and as a result, the character 化 is written in lieu of 花. Therefore, this error can be classified as a combination of orthographic and phonological types.

In addition to the hypothesis of Hatta et al. (1998) that the occurrence of this 花 → 化 type of writing error is caused by activation of the orthographic and phonological assemblies and inactivation of the semantic assembly, this study further hypothesises that this type of error indicates orthographically and phonologically inclined retrieval approaches and deficiency in the semantic screening function. As a number of previous studies have pointed out (to be described in Chapter 3), there are predominant strategies and notable difficulties that each learner/user group has, such as recognition via orthographic decoding and weakness in processing abstract words (Tamaoka, 1989; Komori, 2009). In the mental lexicon of those who prefer orthographic decoding, the ripple of activation in search of the target character tends to be oriented towards the

orthographic assembly. Semantic inactivation is to be regarded as a malfunction of the semantic screening function. The candidate character (化), picked from the activated orthographic-phonological assembly, should be disqualified if the meaning does not match that of the target character's meaning ("flower"), but the semantic screening fails for some reason, possibly due to insufficient knowledge or temporary negligence. The choice of retrieval method may be unconscious or strategic, either due to L1 transfer or after explicit instructions that it is a "good bet" for a fortuitous success. In this hypothesis, if the intended 花 is erroneously replaced with the semantically related but orthographically and phonologically dissimilar character 竹 (with *on*-reading *chiku*, *kun*-reading *take*, and meaning "bamboo"), it can be interpreted as a reflection of a semantically inclined retrieval approach and as an orthographic and phonological screening deficiency.

The cases discussed above involve the substitution of the target character with another existing character with a certain similarity or association, but the presumptions regarding retrieval approach and screening deficiency in relation to existing characters can also be applied to cases in which pseudokanji is written in place of the target character. A pseudokanji is a non-existing kanji character deviating from any existing character (with additional or missing strokes; additional, missing, replaced or switched components; disproportionate assemblage of components, etc.). Depending on the degree of deviation, a pseudokanji can be a wrong combination of existing components, or may include a non-existing component. It can then be presumed that the latter is a manifestation of a weaker orthographic screening function than the former.

For example, 亠 can be a kanji component, but not △. Therefore, a pseudokanji which is a combination of ⺾ and △ represents a weaker orthographic screening function than another erroneous combination ⺾ and 亠, since, in the former, the screening function failed not only on a whole-character basis (to have written a pseudokanji) but also on a component basis (to have used a non-existing component). This deficiency of orthographic screening function can be interpreted as an embodiment of an underdeveloped configurational awareness. Learners with underdeveloped configurational awareness are uncertain of what kind of lines and dots can be kanji strokes, which configuration of strokes can form a kanji component, and which assemblage of components can constitute a kanji character.

2.5. Problems with multiple readings

This section will describe problems with multiple readings, providing facts and examples.

Error type occurrence patterns can be affected not only by the intrinsic features of the character but also by extrinsic factors. The readings of the character (the sound element of kanji) are one of the intrinsic features of kanji, but if a character has multiple readings, it is the extrinsic factors that decide which one of the readings is to be applied. In fact, only 34% of the Jōyō Kanji characters have one-to-one grapheme-sound-correspondence, and the rest of the characters have multiple readings (Nomura, 1984). Which of the multiple readings should be applied is determined by the word involving the character. Single-character words often take the *kun*-reading and compound words the *on*-reading, but this tendency hardly serves as a reliable guide for reading application, as the example of reading distribution of the character 花 ("flower") indicates in Table 3. The hyphen in the reading (italicised) indicates the break between the characters within the word, and the bold type the reading of the character 花:

Table 3 Reading distribution of the character 花 ("flower")

Word type	**Single-character**	**Compound**	**Compound**
Character position		Word-initial	Non-initial
***On*-reading** *ka*	-----	花器 ***ka-ki*** "flower vase"	造花 *zō-**ka*** "artificial flower"
***Kun*-reading** *hana* (alternatively *-bana*)	花 ***hana*** "flower"	花屋 ***hana-**ya*** "flower shop"	生け花 *i-ke-**bana*** "flower arrangement"

As shown in Table 3, the character 花 in the example words has the same meaning "flower," but different readings are applied depending on the word. The position of the character within the word does not influence the reading, except in the cases of *rendaku* (連濁) or sequential voicing, as in the *kun*-reading of non-initial position ***ike-bana*** **(花** *hana* → *bana***)**.[4]

In some cases, the word meaning does affect the reading, since it is often the case that a polysemic character has different *kun*-readings assigned to each of its meanings. For example, a single-character word 生 is pronounced *sei* in *on*-reading when it has the prototypical meaning of "life," and different *kun*-readings (translations into native Japanese) are applied to its derivative meanings (e.g. 生: *life* → *alive* → *fresh* → *raw* → *untreated* → *undiluted* → *pure*), because these notions are expressed with different native Japanese words (*nama* = "raw, fresh" *ki* = "pure, undiluted"). Polysemic kanji compounds, too, can have multiple readings according to the meaning. The two-character compound 水色, consisting

of 水 (“water”) and 色 (“colour”), literally means “the colour of the water” when the *on*-reading *sui-shoku* is applied, while the same compound means “light blue” (conventionally perceived as the colour of water) when the *kun*-reading *mizu-iro* is applied. In the case of the compound 声明 (声 “voice” + 明 “clear”), *shō-myō*, which is the southern dialect-based *on*-reading, imported from China together with Buddhism, is a Buddhist term for “chant” (rhythmic religious phrases sung in unison in a “clear voice”), whereas the central-dialect based *on*-reading, brought in by envoys and students, means “official statement,” or a “clear voice” of the authorities on a subject.

The abovementioned examples of meaning-based reading applications are listed in Table 4:

Table 4 Examples of meaning-based reading applications

Word type	Single-character	Compound	
***On*-reading**	生 ***sei*** “life, living”	水色 ***sui-shoku*** “the colour of the water”	声明 ***shō-myō*** “Buddhist chant”
			声明 ***sei-mei*** “(official) statement”
***Kun*-reading**	生 ***nama*** “raw, fresh”	水色 ***mizu-iro*** “light blue”	
	生 ***ki*** “pure, undiluted”		

When presented in a written form, the words listed above therefore need context in order to determine the meaning, which, in turn, specifies the reading, as follows:

a. 生 の 喜び
sei *no* *yoroko-bi*
life GEN joy
“joy of **life**”

b. 魚 を 生 で 食べる
sakana *o* ***nama*** *de* *ta-beru*
fish ACC **raw** INS eat
“eat fish raw”

c. ウイスキー を 生 で 飲む
uisukii *o* ***ki*** *de* *nomu*
whiskey ACC **undiluted** INS drink
“drink whiskey straight”

d. ダージリン紅茶 の 明るい オレンジ の 水色
dājirin-kōcha no aka-rui orenji no ***sui-shoku***
Darjeeling tea GEN bright orange GEN **colour of water**
"the bright orange colour of (the water of the brewed) Darjeeling tea"

e. 水色 の シャツ
mizu-iro *no shatsu*
light blue GEN shirt
"a light blue shirt"

f. 政府 が 声明 を 発表した
seifu ga ***sei-mei*** *o happyō-shita*
government NOM **statement** ACC issue-PAST
"the government issued a statement"

g. 声明 が 寺中 に 響いた
shō-myō *ga tera-jū ni hibiita*
Buddhist chant NOM temple-around LOC resonate-PAST
"Buddhist chant sounded through the temple"

For words with inflections, kanji is used to transcribe the stem (the root of the word), and the inflectional endings are written in *hiragana,* which is called *okurigana*. In some cases, inflectional endings can be an indicator of which reading should be applied. For example, the polysemic character 行 ("to go, to carry out") takes −く (*-ku*) as *okurigana* and is read as ***i-ku*** when it means "to go", but when it means "to carry out", it takes the *okurigana* −う (*-u*) and is pronounced ***okona-u***. It is therefore clear from the *okurigana* (*-ku* or *-u*) what the character means and how it should be pronounced, as exemplified below:

(1) 行く
i-ku
"to go"

(2) 行う
okona-u
"to carry out"

In the past tense forms, however, they become visually identical with each other; both take the inflectional ending "*-tta*" and are spelled 行った, as (1)' and (2)' below:

(1)' 行った
i-tta
go-PAST
"went"

(2)' 行った
okona-tta
carry out-PAST
"carried out"

Even in such a case, a context that specifies the meaning of the word can clarify the appropriate reading, as exemplified in (1)" and (2)" below:

(1)”	大学	に	行った	(2)”	実験	を	行った
	daigaku	*ni*	***i**tta*		*jikken*	*o*	***okona**tta*
	university	to	**go**-PAST		experiment	ACC	**carry out**-PAST
	“went to the university”				“carried out an experiment”		

There is a type of reading error in which an inappropriate choice of the character’s multiple readings is applied. For instance, both *hana* and *ka* are two of the legitimate readings of the character 花 (“flower”), but it is a mistake to read 花器 (“vase”) as ***hana**-ki* instead of the correct ***ka**-ki*, or 行った in 実験を行った (“carried out an experiment”) as ***i**tta* (“went”) instead of the correct ***okona**tta* (“carried out”). This type of error implies that the learner has at least been able to recognise the character orthographically but was unsuccessful in phonologically identifying it, for the following reasons:

i. insufficient phonological knowledge of the character (the learner knows only one of the readings of the character);

ii. limited metalinguistic awareness (the learner knows the multiple readings of the character, but is ignorant of extra-character factors such as word formation, meaning or the context, which function as reading determinants); or

iii. poor association of multiple information of the character (the learner has sufficient phonological knowledge and metalinguistic awareness, but fails to connect them).

On the other hand, there is a type of error that makes sense of the given circumstances, taking into account the inflectional endings (*okurigana*), the other (non-target) compound constituent character, and the context. In such errors, the erroneous reading applied to the target character can either form another word that shares the inflectional ending or compound constituent, or makes sense in the given phrase/sentence.

An English example of this would be a case in which the first five letters of the target word *grateful* were somehow illegible (*?????ful*) and the illegible part was filled with the morpheme *cheer*, on the basis of the *-ful* ending, to complete the word as *cheerful*, in an attempt to find a morphologically appropriate candidate. If the target word is embedded in the phrase *my ?????ful friend*, *cheerful* makes a contextually appropriate candidate as well.

In the case of kanji, the phonologically opaque grapheme is equivalent to the illegible letters, and the inflectional endings written in *hiragana*, the

non-target compound constituent whose pronunciation is known, or the context, can be the legible clue. Even though the substitute candidate is morphologically and/or contextually appropriate, however, it is counted as an error if it is not any of the legitimate readings assigned to the target character. Table 5 is a list of examples of kanji errors falling into this category, presented according to the circumstantial clue types:

Table 5 Examples of kanji errors based on circumstantial clues

Circumstan -tial clue	Target word	Erroneous answer	Shared element
a. *Okurigana* (verbal)	解く ***to-ku*** solve-NONPAST (“solve”)	書く ***ka-ku*** write-NONPAST (“write”)	–く *–ku* Nonpast verb ending
b. *Okurigana* (adjectival)	美しい ***utsuku-shii*** (“beautiful”)	新しい ***atara-shii*** (“new”)	–しい *–shii* Nonpast adjective ending
c. Compound constituent	親切 ***shin-setsu*** familiar-sincere (“kind”)	大切 ***tai-setsu*** big-sincere (“precious”)	–切 *-setsu* -sincere
d. Context	心　　が　広い ***kokoro ga hiro-i*** heart NOM spacious-NONPAST (“the **heart** is big”)	部屋が　広い ***heya ga hiro-i*** room NOM spacious-NONPAST (“the **room** is big”)	が　広い *ga hiro-i* NOM spacious-NONPAST (“the **…..** is big”)

The contextual example in Table 5 can be argued to be an example with a syntactic-contextual clue: in addition to the contextual clue, the nominative particle が *ga* determines the preceding word to be a subject, which in turn determines the word to be a noun or a noun phrase.

This type of error can be explained in terms of the learner being incapable of recollecting the phonological knowledge of the character in question and making educated guesses based on morphological, syntactic, lexical or contextual clues, although these guesses have turned out to be wrong. Alternatively, learners may have carelessly mistaken the reading of the character and their mental lexicon did not function properly to screen the error, because of its mock consistency on the grammatical, lexical and/or contextual level.

2.6. Visual aspects of intra-character components

This section will describe the visual aspects of kanji's intra-character components and present hypotheses relating to learners' decompositional ability and functional awareness of kanji components.

Although visual representation of a word in the alphabetic writing system is a linear formation of phonemic symbols, each kanji character, which is a logographic script, is a collection of strokes in a square configuration. The number of strokes used in the configuration is a reliable indicator of the visual complexity of the character. Among the 1,945 Jōyō Kanji characters, the number of strokes used to compose a character ranges from one (一 meaning "one") to twenty-three (鑑 meaning "appreciate/appraise/judge"), which demonstrates the varying visual complexity of these characters. Simple characters such as 木 ("tree") and 山 ("mountain") are monomorphemic and have a single structure, whereas complicated characters are multimorphemic and can be divided into two or more morphemes, and each morpheme is represented by an intra-character component. For example, 林 ("woods") is a left-and-right combination of two trees (木), and 岩 ("rock") is a top-and-bottom combination of the two morphemes 山 ("mountain") and 石 ("stone"). Semantically and functionally significant intra-character components are called radicals, according to which characters are classified in dictionaries.

Each radical is morphemic and therefore contains semantic value, and when two radicals are combined to form a character, the combined morphemes can create a further semantic value. For instance, the character 林 is a semantic composite, a combination of two semantic radicals 木 ("tree") denoting "gathering of trees" or "woods." In this case, the radicals' *on*-reading *boku* or *moku* has nothing to do with the character's *on*-reading *rin*. Alternatively, one of the two radicals can specify the character's semantic category and the other can function as an indicator of the *on*-reading of the character. For example, the character 銅 *dō* ("copper") is a phono-semantic composite, or a combination of the semantic radical 金 ("metal") on the left and the phonetic radical 同 ("same") on the right. The character 銅 shares the *on*-reading of the phonetic radical 同 *dō*, while the semantic radical 金 merely functions as a category indicator and its *on*-reading *kin* has no phonological influence on the whole character 銅.

There are seven basic positions for radicals, namely, (1) left (偏 *hen* "side"), (2) right (旁 *tsukuri* "building"), (3) top (冠 *kammuri* "crown"), (4) bottom (脚 *ashi* "leg"), (5) top-left (垂 *tare* "hanging"), (6) left-bottom (繞 *nyō* "entering"), and (7) exterior (構 *kamae* "structure"), as illustrated in Figure 2:

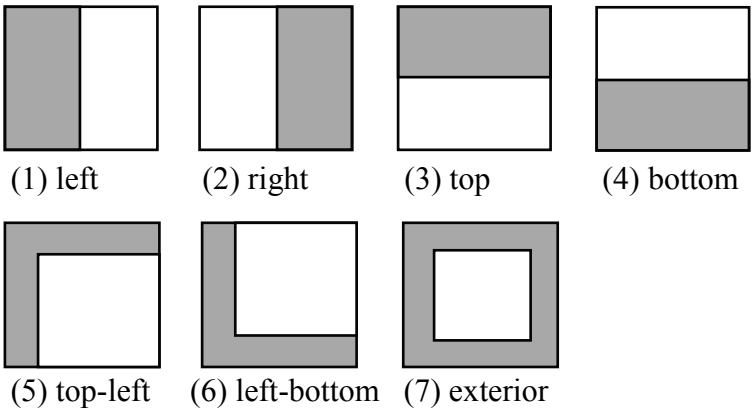

Figure 2 Basic radical positions

These positional constraints affect orthographic similarity. As shown in Table 6, when the character 枯 ("wither") is compared with the four characters 沽 ("sell/buy"), 苦 ("bitter"), 居 ("settle") and 固 ("hard"), which all consist of the radical 古 and another radical comprising three or four strokes, 沽 ("sell/buy") bears the highest graphic similarity to 枯 ("wither") because it shares the target character's left & right assembly pattern.

Table 6 Radical assembly patterns and character examples

left & right	top & bottom	top-left & the rest	left-bottom & the rest	exterior & interior
枯 "wither" (木+古)	苦 "bitter" (艹+古)	居 ("settle") (尸+古)		固 "hard" (口+古)
沽 "sell/buy" (氵+古)				

There is evidence that radicals function as lexical units in kanji processing. Studies have proven semantic radicals' interference in a character categorisation task (Flores d'Arcais, 1992) and their utilisation in the identification of word meaning (Flores d'Arcais & Saito, 1993; Flores d'Arcais et al., 1995). They have also been shown to affect the activation of phonetic radicals in lexical retrieval (Saito et al., 1998), especially for grapheme-sound correspondences in unfamiliar character recognition (Hue, 1992; Leong & Tamaoka, 1995; Seidenberg, 1985).

2.7. Component-based analogy

This section will explain component-based analogy as a type of reading error, providing the background data and examples.

Apart from the positional assembly patterns discussed in section 2.6, there is another traditional way of classifying Chinese characters according to their original formation, namely, pictographs (象形 *shōkei*), diagrammatic characters (指事 *shiji*), semantic composites (会意 *kaii*) and phono-semantic composites (形声 *keisei*) (Saito et al. 1979; Kess & Miyamoto, 1999).[5] Among these, the phono-semantic composites (PSCs) which were created by combining a phonetic radical (音符) with a semantic radical (意符) are the most common type. In fact, over 80% of kanji characters in use today are said to be of this type (Banno et al., 2009), and among the 1,945 Jōyō Kanji characters, 1,286 (66%) are PSCs (Nomura, 1984).

The *on*-reading of a PSC is often identical with that of its phonetic radical. In fact, 57.6% of these 1,286 PSCs are completely consistent with their phonetic radical in terms of *on*-reading, 32.7% are partially consistent, and 9.7% are inconsistent (Nomura, 1984). Table 7 presents examples of PSCs and their phonetic radicals (both with readings) according to consistency, as well as the percentage of each consistency type within the 1,286 Jōyō Kanji PSCs:

Table 7 Jōyō Kanji PSC character-radical on-reading consistency (Nomura, 1984)

	Completely consistent				Partially consistent				Inconsistent		
Percentage in Jōyō Kanji PSC	57.6%				32.7%				9.7%		
PSC examples	理 *ri*	判 *han*	草 *sō*	貨 *ka*	海 *kai*	苦 *ku*	防 *bō*	除 *jo*	肺 *hai*	針 *shin*	読 *doku*
Phonetic radicals	里 *ri*	半 *han*	早 *sō*	化 *ka*	毎 *mai*	古 *ko*	方 *hō*	余 *yo*	市 *shi*	十 *jū*	売 *bai*

Knowledge of the *on*-reading of the phonetic radical, therefore, can provide a reading clue for an unknown PSC containing such a radical. However, this method proves successful only if the character-radical relationship is of the "completely consistent" type. Besides, character-radical *on*-reading consistency is not specific to the radical, i.e. the same radical can be of different consistency types. For example, the radical 反 *han* is completely consistent with the character 版 *han*, partially consistent with 返 *hen* and inconsistent with 仮 *ka*. Such character-radical *on*-reading inconsistency is often a result of official simplification of complicated characters, and 仮 *ka* is in fact an example of this. The 9-stroke phonological radical 叚 *ka* on

the right side of the original character 假 *ka* was replaced with the 4-stroke radical 反 *han* for the sake of orthographic simplification, and therefore the complete *on*-reading consistency between the original character and the radical (假 *ka* vs. 叚 *ka*) has been lost in the simplified version (仮 *ka* vs. 反 *han*).

As stated later, in Chapter 4, there are 90 target characters for LV240, and 24 out of those 90 characters are PSCs. With regard to character-radical *on*-reading consistency, 11 of the 24 PSCs are completely consistent, 3 are partially consistent and 10 are inconsistent. In the reading test, seeking a reading clue in the phonetic radical assuming complete consistency for the partially consistent or inconsistent ones would result in an erroneous answer: e.g. the partially consistent 海 (*kai*) will be misread as *mai*, applying the *on*-reading of its radical 毎 (*mai*), while the inconsistent 体 (*tai*) will be misread as *hon*, based on the *on*-reading of its radical 本 (*hon*).

In order to make such an assumption, however, a learner must be able to (i) decompose the character into components, (ii) recognise the pronounceable component and (iii) know the pronunciation of this component. Therefore, a PSC that can induce a component-based analogy error must have pronunciation that is not completely consistent with its pronounceable component, and this component and its pronunciation must already have been learned. For example, it is possible for a second grader to erroneously read 体 (*tai*) as *hon* (based on the *on*-reading of 本), because the character 本 and its reading (*hon*) have been learned in the first grade. On the other hand, misreading 話 (*wa*) as *zetsu* based on the *on*-reading of 舌 (*zetsu*) is not normally possible for a second grader, since the character 舌 (*zetsu*) and its readings are not learned until the fifth grade. Therefore, in analysing component-based errors, not only the consistency of the *on*-readings of the radical and the character, but also which characters have been learned earlier, should be taken into consideration.

Another type of component-based analogy is mistaking a semantic composite (SC) character for a PSC character. For example, the character 規 is a SC (a combination of two semantic radicals 夫 and 見) and is read as *ki*. Some learners, however, may read this character as *ken*, erroneously assuming it to be a PSC with complete radical-character *on*-reading consistency, and misapplying the *on*-reading of the right-hand radical 見.

Notes

[2] A mora in Japanese phonotactics is a phonological unit conceived to be temporally constant. It consists of /V/ (a single vowel), /CV/ (a vowel preceded by a consonant), /CyV/ (a vowel preceded by a palatalised consonant), /N/ (a moraic nasal), or /Q/ (a

moraic obstruent) coinciding with non-nasal geminate consonants. A mora is represented by a single *kana* symbol with the exception of /CyV/, which is transcribed with a sequence of two symbols (a normal-sized symbol followed by a reduced-sized symbol や (ヤ) *ya*, ゆ (ユ) *yu* or よ (ヨ) *yo*), such as きゃ (キャ) *kya* (non-parenthetical *kana* symbols are *hiragana*, parenthetical ones are *katakana*). The moraic obstruent /Q/ is represented by a reduced-sized symbol っ (ッ) *tsu*. In general, a syllable corresponds to a mora and a mora is represented by a kana symbol. For example, the bisyllabic word *neko* ("cat") has two morae *ne* and *ko*, and is spelled with two *kana* symbols: ねこ in *hiragana,* ネコ in *katakana*. When the mora /CyV/, /N/, /Q/ or long vowel /H/ are involved, however, the numbers of syllables, morae and *kana* symbols are not consistent with each other. A good example including all these irregular morae is the Japanese word for "one week", which has three syllables *is-shū-kan* and six morae *i-s-shu-u-ka-n* (/V-Q-CyV-H-CV-N/), and is written with seven symbols as いっしゅうかん.

[3] Readings of kanji are usually spelled without hyphens. The hyphens in the readings *san-kyaku* and *mitsu-ashi* are inserted here to indicate the break between the two characters 三 and 脚.

[4] *Rendaku* (sequential voicing) is a morphophonological phenomenon in which the initial voiceless consonant of the non-initial morpheme of a compound becomes voiced under certain conditions (Vance, 1987; Tsujimura, 1996; Iwasaki, 2013). The following are a few examples of sequential voicing:

k → g	*maru*	+ ***k**ao*	→ *maru-**g**ao*
	"circle"	"face"	"round face"
s → z	*maki*	+ ***s**ushi*	→ *maki-**z**ushi*
	"roll"	"sushi"	"rolled sushi"
t → d	*ude*	+ ***t**okei*	→ *ude-**d**okei*
	"arm"	"clock"	"wristwatch"
h → b	*ike(-ru)*	+ ***h**ana*	→ *ike-**b**ana*
	"arrange"	"flower"	"flower arrangement"

[5] In addition to these four formation groupings, two groupings of different use of the characters, "derivative cognates (転注 *tenchū*)" and "phonetic loan characters (仮借 *kasha*)", are often added to the classification and are collectively referred to as " six writings" (六書 *rikusho*).

3. PREVIOUS STUDIES

This chapter will review previous studies in the relevant fields. Section 3.1 will take up theories of writing system studies, section 3.2 will deal with kanji recognition and processing, and section 3.3 kanji production. Section 3.4 will be a brief review of phonological studies relating to Swedish learners of Japanese.

3.1. Theories of writing system studies

This section will discuss the following three aspects of writing system studies, namely, writing system and orthography in subsection 3.1.1, to confirm the key terms and basic concepts necessary to analyse the complexity of the Japanese writing system, differences across writing systems in 3.1.2, and error analysis in 3.1.3.

3.1.1. Writing system and orthography

In this subsection, definitions and notions concerning writing systems and orthographies will be reviewed in order to better understand the complexity of the Japanese writing system as stated in Chapter 2, as well as the difference between (i) Japanese and alphabetic writing systems and (ii) the kanji and Swedish orthographies.

A writing system is a system in which written symbols represent certain aspects (e.g. sounds, meanings, phonemes, syllables or words) of a language. Orthography is "the rules for using a script in a particular language (e.g. the English or Italian orthography for the Roman alphabet)" (Cook & Bassetti, 2005:4). The term "writing system" sometimes refers to the specific system in which a particular language is written. For example, the Japanese writing system is a systematic mixture of logographic, syllabic and alphabetic writing systems, which makes use of three types of scripts: logographic kanji; moraic *kana*, which is actually a collective name for the two sets of moraic scripts *hiragana* and *katakana*; and alphabetic *romaji*, which is used only supplementarily. Learners of Japanese have to learn the orthographies of kanji, *kana*, and *romaji* and how to mix them properly in order to master the Japanese writing system, as explained in section 2.1.

Writing systems are divided into three major types according to the mapping principle of their graphemes: meaning-based **logographic** writing systems in which a grapheme represents a word, and two types of sound-based writing systems, namely, **syllabic** writing systems with grapheme-syllable correspondence, and **alphabetic** writing systems with grapheme-

phoneme correspondence (Perfetti & Dunlap, 2008). The logographic writing system is sometimes referred to as a morphographic writing system, depending on the understanding of the semantic unit represented by a grapheme. The system is called logographic if the unit is understood to be a word (a single distinct meaningful unit), and morphographic if it is categorised as a morpheme (the smallest unit of meaning) (see section 2.2 for further details). Japanese kanji and Chinese *hanzi* are examples of the logographic (morphographic) writing system. The syllabic writing system includes the moraic writing system, in which a grapheme corresponds to a quasi-syllabic phonological unit, the mora (see Note 2 in Chapter 2). Tibetan is an example of the syllabic writing system, and Japanese *kana* of the moraic writing system. Graphemes in the alphabetic writing system (e.g. English and Japanese *romaji*) represent both consonants and vowels. Korean *hangul* is an unusual example of the alphabetic writing system. *Hangul* symbols are arranged in a square configuration representing a syllable instead of the customary linear formation. A writing system in which graphemes represent only consonants (vowels can be represented with diacritics, but they are normally excluded), such as Persian and Hebrew, is called a consonantal writing system. This overall division is summarised in Table 8, together with script examples:

Table 8 Major types of writing system (Based on Cook & Bassetti, 2005)

Base of writing system	**Represented unit**	**Type of writing system**	**Script example**		**Orthography example**
Meaning	words/ morphemes	logographic (morphographic)	*Hanzi*	汉字	Chinese
			kanji	漢字	**Japanese (kanji)**
Sound	syllables/ morae	syllabic	Tibetan	བོད་སྐད	Tibetan
		moraic	*kana*	かな	**Japanese (*kana*)**
	phonemes	consonantal	Persian	ﻯﺳﺭﺍﻑ	Persian
			Hebrew	תירבע	Hebrew
		alphabetic	*hangul*	한글	Korean
			Roman	English	English
				romaji	**Japanese (*romaji*)**

As shown in Table 8 in bold type, the Japanese writing system uses three orthographies, covering all three of the major writing system types.

Orthographies with highly regular GSC (grapheme-sound correspondence) are phonologically transparent. One-to-one GSC makes the pronunciation of a given spelling (as well as the spelling of a given pronunciation)

predictable. Orthographies within the same writing system can have differing phonological transparency. In the alphabetic writing system, orthographies such as Italian, German and Japanese *romaji*, that have basically one-to-one GSC, are highly transparent. As the regularity of GSC decreases, the orthography becomes less transparent. According to this principle, Swedish and Dutch, with less regular GSC, are less transparent than Italian and German, while English and French with rather irregular GSC are more opaque than transparent. Orthographies under the consonantal writing system, such as Persian and Hebrew, are highly opaque, for they do not normally represent vowels and the pronunciation is unpredictable without the context. Japanese *kana* orthography has nearly complete one-to-one GSC and therefore is highly transparent. Japanese kanji, on the other hand, is particularly opaque, due to an abundance of characters with multiple readings and homophonous characters intermingling one-to-many and many-to-one GSC (Perfetti & Dunlap, 2008, Wydell, 2008).[6]

The size of the phonetic unit corresponding to a grapheme is called **granularity**. Among the three major writing systems, the alphabetic system is of fine granularity, the syllabic system is medium-grained, and the logographic coarse-grained. The abovementioned assessment of **phonological transparency** of orthographies and the notion of granularity according to the writing system are integrated and illustrated in Figure 3:

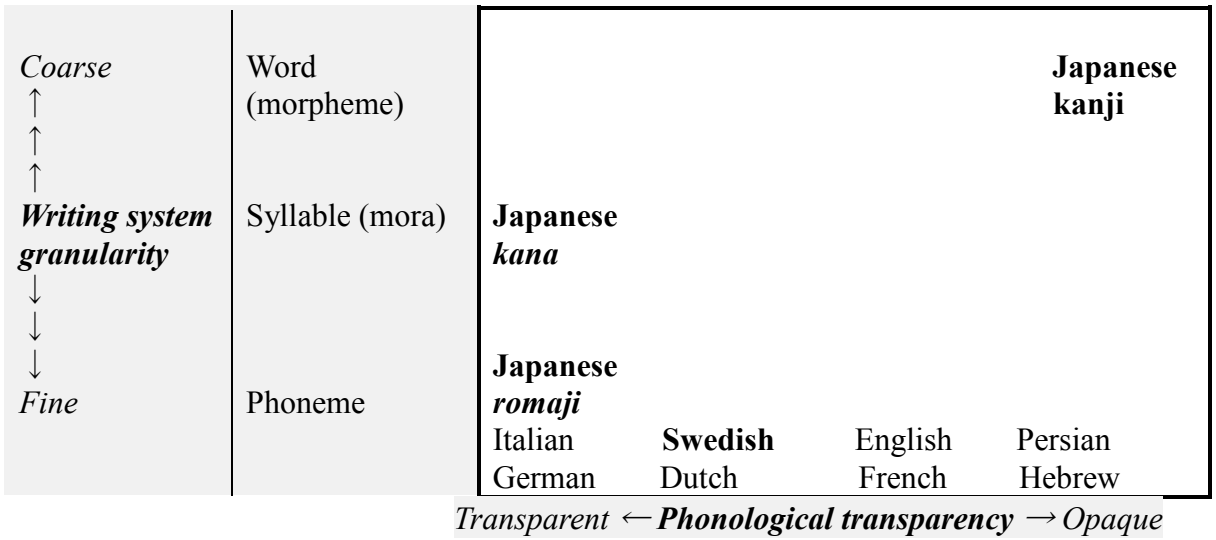

Figure 3 Writing system granularity and phonological transparency (Based on Wydell, 2008; and Perfetti & Dunlap, 2008)

As illustrated in Figure 3, the three orthographies kanji, *kana* and *romaji* are of varying granularity and transparency, verifying the complexity of the Japanese writing system, which intermixes these three. It also indicates that Swedish orthography has the same granularity as *romaji* and is less transparent than *kana* and *romaji* but much less opaque than kanji. In order to understand problems encountered in learning L2WS with different granularity and transparency from L1WS, the next subsection will present a comparison across writing systems.

3.1.2. Differences across writing systems

Although different writing systems have certain universal aspects, there are differences in reading, writing and awareness due to dissimilarities in writing system granularity and phonological transparency. L1 users of different writing systems cope with and are aware of different units of language with different degrees of phonological transparency in decoding and encoding (Cook & Bassetti, 2005). This subsection will compare reading, writing and awareness across writing systems.

The process of reading English words aloud has been commonly conceptualised as a dual-route model, as shown in Figure 4 (Patterson & Morton, 1985, Cook & Bassetti, 2005):

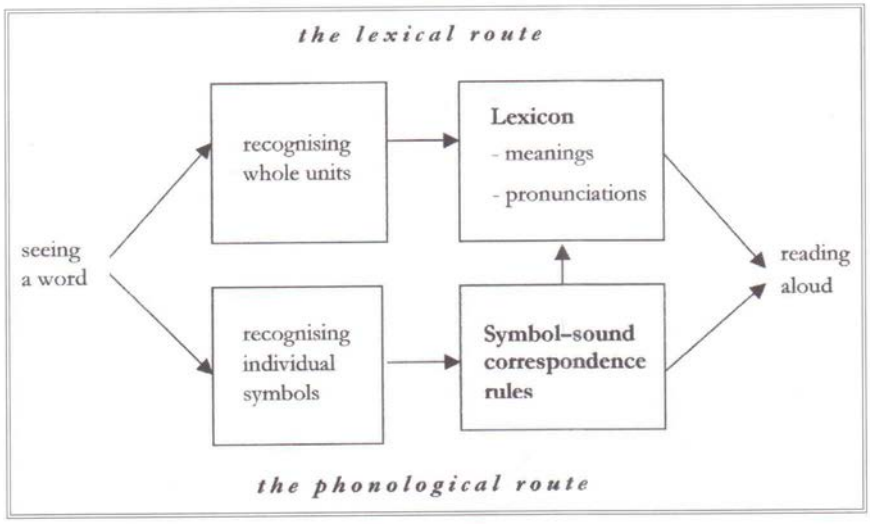

Figure 4 The dual-route model of reading aloud (Patterson & Morton, 1985; Cook & Bassetti, 2005:14)

In the phonological route, words are decoded through GSC, whilst whole-word processing through the mental lexicon takes place via the lexical route. For example, the word "tree", with regular GSC, can be decoded through the phonological route as "t-r-ee" ([t]-[r]-[i:]) and pronounced

[tri:], whereas the whole word of "yacht", with irregular GSC (with the silent word-medial *ch*), is recognised as one unit, the meaning "sailing boat" and the pronunciation [jɒt] being retrieved from the mental lexicon (Cook & Bassetti, 2005). The phonological route decodes the written symbol into sounds and connects them to meanings; the lexical route regards the written symbols as representations of meanings, which may be linked to sounds (Cook & Bassetti, 2005).

Figure 5 is the dual-route model of English spelling (Cook & Bassetti, 2005:18), which parallels its reading counterpart in Figure 4. Words that are frequently used and/or orthographically opaque, i.e. those that have irregular GSC (such as "through"), are spelled via the lexical route, whereas phonologically transparent and/or infrequent words (such as "tooth" or "hippopotamus") take the phonological route.

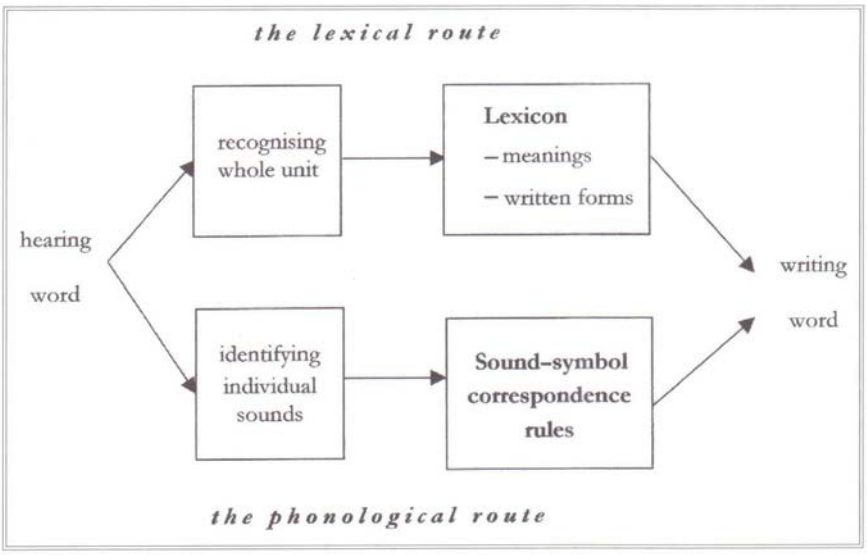

Figure 5 The dual-route model of English spelling (Cook & Bassetti, 2005:18)

When this dual-route model is applied to different writing systems, an expected parallel is observed. Although the meaning-based logographic writing system makes primary use of the lexical route, and the sound-based syllabic and alphabetic writing systems rely greatly on the phonological route, activation of one route completely exclusively of the other is not as common, and simultaneous and interacting activation of both routes has been the leading approach in recent years (Cook & Bassetti, 2005).

In terms of research across writing systems, different writing systems rely on each route to varying degrees. For example, in Japanese kanji and Chinese *hanzi*, which are logographic, a character cannot be written solely via the phonological route. Instead, the whole character has to be retrieved from the mental lexicon. In consequence, morphological/orthographic awareness plays an important role in the acquisition of logographic writing

skills, and even in consonantal and orthographically opaque Hebrew (Hanley et al., 1999), whereas distinctive correlations have been observed between phonological awareness and spelling skills in English (Goswami, 1999), which is a sound-based orthography of medium transparency (Cook & Bassetti, 2005).

Ellis et al. (2004) explored the influence of phonological transparency on L1 learners' reading acquisition in five different orthographies (Japanese *kana*, Albanian, Greek, English and Japanese kanji). They concluded that phonological transparency affected both the accuracy rate in and strategy for L1 reading: the more opaque the orthography was, the greater the proportion of errors was and the more holistic was the approach applied in reading.

Wang, Koda & Perfetti (2003) examined the word recognition strategies of level-matched Korean (alphabetic L1WS) and Chinese (logographic L1WS) groups of learners of English in a semantic category judgement task. The Korean learners made more false positive errors to homophonous exemplars (e.g. *bare* in place of *bear*) than to spelling controls, whereas the Chinese learners made such errors only when homophones were orthographically similar to exemplars (e.g. *creak*/*creek*, but not to *knows*/*nose*). Furthermore, Korean learners showed better phonological awareness than Chinese learners. Wang et al. (2003) suggested that learners' L1WS mapping principles affect strategies in their L2 word recognition: Korean learners with alphabetic L1WS (*hangul*) are used to taking the phonological route following the GSC rules and make use of a sublexical strategy in reading English; whereas L1 logographic Chinese learners who are more dependent on the lexical route in their L1WS read English using a lexical (holistic) strategy, i.e. visually recognising the whole word rather than merely phonologically decoding based on GSC.

Nelson et al. (2005) conducted brain-imaging research and analysed fMRI data from English learners of Chinese and Chinese-English bilinguals (L1 Chinese). After one year of studying Chinese, English learners showed newly gained activation of the brain regions (bilateral visual and visual-temporal cortex) that were used by Chinese readers. On the other hand, the Chinese bilinguals showed the Chinese pattern of bilateral activation in reading both Chinese and English, and the alphabetic pattern of left hemisphere dominance was never observed. This indicates that English speakers had to develop a new brain network for visual recognition to read Chinese that cannot be decoded phonemically, whereas Chinese speakers applied the same whole-word recognition approach to reading both L1 Chinese and L2 English. Since there is a competence level difference between the subject groups in the research (near novice vs. skilled bilingual) and no readers of any writing system are 100% dependent on

only one type of strategy, the observation should not be too simplified or overgeneralised. However, the result suggests a possibility that the lexical (holistic) strategy (whole-word recognition via the lexical route) can deal with both phonologically opaque and transparent orthographic systems but that the sublexical (GSC-based decoding) approach is specific to the phonologically transparent orthography (Perfetti & Dunlap, 2008).

The findings of Wang et al. (2003) and Nelson et al. (2005) can be summarised as Table 9:

Table 9 Summary of L1 phonological transparency and L1 & L2 reading strategies (Wang et al., 2003; Nelson et al., 2005)

L1 orthography • **phonological transparency** • **writing system**	**Dominant L1 reading strategy**	**L2 orthography** • **phonological transparency** • **writing system**	**Dominant L2 reading strategy**	**Source**
Korean • transparent • alphabetic	sublexical (GSC-based decoding)	**English** • less transparent • alphabetic	sublexical (GSC-based decoding)	Wang et al. (2003)
English • less transparent • alphabetic	sublexical (GSC-based decoding)	**Chinese *hanzi*** • opaque • logographic	sublexical (GSC-based decoding) ↓ lexical (whole-word recognition)	Nelson et al. (2005)
Chinese *hanzi* • opaque • logographic	lexical (whole-word recognition)	**English** • less transparent • alphabetic	lexical (whole-word recognition)	Wang et al. (2003) Nelson et al. (2005)

Furthermore, the conclusions and suggestions of Ellis et al. (2004), Wang et al. (2003) and Nelson et al. (2005) can be integrated and summarised in the following two statements:

A. The degree of phonological transparency of L1 orthography is an indicator of the reading strategy/approach taken by the reader: readers with transparent L1 orthographies tend to read via the phonological route using the sublexical strategy and decoding words based on GSC, while those with opaque L2 orthographies depend on the lexical route, preferring the holistic approach and recognising the word as a whole.

B. Readers of L2WS tend to apply their L1WS reading strategy. The applied L1WS strategy can handle the L2WS as long as the L2WS's

degree of transparency is the same as or greater than that of the L1WS. If the L2WS is more opaque than the L1WS, readers must shift the dominance of their reading strategy towards lexical approach. In the case of alphabetic L1WS and logographic L2WS, readers even develop new brain networks to enhance their visual recognition ability for the lexical strategy, so that they can cope with the phonemically undecodable writing system.

In consideration of the findings and observations summarised in Table 9 and statements A and B above, the strategic differences between Swedish learners of Japanese and L1 Japanese learners, as compared in this study, can be synopsised as Table 10. The grey-marked features are the focal point of this study study:

Table 10 Summary of L1 phonological transparency and L1 & L2 reading strategies (present study)

L1 orthography • **phonological transparency** • **writing system**	**Dominant L1 reading strategy**	**L2 orthography** • **phonological transparency** • **writing system**	**Dominant L2 reading strategy**	**Source**
Swedish • moderately transparent • alphabetic	sublexical (GSC-based decoding)	**Japanese kanji** • highly opaque • logographic	sublexical (GSC-based decoding) ↓ lexical (whole-word recognition)	present study
Japanese *kana* • highly transparent • moraic	sublexical (GSC-based decoding)	•		present study
Japanese kanji • highly opaque • logographic	lexical (whole-word recognition)			

Since Swedish has an orthography of moderate phonological transparency under the alphabetic writing system, it is most likely that Swedish speakers' dominant L1WS reading strategy is sublexical, and that they will try to apply it in reading kanji. Being unable to phonemically decode kanji characters, however, they will develop a L2WS lexical strategy, just like the English learners of Chinese in Nelson et al. (2005). On the other hand, Japanese native speakers develop both strategies to read their L1WS, first the sublexical strategy at around six years of age when they have learned *hiragana* and *katakana*, and subsequently the lexical strategy when they start learning kanji at the age of six or seven. It is one of the purposes of this study to explore the differences between Swedish learners' L2WS

strategy and level-matched Japanese native speakers' L1WS strategy in reading kanji, as indicated with grey marking in Table 10.

Apart from the predominance of one route over the other, the following aspects are known to be indicators of the differences relating to phonological transparency between the meaning-based and sound-based writing systems:

- **Phonological activation**
  In a meaning-based writing system, phonological activation is a threshold procedure (i.e. it occurs after the kanji/*hanzi* character is recognised), while it is a cascaded procedure (i.e. occurs from the start and successively) in a sound-based system.

- **Word familiarity and frequency**
  Since word familiarity and frequency affect whole-word recognition in the lexical route but the phonological route is unaffected, phonologically opaque systems are more affected by word familiarity and frequency than transparent systems; hence more effects will be expected in reading highly opaque Japanese kanji than rather transparent Swedish words.

- **Correlated skills for writing system learning**
  Phonemic awareness shows a correlation with reading skills in alphabetic writing systems, whereas visual skills correlate with the logographic writing system. Huang & Hanley (1995) compared Chinese- and English-speaking children and found that the former's competence in reading is correlated with visual skills test results and the latter's with phonological awareness test results.

- **Morphemic awareness** is required in the spelling of less phonologically transparent writing systems (Muter & Snowling, 1997), but this is not necessarily the case with transparent writing systems.

(Cook & Bassetti, 2005)

Since Japanese kanji is a phonologically opaque orthography in the meaning-based logographic writing system and Swedish is a rather transparent orthography in the sound-based alphabetic writing system, the cross-writing system differences reviewed above can be applied to the comparison of the two orthographies as summarised in Table 11:

Table 11 Cross-writing system differences between Japanese kanji and Swedish

Differentiating aspects	Japanese kanji	Swedish
Predominant route for reading/writing	lexical route (whole-word recognition)	phonological route (GSC-based)
Predominant strategy for reading/writing	lexical (holistic) strategy	sublexical strategy
Phonological activation	threshold procedure	cascaded procedure
Word familiarity & frequency	influential	less influential
Correlated skills for learning	visual skills	phonemic awareness
Morphemic awareness	required	optional

The writing of L2WS has been less well explored than the reading component. Oller and Ziahosseiny (1970) compared L1WS users of the Roman alphabet and L1WS users of other writing systems (Chinese, Japanese, Arabic) and found that the former made more spelling errors in their L2WS than the latter, concluding that the L1WS spelling processes affect the L2WS spelling strategy. The latter's L1WS whole-word recognition approach is most likely to have resulted in strict visual scrutiny of the L2WS spelling and reduced the number of spelling deviations. Cook (2004) compared L2WS learners of English and L1WS English children in mastery of the uniform spelling of the written past tense morphology and found that L2WS learners were quicker than L1WS children, possibly due to the difference in age and their own L1WS literacy skills, but most probably because L2WS learners had received explicit instructions on *-ed* as the English past tense morpheme, which was apparently absent in L1WS children's learning process. There are also a limited number of cross-linguistic studies on kanji, which will be reviewed in subsections 3.2.3 and 3.3.3.

As seen above, learners of different writing systems are required to become aware of different linguistic units (phonemes in the alphabetic systems such as English and Swedish, morae in the moraic system such as *kana*, and words and morphemes in logographic systems such as kanji and *hanzi*) (Cook & Bassetti, 2005). Since English has the most studied L2WS, most research on language awareness has focused on phonemic awareness. By comparison, there is a limited amount of L2WS research on orthographic awareness. Wade-Woolley (1999) compared Japanese and Russian learners of L2WS English in a judgement test in which they were asked to decide if a presented sequence of letters was a word or non-word in English, and found that the Japanese learners were quicker in judging than the Russian learners. His explanation of the difference in reaction time was that the Japanese learners could make quicker decisions based on the orthographic information (whole-word recognition) than the Russian learners, who relied more on phonological decoding. Su et al. (2010) tested English-speaking

learners of Chinese as a foreign language (CFL) of different competency levels for radical awareness and correlations between radical awareness and word recognition (radicals are subcomponents of characters that function as semantic and/or phonological cues). The study found that advanced CFL learners showed higher level of radical awareness than novice CFL learners, and that semantic functional radical awareness serves as a predictor of word recognition among CFL learners.

3.1.3. Error analysis

Error analysis is a common approach in L2 research to collecting and describing learners' writing. The technique of error analysis was started by Corder (1974) and replaced contrastive analysis, which was the mainstream approach in the 1960s and early 1970s. Contrastive analysis was based on the theory that L2 learning difficulty is due to the structural differences between L1 and L2, claiming that all errors made by L2 learners were ascribable to L1 interference; however, this claim could not be supported by the empirical evidence of L2 learner error types not corresponding to any of the structures of the learners' L1s (Ozeki, 2010). In error analysis, learners' writings were collected and carefully examined for mistakes, which were subsequently classified into different types for analysis. This approach revealed that many of the errors that could not be explained with L1 interference were the result of learners' fallacious inference of L2 rules (Ozeki, 2010).

Most studies in L2WS error analysis use a single group of L2WS learners as subjects, although some experimental researchers have made comparisons between L2WS learners and L1WS users (Bebout 1985) or between groups of L2 learners with different L1WS backgrounds (Cook, 1997) (Cook & Bassetti, 2005).

One problem with this technique is the difficulty in classification. One error can have several different causes, and therefore it is not always easy to classify errors into the correct categories (Ozeki, 2010). A graver problem is learners' possible avoidance of the forms and expressions they find particularly difficult or are uncertain of. If certain forms are not used, errors involving such forms will not be found, but it by no means indicates learners' mastery of the said forms. Furthermore, there is a fundamental issue that error analysis looks exclusively at learners' incapability, and that the general picture of learners' L2 ability, usage and learning strategies cannot be investigated (Ozeki, 2010). In search of a more holistic approach, the concept of interlanguage has been developed. Interlanguage is a new linguistic system that L2 learners have created in their mental lexicon, which is situated somewhere between their own L1 system and the system possessed by native speakers of the target language. Interlanguage should

be analysed not from the point of view of judging the legality of learners' use of the target language, but in order to elucidate how it is formulated and used by learners (Ozeki, 2010).

The aim of this study is to investigate the characteristics of Swedish learners of Japanese concerning their kanji interlanguage in the developmental process of kanji learning, especially their retrieval methodologies and difficulties in recognition and production. In order to attain this aim, the technique of error analysis was used with good cause. In descriptive and observational research of kanji, analysing correct forms of kanji reading and writing does not provide many insights into the learners' mental lexicon, its activation pattern or retrieval approaches. A kanji quiz answer sheet with full marks would be a proof of the test taker's excellent command of kanji, but not much more than that. On the other hand, an answer sheet with a number of errors, substituting the target characters with homophonous characters, indicates leanings towards the phonological aspect of kanji in a learner's retrieval of characters from the mental kanji lexicon (retrieval via the phonological route). Alternatively, writing a pseudokanji character consisting of 扌 and 虫 for the target character 独 ("lone") suggests orthographic confusion of 扌(the radical for "hand") and 犭(the radical for "beast") in the mental lexicon. In addition, it indicates that this unit of kanji memory is radical-based rather than stroke-based (i.e. the learner is at least capable of writing a pseudo-character consisting only of existing radicals, rather than as an illegal assemblage of strokes). The errors to be analysed in this study were collected from kanji reading and writing test. Unlike free composition, in which writers can avoid using difficult or uncertain characters, the test setting requires the takers to produce the readings or written forms of the target characters. Failure to give an answer is recorded as a blank, which indicates non-mastery of the target character.

3.2. Kanji recognition and processing

This section will review studies on kanji recognition and processing that are closely related to this study. Since this study is a comparison of novice and advanced L2 learner groups with level-matched L1 learner groups, the focus will be on those studies taking informant groups' level differences into careful consideration. Subsection 3.2.1 will be on studies having L1 learners as informants, 3.2.2 on those with L2 learners and 3.2.3 on L1/L2 comparative studies.

3.2.1 Studies on kanji recognition and processing with L1 learners

The Basic Academic Ability Survey Committee of Teaching Skill Sharing Society of Japan (BAASC) (2007) conducted a large-scale survey of Japanese schoolchildren's kanji reading ability, involving second to seventh graders in 462 primary schools and 18 junior high schools all over Japan. The target kanji characters were selected from the List of Kanji by School Year (学年別漢字配当表 *Gakunenbetsu Kanji Haitōhyō*) (Ministry of Education, Culture, Sports, Science and Technology-Japan, 1989).

Pupils were tested at the beginning of a school year for characters on the list for their immediately preceding grade, i.e. the second graders were tested on the Grade 1 kanji list, and the seventh graders on the Grade 6 kanji. The target characters were presented in word form embedded in short sentences and phrases, which provided the context needed to determine the meaning and reading of the words. The task in the reading test was to fill in the blanks with the reading of the target character in the given word and context. Since the purpose of the survey was to grasp schoolchildren's kanji acquisition situation and to seek educational implications, the error and blank rates were calculated and error examples were collected for each kanji, and the patterns according to grade and characters with particular difficulty were discussed. However, the errors were not classified according to cognitive aspects such as phonological or orthographic type.

In their analysis BAASC pointed out general tendencies of higher error and blank rates in higher grades and difficulty in reading Sino-Japanese (*on*-reading) words that are outside of children's everyday language. The higher error/blank rates for higher grades indicates that higher-grade pupils have an increased workload (the characters they learn are orthographically and phonologically more complex and constitute more abstract vocabulary), which leads to more frequent failure to produce correct answers (or answers of any kind) than lower-grade pupils. The difficulty in reading *on*-reading characters that are used for words less familiar to children implies that their kanji reading ability is supported by non-kanji linguistic skills: the sentences and phrases in which the target characters were embedded must have given them reading clues via grammar and context, but if the words involving the characters were outside of their everyday vocabulary, they were unable to make good use of these clues.

BAASC (2007) is of particular importance for this study because the format of the reading test in this study is based on this survey's task type and test format. The task and format of this survey was chosen for a number of reasons. This survey is an error analysis in a test setting, in which the participants are required to produce readings of appropriate

selections of target characters, as opposed to essays and free compositions in which they can avoid using difficult characters. The task type (reading kanji presented in a context) is close to what is required in normal real-life reading (such as reading books and letters), as opposed to more experimental tasks such as character recognition and semantic category judgement. The fill-in-the-blanks format used in this survey is the format to which both L1 and L2 learners are accustomed in the form of exercises and written examinations. By applying cognitive classifications to the analysis of errors, the methodology of this survey can be adapted to suit the aims of this study.

Chan & Nunes (1998) investigated L1 Chinese children's awareness of functions and positional constraints of radicals (kanji components of functional importance). Positional constraint awareness was examined by asking children to judge the legality of non- and pseudo-characters with switched radicals (i.e. if they violate positional rules or not), and functional awareness by creating and pronouncing a character for newly introduced object. Six-year-old subjects were able to judge the legality according to positional rules and made use of semantic radicals to create characters, but did not make use of phonological radicals to pronounce them. On the other hand, nine-year-old subjects were able to use both semantic and phonological radicals according to their functions. Chan & Nunes (1998) concluded that functional awareness of semantic radicals might develop at an early stage in Chinese L1WS acquisition, but that awareness of the phonological function of radicals would develop later in the acquisition process.

The late development of phonological radical awareness function may be due to the difference in cognitive ability between six-year-olds and nine-year-olds, or alternatively, because of the difference in the number of characters they have learned, since the more characters they have learned, the more characters there are in children's mental lexicon sharing a radical and sound to help them become aware of the functions of phonological radicals.

Although this is a study of *hanzi*, which is phonologically less opaque than kanji (*hanzi* characters basically have only one pronunciation), the results and implications can be used as a reference for studies of kanji, which has the same positional constraints and radical functions. Due to kanji's high level of opacity (most characters have multiple readings), however, the phonological functions of radicals may be even more difficult to grasp, and are therefore developed at a later stage of learning.

3.2.2. Studies on kanji recognition and processing with L2 learners

Matsumoto-Sturt (2004) compared novice and advanced L1 English adult learners of Japanese and examined frequency effects in naming latencies (the mean time from presentation to articulation) of kanji compounds. In the experiments, high-frequency words had shorter latency and were more accurately pronounced by novice learners. The results indicated that the quality of early interlanguage lexical representations is affected by word frequency and that low frequency words caused difficulty in L2 reading. Including word frequency as a variable for L2 experimental research is often problematic, since L2 exposure of the learners outside of Japan is usually limited to in-class instructions and activities, and accordingly, the application of L1 word frequency databases is often irrelevant. Therefore, this study has the extra merit of developing an original L2 frequency database, based on the course contents, thereby increasing the reliability of the findings.

Matsumoto (2013) compared three groups of adult L2 learners: beginner's level students with L1 alphabetic (English) background, beginner's level students with L1 logographic (Chinese) background, and intermediate level students with L1 alphabetic (English) background. She examined their kanji recognition strategies and the possible influence of L2 exposure thereon, using lexical judgment tasks involving pseudo-homophones, pseudo-homographs and real words. Both groups with alphabetic background showed poor visual recognition strategies for L2 kanji decoding, which implied underdeveloped orthographic awareness, whereas the group of L1 logographic beginners demonstrated character-based access as a result of sufficient orthographic knowledge. In addition, there were significant differences in reaction time for the judgment task according to level. The findings indicated use of different reading strategies according to the learners' L1 background, as well as the level differences caused by different degrees of L2 exposure. This study would be even more enlightening if comparison were made with an additional group of intermediate level L1 logographic learners.

3.2.3. Studies on kanji recognition and processing comparing L1 and L2 learners

Tamaoka (1992) compared two levels of Canadian (L1 English) learners of Japanese and Japanese college students to examine how their kanji learning experience affected their mental kanji word processing speed. The latency of the Canadian learners was affected by the orthographic complexity (number of strokes) of kanji included in the words, and the lower level students were more affected than the higher level students. On the other

hand, the Japanese students did not show any significant difference depending on the complexity of kanji. In his analysis, Tamaoka suspects that the difference in latency is caused by different degrees of character familiarity developed by the two groups of students. Japanese students who had had longer learning experience were familiar with all the characters used in the test and had developed whole character-based automatic processing skills even for complicated characters, whereas the Canadian student groups, who had 1-2 or 2-3 years of learning experience, were less familiar with complex characters and they had to decode the kanji configuration on a stroke or component basis, which resulted in longer latency. The findings of Tamaoka (1992) suggest that skilfulness in reading is under the sequential influence of the duration of study, the extent of exposure to kanji, character familiarity, unit of kanji perception, and processing speed.

Komori (2009) compared intermediate- and advanced-level adult L2 learners of Japanese (with L1 English background) with adult L1 Japanese speakers concerning recognition and processing of two-character kanji compounds. She used word/non-word lexical decision tasks to assess form identification, word recognition, meaning access and memory-retaining abilities, as summarised below:

- In the form identification task, significant differences were observed between all three groups. The intermediate learners were slower and less accurate in identifying words consisting of more complex characters (with more than 25 strokes in total) than simple-character words (with max 9 strokes in total), the advanced learners were slower but not less accurate, and L1 users were quick and accurate in both simple and complex characters.

- The lexical decision task with sound prime showed a significant difference between L2 learners and L1 users: both L2 learner groups were influenced by the sound information, but the L1 users were unaffected.

- In the meaning access task, the intermediate learners were slower and less accurate in abstract words than concrete words, although advanced learners showed no difference between abstract and concrete words, similarly to L1 users.

- The retention task examined the ability to retain, recall and produce visually presented kanji words of different abstractness. Ten kanji words were visually presented in sequence for four seconds each, and immediately after the last word, the subjects were asked to recall as many words as possible and write them down in any order within

two minutes. No significant differences were observed among the three groups in terms of abstractness, although there were specific differences in the L2 learner groups' approaches: the intermediate learners made mostly orthographic errors; and the advanced learners used phonological information for retention. The intermediate learners could not read the orthographic forms and tried to memorise and reproduce them without success. On the other hand, the advanced learners' errors were mostly written in *kana* or *romaji*, indicating that they converted the orthographic representations of the kanji words into their phonological representations to retain them in the memory, but could not convert them back to their orthographic forms. In other words, they could read the words, but could not write them.

Based on the abovementioned results, Komori concluded as follows:

A. The intermediate and advanced learners share an analytical decoding strategy for the orthographic and phonological aspects of kanji recognition process;

B. L2 learners are more dependent on phonological information than L1 users; and

C. The advanced learners' kanji word recognition process is similar to that of L1 users' in many aspects but is not the same (e.g. there was no evidence of the advanced learners' establishment of the automatic recognition pattern that was characteristic of L1 users).

Komori (2009) used the same sets of target words (consisting of characters selected from approximately 1,400 characters on Grades 1-7 lists of the List of Kanji by School Year) for all three groups, which required the level settings of the L2 learner groups to be rather high: the advanced learners had had approximately 900 class hours, had learned 2,000 kanji characters and had passed Level 1 of the Japanese Language Proficiency Test (JLPT), and the intermediate learners had had 600 class hours, had learned 1,000 characters and had passed Level 2 of the JLPT. The L1 users were university students, who had also studied approximately 2,000 characters of the Jōyō Kanji by the end of their secondary education. The differences between the intermediate and advanced learners' characteristics are ascribable to the difference in the number of learned kanji and their competence in Japanese, whereas the differences between the advanced learners and L1 users might be any one or combinations of several reasons: the advanced learners' dependence on phonology and tendency towards analytical decoding may be due to L1WS transfer, while L1 users'

unconditional accuracy and speed are probably thanks to their long and extensive processing experience.

3.2.4. Summary of kanji recognition/processing studies

Table 12 is a summary of the kanji recognition/processing studies reviewed in subsections 3.2.1, 3.2.2 and 3.2.3.

Table 12 Summary of kanji recognition/processing studies

Study	Subjects	Tested aspect	Experiment	Comparison
Studies with L1 subjects (subsection 3.2.1)				
BAASC (2007)	L1 Japanese children (Grades 2 - 7)	Reading ability of kanji learned previous year	Fill-in-the-blanks with reading of target kanji in given word and context	Reading types (*on* vs. *kun*)
	• Higher error and blank rates in higher grades. • Difficulty in reading *on*-reading words (unfamiliar vocabulary).			
Implications	➢ Increased workload due to quantity and complexity of characters for higher grades. ➢ Dependence on non-kanji linguistic skills.			
Chan & Nunes (1998)	L1 Chinese children - 6-year-olds - 9-year-olds	Radical positional constraint awareness	Positional legality judgement of switched radicals	Non-characters vs. pseudo-characters
		Radical function awareness	Creating and pronouncing a character for a newly introduced object	Semantic radicals vs. phonological radicals
		6-year-olds	9-year-olds	
	Positional constraints	Yes	Yes	
	Semantic function	Yes	Yes	
	Phonological function	No	Yes	
Implications	➢ Early development of semantic radical awareness ➢ Late development of phonological radical awareness ➢ A large stock of characters sharing a radical and sound is necessary to become aware of the functions of phonological radicals.			

(continued)

Table 12 *Continued*

• Study	• Subjects	• Tested aspect	Experiment	• Comparison
• **Studies with L2 subjects (subsection 3.2.2)**				
Matsumoto-Sturt (2004)	L1 English adults - novice - advanced	Frequency effects on naming latencies	Naming of kanji compounds	Word frequency (high vs. low)
	• Shorter latency and more accurate pronunciation by novice learners for high-frequency words.			
Implications	➢ Word frequency affects quality of early interlanguage lexical representations. ➢ Low frequency words cause difficulty in L2 reading.			
Matsumoto (2013)	L1 English/Chinese adults - L1 English beginners - L1 Chinese beginners - L1 English intermediate	• Kanji recognition strategies • Influence of L2 exposure	Lexical judgment	Pseudo-homophones / pseudo-homographs / real words
	• Poor visual recognition strategy for decoding by both L1 English groups. • Character-based access due to sufficient orthographic knowledge by L1 Chinese group. • Significant level differences in reaction time for judgement task.			
Implications	➢ L1 English groups' underdeveloped orthographic awareness. ➢ Use of different reading strategies according to learner's L1 background. ➢ Level differences caused by different degrees of L2 exposure.			
• **Study**	**Subjects**	**Tested aspect**	**Experiment**	**Comparison**
Comparative studies with L1 & L2 subjects (subsection 3.2.3)				
Tamaoka (1992)	L1 English adults - intermediate - advanced L1 Japanese adults	Influence of kanji learning experience	Mental kanji processing speed	Orthographic complexity (simple vs. complex)
	• Orthographic complexity affected L1 English learners' latency (lower level students more than higher level students), but not L1 Japanese adults'.			
Implications	➢ Different degrees of character familiarity caused difference in latency. ➢ Japanese students, with longer learning experience, developed whole character-based automatic processing skills even for complex characters. ➢ Longer latency of L1 English groups (less familiar with complex characters) due to kanji decoding on a stroke or component basis. ➢ Reading skill is influenced by: (1) duration of study; (2) extent of exposure to kanji; (3) character familiarity; (4) unit of kanji perception; and (5) processing speed.			

(continued)

Table 12 *Continued*

• **Study**	**Subjects**	**Tested aspect**	**Experiment**	**Comparison**
Komori (2009)	L1 English adults - intermediate - advanced L1 Japanese adults	Recognition and processing of two-character kanji compounds	Form identification	Orthographic complexity (simple vs. complex)
			Word recognition	Sound prime (with vs. without)
			Meaning access	Word abstractness (concrete vs. abstract)
			Memory retention	Word abstractness (concrete vs. abstract)
		L1 Eng. Int.	L1 Eng. Adv.	L1 Japanese
	Form	Slow and inaccurate with complex characters	Slow with complex characters	Quick and accurate
	Word	Affected by sound prime	Affected by sound prime	Unaffected by sound prime
	Meaning	Slow and inaccurate with abstract character	Unaffected by abstractness	Unaffected by abstractness
	Memory	Unaffected by abstractness; mostly orthographic errors	Unaffected by abstractness; use of phonological information	Unaffected by abstractness
Implications	➢ Characteristics shared by L1 English groups: - Analytical decoding strategy for the orthographic and phonological aspects of the kanji recognition process - More dependent on phonological information than L1 Japanese group ➢ Many similarities between L1 English advanced and L1 Japanese, but no automatic recognition pattern is established even by advanced learners.			

The implications of the previous kanji recognition/processing studies reviewed above in Table 12 can be summarised and reorganised according to learner groups and aspects of kanji competence, as listed in Table 13:

Table 13 Summary of implications of kanji recognition/processing studies

Subjects	Approach	Ability	Difficulty
L1 children	Dependence on non-kanji linguistic skills (BAASC, 2007)	Early development of semantic radical awareness and late development of phonological radical awareness (large stock of kanji homophones needed) (Chan & Nunes, 1998)	*On*-reading words (unfamiliar vocabulary) Increased quantity and complexity of characters for higher grades (BAASC, 2007)
L1 adults		High character familiarity; whole character-based automatic processing skills (Tamaoka, 1992) Automatic recognition (Komori, 2009)	Unaffected by complexity/abstractness of character (Komori, 2009)
L2 novice with alphabetic L1WS		Underdeveloped orthographic awareness; L1WS-influenced reading strategy; lesser L2 exposure (Matsumoto, 2013)	Low frequency words (Matsumoto-Sturt, 2004)
L2 novice with logographic L1WS		Developed orthographic awareness; L1WS-influenced reading strategy; lesser L2 exposure (Matsumoto, 2013)	
L2 intermediate with alphabetic L1WS	Dependent on phonological information; orthographic and phonological decoding (Komori, 2009)	Underdeveloped orthographic awareness; L1WS-influenced reading strategy; greater L2 exposure (Matsumoto, 2013) Low character familiarity; longer processing latency (Tamaoka, 1992)	Affected by character complexity/abstractness (Komori, 2009)
L2 advanced with alphabetic L1WS	Dependent on phonological information; orthographic and phonological decoding (Komori, 2009)	Limited character familiarity; longer processing latency (Tamaoka, 1992) Limited orthographic awareness (Komori, 2009)	Unaffected by abstractness of character (Komori, 2009)

3.3. Kanji production

This section will review studies on kanji production. Subsection 3.3.1 will focus on those having L1 learners as informants, 3.3.2 on those with L2 learners and 3.3.3 on those comparing L1 and L2 learners.

3.3.1 Studies on kanji production with L1 learners

Koike et al. (2003) have hypothesised the following five normal developmental stages of L1 learner's *hiragana* and kanji writing as a benchmark for the analysis of dyslexic children's writing disorders as follows:

A. Pre-GSC stage

There is no letter-by-letter recognition. A whole word is recognised as one unit. Writing of own name and names of familiar objects can be performed, although as mere imitation without awareness of GSC (grapheme-sound correspondence).

B. *Hiragana*-word stage

Learner is able to write words with 46 basic *hiragana* symbols, excluding special mora symbols (see subsection 6.1.3 for detailed explanation). Cases of mirror writing may be observed.

C. Special mora word stage

Learner is able to write words in *hiragana* including special mora symbols (with or without difficulty).

D. Basic kanji stage

Learner is able to write basic kanji with concrete meaning, but lacks awareness of radicals.

E. Kanji expansion stage

Learner is able to write kanji, being aware of radicals.

The delayed mastery of special morae (after basic symbols) and later development of radical awareness indicate the greater difficulty of these aspects, which can be observed in L2 learners' *hiragana* and kanji acquisition, as the studies reviewed below indicate (e.g. Toda, 2003; Kano et al., 1989). Since there are very few preschool children among L2 learners of Japanese, it is most likely that Stage A (pre-GCS) is bypassed. The learner groups compared for this study are expected to fall into Stages C, D and E.

The kanji reading survey by BAASC (2007) described in subsection 3.2.1 also examined writing ability for the identical sets of target kanji in the

same manner. In addition to the reading test subjects (second to seventh graders), the writing test included first graders as well, who were tested for their ability to write *hiragana*, not having learned kanji in the previous year (they had just started receiving formal instructions on *hiragana*, but "it is a known fact that many children are able to read and write at least some *hiragana* before the start of their schooling" (BAASC, 2007)). The target was limited to the 46 basic symbols "exclusive of symbols with diacritics, palatalised syllables, geminate consonants, and long vowels that are known to be difficult for children" (BAASC, 2007). The average rate of correct answers was as high as 86.6%.

The task in the writing test was parallel to that in the reading test, which was to fill in the blanks with appropriate characters to complete the phrases/sentences. Corresponding to the reading test, the analysis concentrated on the error and blank rates and problematic characters for each grade, and no cognitive error classification was involved. Tendencies similar to those in the reading results were observed, which were higher error and blank rates in higher grades and difficulty in writing Sino-Japanese (*on*-reading) words outside of children's everyday language.

An interesting finding is the increase in the type of error in which the target characters were substituted with homophonous characters in the results concerning Grade 4 kanji. This indicates that L1 learners have developed a sizable stock of homophonous characters by the end of the fourth grade, when they have learned a cumulative total of 640 characters. By that time the stock is large enough for attempts to retrieve kanji phonologically (i.e. via its pronunciation), but the attempts often result in confusion among several candidate characters.

For the same reasons as for the reading part of this survey, the writing part of the BAASC survey can be adapted to suit the aims of this study by applying cognitive classifications to the analysis of errors.

3.3.2. Studies on kanji production with L2 learners

In their analysis of kanji knowledge of novice L2 learners with different L1WS background learning Japanese at a university, Kano et al. (1989) investigated the patterns of kanji retrieval from the mental lexicon in free recall writing. The learners were asked to write as many characters as possible in five minutes on three occasions during the course of study, when they had learned approximately 50, 300 and 500 characters, and the number of written characters and the association patterns of the character sequences were examined.

As the students learned more characters, the number of characters they wrote increased steadily, and the predominant association patterns of the character sequences shifted from semantic to orthographic, in which there was a shift from the whole-character resemblance type (e.g. 木 “tree” → 本 “book”) to the shared component type (e.g. 木 “tree” → 休 “rest”). There were very few cases of phonological association (sequences of phonologically identical characters).

The semantic association included characters in the same semantic category (e.g. days of the week, directions, nature) or grammatical category (e.g. verbs, adjectives), synonyms and antonyms, constituents of the same compound (e.g. 日 “sun/day” → 本 “book” for 日本 “Japan”), and contextually concurrent characters (e.g. 本 “book” → 読 “to read”). As for the orthographic associations, there was a difference in patterns between L2 learners and L1 users. L2 learners made mostly the shared radical type of errors in which Character 1 is used as a component of Character 2 (e.g. 食 → 飲), whereas L1 users usually made a sequence out of characters sharing radicals (e.g. 飲 → 飯 → 館).

The shift of predominance from semantic association to orthographic association and then from whole-character resemblance to shared radical can be explained as follows:

a. At an early stage, semantically related characters are often introduced simultaneously, and kanji is primarily learned as vocabulary and sorted according to its semantic value;

b. As the number of learned characters increases, learners develop configurational awareness; and

c. As learners learn more characters and possibly receive instructions on radicals, they develop character decompositional ability and intra-character structural awareness.

The possible reasons for the very limited occurrences of phonological association are (i) the task type (character writing by free recall without phonological cues), which requires no phonological memory activation, and (ii) that the characters they had learned (approximately 500 maximum) did not include sufficient numbers of homophonous characters to make phonological character sorting (sorting by pronunciation) meaningful.

Since Kano et al. (1989) did not intend to investigate L1 transfer in particular, the learners had different L1s with varying phonological transparency, and it was therefore impossible to investigate the possible L1 influence on their kanji acquisition process.

Okita (2001) examined kanji writing errors of novice to intermediate level L2 learners of Japanese from seven Asian countries with mixed L1WS backgrounds (logographic and alphabetic). The errors were collected from learners' compositions and classified into the error types graphic, phonological and semantic.

Most errors were of the graphic type (inclusive of pseudokanji and substitution with graphically similar characters), while phonological and semantic errors were few. Learners with logographic backgrounds made fewer graphic errors, most of which were L1-influenced errors (e.g. substitution with 簡体字 or 繁体字, i.e. the simplified or more complicated variations of a character used in China and Taiwan, respectively). Learners with alphabetic backgrounds made errors purely consisting of incorrect forms, sometimes even mistaking the character unit (writing two characters in one box or one character divided into two components in two adjacent boxes). All graphic errors were due to inaccuracy in subtle or partial geometric features (e.g. missing strokes, replaced radicals) rather than total graphic anomaly.

Among the few studies involving L2 kanji writing error analysis, Okita (2001) dealt with a rare comparison of Asian students with different L1 writing systems, which involved more aspects of L1 transfer-related problems than comparisons between learners with a single L1 writing system background: clear examples of L1WS transfer were observed among learners with logographic backgrounds, whereas learners with alphabetic backgrounds showed underdeveloped orthographic awareness. Since the learner groups were of mixed levels and analysis was made according only to learners' L1, the possible influence of level difference was not clarified.

3.3.3. Studies on kanji production comparing L1 and L2 learners

Hatta et al. (1998) compared Japanese college students and Australian university students learning Japanese at beginner's level and analysed their writing errors in two-kanji compound words. Errors were collected from Japanese students' academic writing and Australian students' weekly kanji quizzes. They proposed cognitive models explaining the error generation mechanisms and classified the errors into categories based on such a model, i.e. phonological, semantic and orthographic types. One of their models of error generation mechanisms is presented in Figure 6.

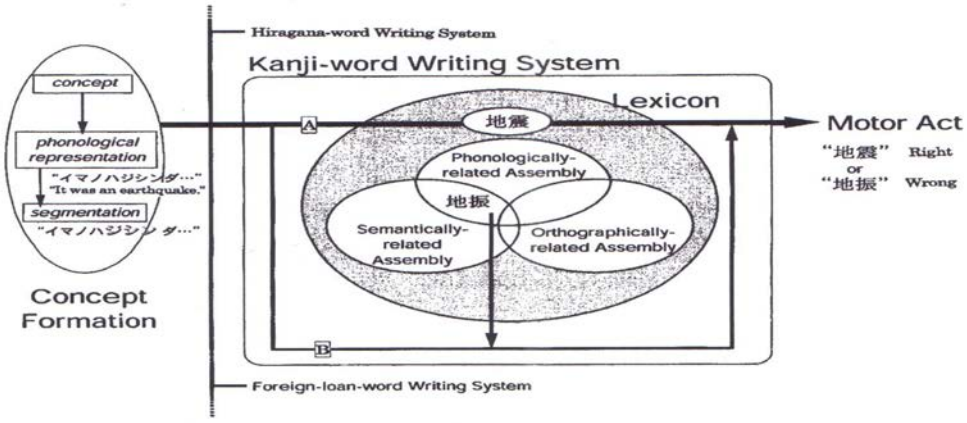

Figure 6 Cognitive model explaining the occurrence mechanism of phono-semantic type writing errors (Hatta et al., 1998:464)

The phonological type (substitution of target characters with homophonous characters) was most common among the Japanese students' errors, whereas the non-kanji (i.e. non-existing pseudokanji) type was predominant in Australian students' errors. Breakdown of non-kanji type errors made by the two groups indicated that kanji characters might be stored as radical assemblage in the Japanese students' mental lexicon, and as stroke assemblage in the Australian students' mental lexicon.

In order to identify the influential factors involved in the error generating process, Hatta et al. (2002) made a threefold comparison by adding a third group to the two participant groups in Hatta et al. (1998). The new group was Japanese seventh graders, whose errors were collected from their work sent to a correspondence course in which they participated. The Japanese college students represented advanced learners, the seventh graders intermediate learners, and the Australian students beginners.

The results showed different error generation patterns depending on the levels of kanji acquisition. The most common error type was non-kanji for the Australian students, orthographic (substitution by orthographically similar characters) for the seventh graders, and phonological for the Japanese college students. The results indicate that Australian students that are at the beginner's level have underdeveloped configurational awareness, Japanese seventh graders (intermediate-level learners) have developed limited radical awareness, while Japanese college students (advanced users) have well-developed radical awareness and phonologically inclined character retrieval aproaches.

Although the studies of Hatta et al. (1998, 2002) are greatly inspiring for the purposes of this study, they are not without limitations, namely:

A. use of non-identical data collection methods among the groups;

B. use of close equivalents of free composition (academic essays and course work) as a data source for two of the three groups, which resulted in a lack of target character setting and participants' possible avoidance of use of error-prone characters; and

C. lack of level matching between L2 and L1 participant groups, i.e. the absence of intermediate/advanced level L2 learners or beginner level L1 learners.

As a result, it is not clear if the observed differences were level-driven or due to the different conditions such as the learners' L1 and the data collection method.

Chikamatsu (2005) explored the issue of L2 kanji retrieval through a phenomenon called "tip-of-the-pen" (TOP) (in which the subject experiences a phenomenon described as "I think I know the kanji character, but cannot write it accurately, as if the character is stuck on the tip of the pen"). She compared L1 Japanese college students and American college students (L1 English) of Japanese at the intermediate level, using different sets of target characters that were of appropriate degrees of difficulty to induce TOP phenomenon for each group.

The subjects were asked to write two-character words based on given pronunciation and meaning, and then to provide their best guesses of the componential structure, stroke number and configuration of the target word, if they felt they were in a TOP-state. The occurrence rates of different TOP types (correct TOP, incorrect TOP with correct radical, incorrect TOP with incorrect radical, and non-TOP) were examined. In addition, error analysis was made, classifying errors into phonetic, graphic, semantic, compositional and contextual types.

In TOP analysis, the American subjects made significantly lower rates of correct TOP and incorrect TOP with correct radicals than the Japanese subjects, which implies that they were overconfident about their kanji production ability: being used to their alphabetic L1WS with configurationally simple graphemes and more transparent GSC, they felt they could spell the word when the pronunciation was given, but failed to produce a correct form, being oblivious to kanji's configurational details.

The American subjects' error analysis results were characterised by a relatively low rate of phonetic errors (which consisted mostly of substitution with characters sharing no phonological radical with the target character), a low rate of phonetic-graphic errors, which consisted of substitution with characters sharing the pronunciation and phonological radicals, and a high rate of compositional errors (substitution with the other constituent character of a compound for the target character, e.g. 究 for 研 from the compound 研究 "research"). These characteristics indicate the American subjects' underdeveloped phonological radical awareness, automatic interactive phonetic activation (the given pronunciation of the target character activating homophonous candidate characters in the mental lexicon), and unestablished orthographic-phonological linkage for each character of the compound.

Based on the differences observed in the two results, Chikamatsu (2005) pointed out the following characteristics of L2 learners:

A. large gap between recognition and production skills;

B. L1 orthographic transfer on L2 memory mechanism and retrieval strategy;

C. lack of intra-character structural awareness and decompositional ability with kanji;

D. Unestablished radical-based memory unit in the mental lexicon; and

E. Weak association of multiple character information in the mental lexicon.

The limitation of this study derives from the lack of level matching between the L2 and L1 groups. Because of the level difference, different sets of target characters had to be used, which resulted in an uncontrolled balance of features (semantic/phonological transparency, complexity, familiarity, etc.) between the two sets of characters.

3.3.4. Summary of kanji production studies

Table 14 is a summary of kanji production studies reviewed in subsections 3.3.1, 3.3.2 and 3.3.3.

Table 14 Summary of kanji production studies

Study	Subjects	Tested aspect	Experiment	Comparison
Studies with L1 subjects (subsection 3.3.1)				
Koike et al. (2003)	L1 Japanese children	Developmental stages of *hiragana* and kanji writing	(N/A)	(N/A)
	A. Pre-GSC stage (whole-word writing without GSC awareness) B. *Hiragana*-word stage (write words with basic *hiragana* symbols) C. Special mora word stage (write words in *hiragana* including special morae) D. Basic kanji stage (write basic kanji without radical awareness) E. Kanji expansion stage (write kanji with radical awareness)			
Implications	➢ Difficulty of special morae indicated by later mastery. ➢ Later development of radical awareness.			
BAASC (2007)	L1 Japanese children (Grades 1 - 7)	Writing ability of (*hiragana* and) kanji learned in previous year	Fill-in-the-blanks with target kanji to complete given phrase/sentence	Reading types (*on* vs. *kun*)
	• Higher error and blank rates in higher grades. • Difficulty in writing *on*-reading words (unfamiliar vocabulary). • Substitution with homophonous kanji increases at Grade 4			
Implications	➢ Increased workload due to quantity and complexity of characters for higher grades. ➢ Dependence on non-kanji linguistic skills. ➢ Sufficiently large stock of kanji homophones can be formed after learning a total of 640 characters.			

(continued)

Table 14 *Continued*

Study	Subjects	Tested aspect	Experiment	Comparison
Studies with L2 subjects (subsection 3.3.2)				
Kano et al. (1989)	L1 non-Japanese adults (novice) at 50-, 300- & 500-character levels	Kanji retrieval patterns	Free recall writing	Number of written characters Association patterns (semantic/ orthographic/ phonological)
	• Increase in number of written kanji for higher level • Association pattern shifts from semantic to orthographic • Orthographic resemblance type shifts from whole character to componential • Few cases of phonological association			
Implications	➢ Semantic association at early stage due to kanji introduction as categorised vocabulary. ➢ Development of configurational awareness, character decompositional ability and intra-character structural awareness after learning sufficient number of characters. ➢ Writing task without phonological cue triggers little phonological activation.			
Okita (2001)	L1 non-Japanese (L1WS logographic & alphabetic mix) adults (novice & intermediate mix)	Writing error occurrence patterns	Writing errors collected from compositions	Error types (orthographic incl. pseudokanji/ phonological/ semantic)
	• Absolute overall predominance of graphic errors (incl. pseudokanji and graphically similar characters) over phonological and semantic errors. • Learners with logographic L1WS made fewer graphic errors, most of which were L1-influenced stylistic errors. • All graphic errors were subtle inaccuracies rather than total anomalies.			
Implications	➢ Clear L1WS transfer observed among learners with logographic L1WS. ➢ Underdeveloped configurational awareness among learners with alphabetic L1WS			

(continued)

Table 14 *Continued*

<table>
<tr><th>Study</th><th>Subjects</th><th>Tested aspect</th><th>Experiment</th><th>Comparison</th></tr>
<tr><td colspan="5">Comparative studies with L1 & L2 subjects (subsection 3.3.3)</td></tr>
<tr><td rowspan="5">Hatta et al. (1998, 2002)</td><td>L1 English adults (novice)</td><td rowspan="3">Writing error occurrence patterns</td><td>Errors collected from kanji quizzes</td><td rowspan="3">Error types (phonological/ orthographic/ semantic/ non-kanji)</td></tr>
<tr><td>L1 Japanese 7th graders (intermediate)</td><td>Errors collected from course work</td></tr>
<tr><td>L1 Japanese adults (advanced)</td><td>Errors collected from essays</td></tr>
<tr><td colspan="4"><table>
<tr><td>Subject groups</td><td>Predominant type</td><td>Character storage</td></tr>
<tr><td>L1 English adults</td><td>Non-kanji</td><td>Stroke assemblage</td></tr>
<tr><td>L1 Japanese 7th graders</td><td>Orthographic</td><td>Radical assemblage (uncertain combination)</td></tr>
<tr><td>L1 Japanese adults</td><td>Phonological</td><td>Radical assemblage</td></tr>
</table></td></tr>
<tr><td>Implications</td><td colspan="4">➢ Different error occurrence patterns for different levels of kanji acquisition.
➢ L1 English learners (novice) have underdeveloped configurational awareness.
➢ Japanese 7th graders (intermediate) have developed limited radical awareness.
➢ Japanese adults (advanced) have well-developed radical awareness and phonologically inclined character retrieval approaches.</td></tr>
<tr><td rowspan="2">Chikamatsu (2005)</td><td>L1 English adults (intermediate)
L1 Japanese adults</td><td>Tip-of-the-pen phenomenon
Writing error occurrence patterns</td><td>Writing two-kanji words based on given pronunciation and meaning</td><td>TOP state (correct/ incorrect/non)
Error types (phonetic/ graphic/ semantic/ compositional/ contextual)</td></tr>
<tr><td colspan="4">L1 English subjects had
• lower rates of correct TOP and incorrect TOP with correct radical
• low rate of phonetic errors and phonetic-graphic errors
• high rate of compositional errors</td></tr>
<tr><td>Implications</td><td colspan="4">➢ L2 learners' large gap between recognition and production skills
➢ L1 orthographic transfer on L2 memory mechanism and retrieval strategy
➢ L2 learners' lack of intra-character structural awareness and decompositional ability
➢ Unestablished radical-based memory unit in L2 mental lexicon
➢ Weak association of multiple character information in L2 mental lexicon</td></tr>
</table>

The implications of the previous kanji production studies reviewed above in Table 14 can be summarised and reorganised according to learner groups and aspects of kanji competence, as listed in Table 15:

Table 15 Summary of implications of kanji production studies

Subjects	Orientation	Ability	Difficulty
L1 children	Dependence on non-kanji linguistic skills (BAASC, 2007)	Later development (still limited at Grade 7) of radical awareness (Koike et al., 2003; Hatta et al. 2002)	Special morae (Koike et al., 2003)
		Large enough kanji homophone stock at 640 character level (BAASC, 2007)	Increased workload due to quantity and complexity of characters (BAASC, 2007)
L1 adults	Phonologically inclined character retrieval (Hatta, et al., 1998; 2002)	Well-developed radical awareness (Hatta, et al., 1998; 2002)	
L2 novice/intermediate with mixed L1WS	Semantic association at early stage (Kano et al., 1989)	Development of configurational awareness, character decompositional ability and intra-character structural awareness after learning sufficient number of characters (Kano et al., 1989)	Clear L1WS transfer among learners with logographic L1WS (Okita, 2001)
	Little phonological activation in writing task without phonological cue (Kano et al., 1989)		
L2 novice/intermediate with alphabetic L1WS	L1 orthographic transfer on L2 memory mechanism and retrieval strategy (Chikamatsu, 2005)	Underdeveloped configurational awareness (Okita, 2001; Hatta, et al., 1998; 2002)	
		Lack of intra-character structural awareness and decompositional ability (Chikamatsu, 2005)	
		Large gap between recognition and production skills (Chikamatsu, 2005)	

3.4. Phonological studies regarding Swedish learners of Japanese

This section will focus on a review of phonological studies regarding difficulties experienced by Swedish learners of Japanese.

Japanese special morae (long vowels, geminate consonants and moraic nasals) are known to be difficult to acquire for L1 learners (Japanese children) for their relative complexity (BAASC, 2007; Toda, 2003); they do not bear an accent kernel, do not take word-initial position, and have various allophones, and they are difficult to acquire even for L2 learners (Toda, 2003). Phonology is one of the areas of L2 learning where L1 transfer is most distinctive. In addition to the abovementioned obscurity, the differences between the phonological structures of Japanese and the learners' L1 can have a negative influence on L2 learners' acquisition of special morae (Toda, 2003).

Both Swedish and Japanese are quantity languages that have a phonological long/short contrast (or "quantity") in both vowel and consonant (Elert, 1964 for Swedish; Han, 1965 for Japanese; Inoue, 2009, for both). Swedish is a stress language that also shows a phonological length contrast in both vowels and consonants, but Swedish quantity is present only in stressed syllables, in which a vowel/consonant length contrast manifests itself complementarily. The vowel is long in an open syllable or when followed by a short consonant, and it is short when followed by a long consonant (either in the form of /V:(C)/ or /VC:/) (Inoue, 2008). For example, the vowel *e* [e:] is long in the words *ge* [je:] ("give") and *get* [je:t] ("goat"), but short in the word *gett* [jet:] ("given"). Although Japanese tolerates CVCV-type of words with various combinations of long and short consonants and vowels (e.g. 外 /soto/ "outside," ソート /so:to/ "sort," 相当 /so:to:/ "considerably," そっと /sot:o/ "quietly," そうっと /so:t:o/ "stealthily," 卒倒 /sot:o:/ "faint"), Swedish allows only /CV:CV/ (long vowel & short consonant) and /CVC:V/ (short vowel & long consonant) word-medial combinations such as *mata* [ma:ta] "to feed" and *matta* [mat:a] "mat." In her experimental research, Inoue (2008) reported that half of the Swedish learners' pronunciation of Japanese /CVCV/ words (e.g. a family/place name 瀬戸 /seto/) was perceived by Japanese native speakers either as /CV:CV/ (/se:to/ 生徒 "pupil") or /CVC:V/ (/set:o/ セット "set"), suggesting L1 transfer of Swedish complementarity of the VC-sequence.

When there are problems in L2 learners' pronunciations of certain sounds of the target language, there are three possible stages for the cause: (i) lack of phonetic knowledge due to lack of the L2 phonetic contrast in the L1; (ii) failed recognition (and pronunciation); and (iii) failed articulation

(recognisable but not pronounceable) (Toda, 2003). The case of Inoue (2008) corresponds to (iii), for the failed articulation is due to the difference in vowel-consonant length contrast rules between Swedish and Japanese. In the case of Swedish learners' transcription errors involving long vowels and geminate consonants, there is another stage: (iv) failed transcription due to incomplete acquisition of the L2 orthographic rules (pronounceable but not transcribable). In addition to the difficulty in vowel-consonant length contrast, Swedish learners often have trouble differentiating consonant pairs /s/ and /z/, /sh/ and /j/, and /sh/ and /ch/ as well.[7] Since the Swedish phonetic system lacks the latter consonant in each of the above pairs (/z/, /j/ and /ch/) (Yamashita, 1990), the cause of problems in transcribing these consonants is deemed to lie in stage (i).

Notes

[6] No orthography is 100% transparent or 100% opaque, except for phonetic alphabets artificially developed to describe speech sounds such as IPA (International Phonetic Alphabet) (Cook & Bassetti, 2005). The degree of phonological transparency is difficult to indicate as a numerical value, and therefore, it is expressed only relatively in comparison with other orthographies.

[7] Since the focus of this study is not on minute phonetic differences in the Japanese speech sounds, Japanese pronunciations will be described in this dissertation with phonological representations of phonemes based on the Hepburn system of Romanisation (ヘボン式ローマ字) (e.g. *sh*, *j* and *y*) instead of phonetic representations (e.g. [ɕ], [ʑ] and [j]). When Japanese pronunciations are stated as readings of kanji words and characters without particular focus on their phonological values, they will be indicated in italicised Hepburn Romanisation (e.g. 山 *yama*).

4. EXPERIMENTS

Chapter 4 will give descriptions of the experiments used in this study. First, an overview of the experiments will be given in section 4.1. Section 4.2 will clarify the essential points in designing the experiments. Sections 4.3 and 4.4 will describe the experiments for the 240-character level (LV240) and 800-character level (LV800) respectively.

4.1. Overview of the experiments

Tables 16 and 17 are overviews of the experiments conducted for this study. Table 16 summarises the details of the participants, who will be described in subsections 4.3.1 and 4.4.1, while Table 17 presents the experimental material and procedures described in subsections 4.3.2, 4.3.3, 4.4.2 and 4.4.3.

Table 16 Summarised descriptions of participant groups

Group	L1	Learned characters	Finished class hours	Educational institution	Ages	Number of participants
SJ240	Swedish	approx. 240	approx. 250	university	19-30	49
NJ240	Japanese	approx. 240	approx. 530	primary school	7-8 (Grade 2)	191
SJ800	Swedish	over 800	over 700	university	20-30	20
NJ800	Japanese	approx. 800	approx. 1,200	primary school	10-11 (Grade 5)	135

Table 17 Summarised descriptions of kanji test specifications

Level	Test	Target kanji	Number of blanks to fill in	Number of sentences/phrases	Time limit
LV240	Reading	90 characters (113 readings) that are commonly included in the first 240 characters for SJ240 and the 160 Grade 2 School Year characters	113	46	15 min.
	Writing		113	46	20 min.
LV800	Reading	118 characters (143 readings) that are commonly included in the 1,023 Level 2 JLPT characters and the 185 Grade 5 School Year characters	143	98	15 min.
	Writing		143	98	20 min.

4.2. General design

This section will clarify the essential points in the general design of the experiments. Subsection 4.2.1 will organise the variables and refer to statistical treatment, subsection 4.2.2 will specify the type and format of the experiments, and subsection 4.2.3 will focus on the concept of level matching.

4.2.1. Variables and statistical treatment

This study aims to investigate the characteristics of Swedish learners' kanji retrieval approaches and difficulties in reading and writing in the developmental process of kanji learning. It attempts to achieve this objective through the analysis of kanji reading and writing errors of novice and advanced groups of both Swedish and Japanese learners.

In designing the experiments for the error analysis that will constitute the major part of this study, reference was made to three of the previous studies discussed in Chapter 3 (Hatta et al., 2002; Chikamatsu, 2005; and BAASC, 2007). For the purpose of providing a comprehensive description of the characteristics of Swedish learneres' kanji learning process, however, changes, additions and adjustments have been made to the methods and conditions of these earlier investigations. Accordingly, the experiments for this study should, inter alia:

A. compare Swedish and Japanese learners to illuminate the Swedish learners' characteristics;

B. examine both reading and writing skills at two different levels for a comprehensive investigation of the developmental process of the learner groups;

C. analyse errors collected from the data source in test format, in which target characters can be set and the participants cannot avoid using error-prone characters; and

D. use cognitive error type classification, which can effectively describe learners' retrieval approaches and difficulties.

In order to meet the above specifications, the following conditions should be matched:

A. the data collection method and the error type classifications between skills (reading/writing), levels (LV240/LV800) and L1s of the groups (Swedish/Japanese);

B. the level settings of the Swedish and Japanese groups; and

C. the target characters (and in consequence the character-related features as well) at each level.

With these points in mind, the experiments were designed to verify each aspect of the error occurrence patterns compared, by dismissing the null hypothesis "there is no significant difference in error type occurrence patterns between the compared groups." For this verification, the level of significance was set at a total of 0.05 (5%), which is the figure prevalent in language acquisition research in general. Table 18 is a list of independent, dependent and intervening variables of the experiments:

Table 18 Independent, dependent and intervening variables of the experiments

Variables \ Tests		LV240 reading test	LV240 writing test	LV800 reading test	LV800 writing test
Independent variables	**Learners' L1**	Swedish (subject) vs. Japanese (control)			
	Learners' L1WS	alphabetic (subject) vs. syllabic-logographic (control)			
	Learners' level	novice (240 character level)		advanced (800 character level)	
	Target characters	90 novice level characters learned by both participant groups		118 advanced level characters learned by both participant groups	
	Test procedure	• written examination • target character words embedded in phrases or sentences • parallel tasks for reading & writing			
Dependent variables		Error type occurrence patterns and tendencies			
Intervening variables	**Intrinsic**	L1/L1WS transfer, cognitive abilities, age, motivation, etc.			
	Extrinsic	(1) duration, method and environment of study (2) extent of exposure to kanji (3) education style, learning order of verbal and literal skills			

4.2.2. Type and format of the experiments

Errors must be collected from written material, since handwriting errors of kanji are to be examined. Reading errors must be collected from written material of the same style, for the purpose of matching the collecting method for both skills. The written material must be in test format, in which the participants are compelled to confront the target characters. If free writing such as essays and letters were used, the writers could avoid

writing the characters of which they feel uncertain, thereby avoiding making errors. Besides, collecting reading errors from free writing would be difficult, if not impossible. Furthermore, with written examinations, a large quantity of data can be collected on a single occasion.

In the written examinations, the target characters are to be presented within a word that is embedded in a short and simple sentence or phrase, thereby creating the context that will specify the meaning and eliminate homophonous characters. In this way, patterns, methodologies and difficulties pertaining to reading and writing of kanji can be investigated in a more practical task setting than in experiments concentrating on single characters or words independent of context.

The task will be fill-in-the blanks questions for both reading and writing. The sentences and phrases will be presented in a proper mixture of kanji and *kana*. *Furigana* (pronunciation of kanji shown in *hiragana*) will be provided for non-target kanji, so the participants have good verbal access to the context, regardless of their knowledge of non-target characters. Both participant groups are familiar with this type of task, since it is often used in both L2 and L1 kanji education for workbook exercises and written tests, etc. In fact, this task type and manner of presentation are the same as those used in the BAASC (2007).

Familiarity of the task type is an aspect that should not be neglected, especially in experiments comparing adults and children. If unfamiliar tasks are used, there is a risk that children's performances will suffer unjustifiably because they often have difficulty in understanding the intention of the examination questions and fail to tackle the task properly (Shirahata, Wakabayashi & Muranoi, 2010).

4.2.3. Level matching

Thus far, kanji-related studies comparing L2 and L1 learner groups have dealt with comparisons between L2 learners who have learned no more than a few hundred characters and L1 learners (users) who have acquired 1,000 to 2,000 characters. In other words, they have involved comparisons between the novice and the advanced, and the observed differences may well have been based on the level difference rather than the difference in the L1WS background.

In order to properly investigate the influence of L1WS background, L2 and L1 groups must be compared at the same developmental stage. Since the number of learned characters is a basic indicator of the level of kanji learning, level matching can be achieved by matching the number of characters the two groups have learned. For the purposes of the present

research, the term "learned characters" means kanji characters that had been introduced in class and practised by the learners prior to the time of the experiments. Learners are nevertheless expected to make errors in reading and/or writing such characters or even to totally fail to recall how to read or write them. Learned characters, therefore, are not necessarily "mastered" or "acquired" characters.

Conventional kanji instructions start with characters that are simple in structure, basic in meaning and/or used frequently, and move on to more complicated, less basic and/or less frequently used ones. Therefore, if there are two groups of learners who have learned about the same number of characters, the two sets of characters these groups have learned are most likely to be largely overlapping, even if learned at different educational institutions using different textbooks. Consequently, both groups can be estimated to have comparable stocks of phonologically identical, orthographically similar and/or semantically related characters in their mental lexicon.

On the other hand, a substantial difference in the number of learned characters affects the cognitive mapping within the mental lexicon to a great extent. An example can be taken in terms of the growth of the stock of phonologically identical characters: the larger the stock, the greater the possibility of making an error in which the intended character is erroneously replaced with a homophonous character. For instance, Japanese learners learn in the first grade at the primary school a total of 80 characters, only one of which is pronounced *kō* (校 "school").[8] When they have learned a cumulative total of 1,006 characters at the end of the sixth grade, the characters that can be pronounced *kō* amount to 31. Therefore, the possibility of first graders erroneously writing a homophonous character when the intended character is 校 (*kō*) is virtually none, whereas sixth graders, with thirty homophonous characters stored in their mental lexicon, have an arguably higher risk of making such an error when trying to write the character 効 (*kō* "effect").

Another important merit of level matching is that it makes testing of the participant groups on an identical set of target characters appropriate, because a number of commonly learned characters can be found in their respective lists of learned characters. Use of an identical set of target characters ensures that the two groups deal with the same intra-character features and extra-character properties.

If different sets of target characters were to be used, the attributes of the characters, such as morphological features, semantic or phonological transparency, orthographic complexity and assemblage, which can affect the tendency to produce a particular error type, would be nearly impossible

to equalise. Varied attributes of the characters can be a cause for different error type occurrence tendencies. Imagine the case in which the compared groups had different sets of target characters of unbalanced orthographic complexity, one group having only characters with at least 15 strokes (such as 議, 燃, 織, 潔, 難, 機, etc.), and the other level-matched group only characters with a maximum of 5 strokes (such as 人, 川, 口, 山, 木, 田, etc.). If the former has made a higher percentage of pseudokanji type of errors (i.e. writing an illegitimate kanji-like figure resembling the target character with a missing or additional stroke or a wrong combination of components) than the latter, it can be argued that this tendency is not a result of the inherent characteristics of the participant groups but of the disproportionate selection of characters with certain attributes.

The use of an identical set of target characters enables the use of an identical set of sentences and phrases as well, which provides perfect control over the extra-character properties (e.g. the character's intra-word position and the part of speech of the word involving the character), thereby preventing such properties from affecting the experimental results.

Although level matching is the core condition, using the same set of target characters for different levels would be inappropriate. Even though the advanced learners have in principle learned all the characters novice learners have learned, those characters should not be shared as target characters by both levels. The error-inducing potentiality of such a list of characters may be adequate for the novice learners, but would be quite insufficient for the advanced learners; the latter would make very few errors in such elementary characters. On the other hand, if characters that are difficult enough for the advanced learners were used, the test would be well beyond the knowledge of the novice learners and they would be compelled to leave the answers blank, failing to produce errors nonetheless.

Since level matching has priority over all other conditions in this experiment, it is inevitable to compromise on the other normally matched experimental conditions such as the participant groups' age ranges and their educational backgrounds. For the purposes of this research, Swedish participants must have learned at least a couple of hundred kanji characters, which is a large enough number to provide a statistically significant number of target characters, but not exceeding one thousand kanji characters, because Swedish learners who know over a thousand characters are so few that it would be extremely difficult to form a statistically large enough group of such learners. Swedish learners of this knowledge range can be found only at the university level, but their Japanese equivalents are primary school pupils.

Taking these matters into consideration, this study compared Swedish university students at the novice and advanced levels and Japanese schoolchildren at the respective corresponding levels. The number of learned characters for the novice level was set at 240, which corresponds to the mid-beginner's level for L2 learners in general and to the second semester in a full-time university course in Sweden. This level corresponds to the end of second grade in the Japanese primary school. For the advanced level, the figure was set at 800 characters and over. Since the number of advanced Swedish learners of Japanese is very limited and those who know over 1,000 characters are even fewer,[9] the 800-character level is a reasonable compromise: it is regarded as (early) advanced level; and the number of learners at that level who can be practicably gathered for the experiments is large enough for statistical analysis. The corresponding level at the Japanese primary school is the fifth grade, at the end of which 825 characters have been learned.

In sections 4.3 and 4.4, the participants, materials and procedures of the experiments will be described according to level.

4.3. 240-character level (LV240) experiments

4.3.1. Participants

Forty-nine Swedish university students and 191 Japanese schoolchildren participated in the experiment. All the Swedish participants ("SJ240") were native Swedish speakers enrolled in the second semester course of the first year of a degree in Japanese at two different universities in Sweden (32 participants from a large western university and 17 from a large southern university). Their ages ranged from 19 to 30, but 81.6% of them were aged 19 to 23. At the time of the experiment they had had approximately six months of formal education in the Japanese language, finished a total of approximately 250 class hours in Japanese offered in the respective university course and learned approximately 240 kanji characters.

All the Japanese participants ("NJ240") were native speakers of Japanese enrolled in the second grade of three Japanese primary schools (76 participants from a university-attached school in eastern Japan, 41 from a municipal school in central Japan, and 74 from a municipal school in southern Japan). Their ages ranged from seven to eight, but 91.7% of them were aged eight. At the time of the experiment they had had nearly two years of formal education in the Japanese language, finished a total of approximately 530 class hours in the subject of the Japanese language offered in the first and second grades, and learned nearly 240 kanji characters (Ministry of Education, Culture, Sports, Science and Technology, 1998).[10]

4.3.2. Materials

Ninety characters (with 113 readings) were used as target kanji for the experiment. Some of the characters have multiple readings, and hence the number of readings exceeds the number of characters. The 90 target characters (see Appendix 1) were selected from the characters commonly included in (i) the first 240 kanji characters learned in SJ240's respective courses (extracted from Banno et al., 2009; Banno et al., 1999a; and Banno et al., 1999b) and (ii) the 160 characters in the List of Kanji by School Year (学年別漢字配当表 *Gakunenbetsu Kanji Haitōhyō*) (Ministry of Education, Culture, Sports, Science and Technology-Japan, 1989) for Grade 2 of Japanese primary schools. The target kanji characters/readings were presented as words or parts of words embedded in short sentences and phrases, which provided the context to determine the meaning and reading of the words.

4.3.3. Procedure

The written examinations were conducted in class-based groups at each institution on single occasions between February and April 2011. The task in the reading test was to fill in the round brackets with the reading of the target kanji in the given word and context. The reading test (see Appendix 3) consisted of 46 short, simple sentences and phrases including the 90 target characters and 113 brackets to be filled in with appropriate readings in *hiragana* within the 15 minutes' time limit. Each sentence/phrase included one to three target characters. The sentences/phrases were written in a proper mixture of kanji, *hiragana* and *katakana*, and the readings of non-target kanji were provided as *furigana* (printed in a small font in *hiragana* above the kanji).

The writing test (see Appendix 5) consisted of the same 46 sentences/phrases, but the target characters were replaced with 113 boxes. The task was to fill in the boxes based on the readings given above the boxes as *furigana*, thereby completing the given sentences/phrases. The readings of non-target kanji were provided as *furigana*, following the style in the reading test.

Consideration for internal validity had been taken so that the results would not be affected by possible differences in task familiarity between the groups or in handwriting speed between the skills. Both SJ240 and NJ240 were familiar with such tasks through in-course activities, such as workbook exercises and written examinations. Although the two tests had parallel tasks of filling in the same number of blanks embedded in the identical set of sentences/phrases (differing only in the manner of

presentation and completion), the writing test was given a longer time limit of 20 minutes. This was planned to account for the fact that writing kanji requires longer time than writing *hiragana* (Tainaka, 1979).

4.4. 800-character level (LV800) experiments

4.4.1. Participants

Twenty Swedish university students and 135 Japanese schoolchildren participated in the experiment. All the Swedish participants ("SJ800") were native Swedish speakers. They were either enrolled in the second or third year of a degree in Japanese or recently graduated with a degree in Japanese from two different universities in Sweden (the same two universities as for SJ240, 11 participants from the western university and nine from the southern university). Their ages ranged from 20 to 30, but 75% of them were aged 21 to 25. At the time of the experiment they had had approximately two to four years of formal education in the Japanese language, had finished a total of at least 700 class hours in Japanese offered in the respective university course and had learned at least 800 kanji characters.

All the Japanese participants ("NJ800") were native speakers of Japanese enrolled in the fifth grade of three Japanese primary schools (the same three schools as for NJ240, 77 participants from the school in eastern Japan, 31 from central Japan, and 27 from southern Japan). Their ages ranged from ten to eleven, but 93.4% were aged eleven. At the time of the experiment they had had nearly five years of formal education in the Japanese language, finished a total of almost 1,200 class hours in the subject of the Japanese language offered in the first to fifth grades, and learned approximately 800 kanji characters (Ministry of Education, Culture, Sports, Science and Technology, 1998).[11]

4.4.2. Materials

One hundred and eighteen characters (with 143 readings) were used as target kanji for the experiment. Some of the characters have multiple readings, and hence the number of readings exceeds the number of characters. The 118 target characters (see Appendix 2) were selected from the characters commonly included in (i) the cumulative total of 1,023 kanji characters for Level 2 of the Japanese Language Proficiency Test[12]

(extracted from Alc Japanese Language Publishing Editors, 1994), and (ii) the 185 characters in the List of Kanji by School Year (Ministry of Education, Culture, Sports, Science and Technology-Japan, 1989) for Grade 5 of Japanese primary schools. The target kanji characters/readings were presented as words or parts of words embedded in short sentences and phrases, in conformity with the manner of presentation used in the experiment for LV240.

4.4.3. Procedure

The written examinations were conducted at each institution on single occasions, in class-based groups for NJ800 and in ad hoc groups for SJ800, between February and April 2011. Although the manner of the task in the LV800 reading test was identical with that in the LV240 reading test, LV800 had more target characters and sentence/phrases to work on. The LV800 reading test (see Appendix 4) consisted of 98 short, simple sentences and phrases including the 118 target characters, and 143 brackets to be filled in with appropriate readings in *hiragana* within the 15 minutes' time limit. Each sentence/phrase included one to three target characters. The sentences/phrases were written in the same manner as in the LV240 reading test, in a kanji-*kana* mixture with *furigana* for the non-target kanji.

The writing test (see Appendix 6) consisted of the same 98 sentences/phrases, although the target characters were replaced with 143 boxes. The task was to fill in the boxes based on the readings given above the boxes as *furigana*, thereby completing the given sentences/phrases. The readings of non-target kanji were provided as *furigana*, following the style in the reading test.

Consideration for internal validity had been taken in terms of task familiarity and handwriting speed, just as was the case for LV240. Both SJ800 and NJ800 were familiar with the task type, and the writing test was given the time limit of 20 minutes as against 15 minutes for the reading test. In spite of the greater number of tasks for LV800, the same examination time limits (15 minutes for reading and 20 minutes for writing) were given as for LV240, because the LV800 learners were expected to have developed greater handwriting speed than LV240 learners.

Notes

[8] The character 口 ("mouth") is also learned in the first grade, but the pronunciation learned then is the *kun*-reading *kuchi,* and the *on*-reading *kō* is not learned until a later grade.

[9] As of 2009, there were no more than 990 learners of Japanese enrolled in Swedish higher education (Japan Foundation, 2011), a great majority of whom were novice level learners.

[10] Since the tests were carried out three to four weeks prior to the end of the school year, the figures for the completed class hours and learned characters are only approximations of the exact total figures for Grades 1 and 2 (552 class hours and 240 characters) according to the 1998 enforcement regulations of the School Education Law (Ministry of Education, Culture, Sports, Science and Technology, 1998) then in effect.

[11] Since the tests were carried out three to four weeks prior to the end of the school year, the figures for the completed class hours and learned characters are only approximations of the exact total figures for Grades 1 to 5 (1,202 class hours and 825 characters) according to the 1998 enforcement regulations of the School Education Law (Ministry of Education, Culture, Sports, Science and Technology, 1998) then in effect.

[12] This was the format prior to the change in 2010, in compliance with which the participants had studied.

5. ERROR CLASSIFICATIONS

This chapter will give comprehensive descriptions of kanji error classifications in four sections: 5.1 will provide an overview of reading and writing error classifications, 5.2 will explain the error types for reading and 5.3 for writing, and both will be summarised and exemplified in 5.4.

5.1. Overview of the error classifications

Table 19 is a brief overview of the error classifications. The reading and writing errors are based on the same basic classification principles, which enables comparisons to be made between the two skills.

Table 19 Overview of kanji reading and writing error classifications

Types	Reading error descriptions	Writing error descriptions
Phonological	Application of an inappropriate/ wrong reading of a character.	Substitution with a phonologically identical character.
Orthographic	Misreading of a character as an orthographically similar character	Substitution with
Semantic	Misreading of a character as a semantically related character	Substitution with
Circumstantial	Misreading of a character as a circumstantially associated character	Substitution with
Pseudokanji	N/A (for writing errors only).	Substitution with a pseudokanji character that deviates from any existing character.
Others	Errors not falling into any of the above categories.	

Section 5.2 will give detailed descriptions of each reading error type, and section 5.3 of each writing error type.

5.2. Classification of kanji reading error types

Based on the concepts discussed in Chapter 2, reading errors obtained in the experiments were classified into four categories: **phonological**, **orthographic**, **semantic** and **circumstantial**. Definitions, characteristics and subtypes (if any) will be described in subsections 5.2.1 to 5.2.4.

5.2.1. Phonological reading errors

Phonological reading errors are caused by the application of an inappropriate/incorrect reading of a kanji character. They are divided

further into three subtypes, namely, (a) **transcription errors**, (b) **alternative reading**, and (c) **component-based analogy**, which will be explained and exemplified below.

(a) **Transcription errors** are incorrect *hiragana* transcriptions of kanji readings as a result of either an interchange of phonemes that are difficult for learners to distinguish, or erroneous transcriptions of phonologically less transparent morae (see Note 2 in Chapter 2 for a definition and examples of morae). Examples of the former are mistaking the long vowel ***o:*** for the short ***o***, or ***sh*** ([ɕ]) for ***ch*** ([tɕ]); and of the latter are phonemes with multiple grapheme-sound correspondence, such as a long vowel ***o:*** corresponding to the different *hiragana* sequences おう and おお. This type of error is based on an incorrect memory of the reading of the target character due to limited phonological awareness, and/or incorrect transcription as a consequence of an incomplete knowledge of the *kana* orthography.

(b) An **alternative reading** is a misapplication of one of the multiple readings of the target character. For example, *bu* (ぶ) is a correct reading of the character 武 as in the word 武士 ***bu****-shi* ("samurai"), but it is incorrect when the word is 武者 ***mu****-sha* ("warrior"). Making this type of error indicates that either the learners know only one reading of the target character and apply it whenever the character is used, or they have knowledge of the multiple readings without proper awareness of the circumstantial constraints (i.e. which reading should be applied in what circumstance).

(c) **Component-based analogy** is an overgeneralisation of analogy of the character's reading based on its phonological radical (e.g. 海 *kai* is misread as *mai*, assuming complete phonological consistency with its phonological radical 毎 *mai*) or semantic radical-phonological radical confusion (e.g. 規 *ki* is read as *ken*, mistaking the semantic radical 見 *ken* for a phonological radical and assuming its complete phonological consistency with the whole character), as explained in section 2.7. This type of error can be made by learners who (i) lack or fail to recollect the phonological knowledge of the character as a whole, (ii) have the ability to decompose the character into components and have phonological knowledge of such components, (iii) tend to seek intra-character phonological clues rather than circumstantial clues, and (iv) have imprecise knowledge of character-radical *on*-reading consistency.

The shared feature of these three subtypes is that the phonological aspect of kanji is disconnected from the orthographic and semantic aspects in one way or another. However, they differ in the unit of error occurrence: **transcription errors** are phoneme-/mora-based, **alternative readings** are

character-based, as they deal with multiple readings of a whole character, while **component-based analogies** decompose a character into components.

5.2.2. Orthographic reading errors

An **orthographic** reading error is a misapplication of (one of) the reading(s) of a character that is confusingly similar in shape, configuration and/or assemblage (but not a visually identical character, since kanji characters are, with very rare exceptions, visually different with each other). For example, the character 西 in 西口 ***nishi***-*guchi* ("**west** exit") was erroneously read as ***yon***-*guchi*, misapplying one of the readings of the character 四 ("four") (target 西 vs. error 四); the character 衛 in 衛星 ***ei***-*sei* ("satellite") was misread as ***machi***-*sei*, misapplying one of the readings of the character 街 ("town") (target 衛 vs. error 街). The term used in this classification is "orthographic" rather than "graphic," because formation of a kanji character is not a mere composition of graphic items. A kanji character is a square configuration of intra-character components, which in turn are conventional arrangements of strokes. Therefore, they should be compared to orthography, or to how words are conventionally spelled.

Occurrence of this type of error implies an orthographic orientation of the retrieval process (activation of the "form" assembly of the target character in the mental lexicon). In addition, it indicates that learners making this type of error process kanji on a whole-character basis, but with uncertainty in orthographic details of the target character. Although they have correct knowledge of the reading of the mistaken character, they are oblivious of circumstantial invalidity (the reading of the mistaken character leads to the creation of nonsensical words). In the abovementioned examples of 西口 ***nishi***-*guchi* /四口 ***yon***-*guchi* and 衛星 ***ei***-*sei* /街星 ***machi***-*sei*, the learners who made these errors were oblivious to the fact that there are no such words as ***yon***-*guchi* or ***machi***-*sei*.

5.2.3. Semantic reading errors

As kanji are often referred to as ideograms, each character represents a certain concept/meaning, each of which takes a different form. Therefore, there are very few characters sharing exactly the same meaning, with the exception of historical and geographical variations of the same character, as mentioned in section 2.3 with the example of 萬 and 万, both representing the word *man* ("ten thousand"). Characters that are synonymous with each other are rather limited, too, in contrast with the abundance of phonologically identical or orthographically similar characters. Consequently, it is deemed appropriate to set a somewhat loose

precondition of "relatedness" (rather than the narrower "identity" or "similarity") for an error to be classified as semantic.

In the light of the reflection above, a **semantic** reading error has been defined as misapplication of (one of) the reading(s) of a character that is semantically related to the target character. Characters that are regarded as being semantically related with each other include: synonyms (e.g. 全 "all" vs. 総 "total"); antonyms (e.g. 強 "strong" vs. 弱 "weak"); and terms belonging to the same semantic field, such as seasons, directions and kinship. Within the same category, some term-pairs have an antonymous relationship (e.g. 夏/冬 "summer/winter," 東/西 "east/west" and 親/子 "parent/child"), but others are not quite opposite of each other (e.g. 秋/冬 "autumn/winter," 東/北 "east/north" and 母/兄 "mother/elder brother").

In an example of a semantic reading error in which the character 首 *kubi* ("neck") was misread as *nodo* "throat", the grapheme-meaning correspondence "首-neck" (as in (1) in Figure 7) must have been activated but not the grapheme-sound correspondence "首-*kubi*" ((2) in Figure 7). The meaning-sound correspondence "neck-*kubi*"((3) in Figure 7) was therefore not established, but the word *nodo* for "throat," which is synonymous with "neck," was selected instead. In order to make this type of error, learners have to know the word *nodo* (and its *hiragana* transcription のど), but not necessarily how it is written in kanji (喉). This type of error is characterised by a semantic approach to retrieval, based on the whole character as the processing unit. It also implies the use of the direct access route from the orthographic representation to the semantic representation.

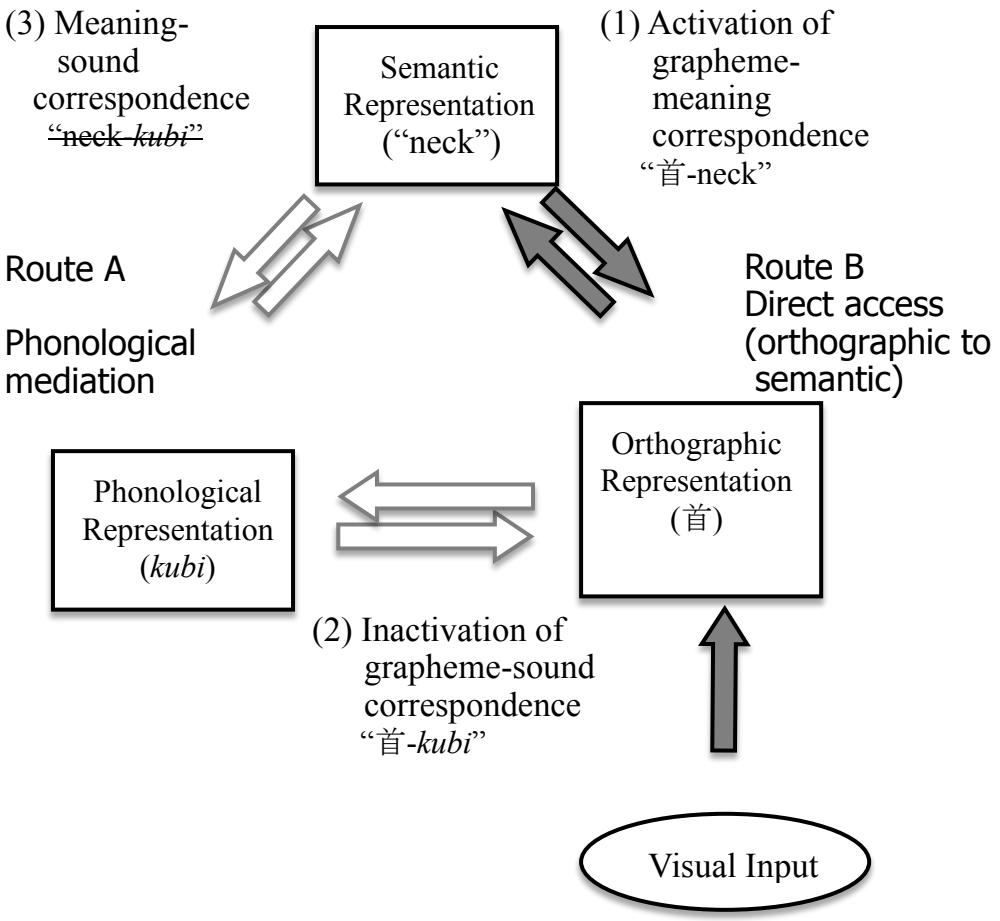

Figure 7 Dual access model for access to semantic information (Based on a model by Kadota, 1998)

As exemplified in the cognitive model of the character 花 in the mental lexicon (Figure 1 in section 2.3), the semantically related assembly of a character is usually the smallest (with the fewest items) of the three assemblies, and therefore provides the fewest number of candidate characters that can erroneously replace the target character. The error-inducing factor of this category, if any, lies in the fact that semantically related characters are often introduced simultaneously. As a result, learners get easily confused, being uncertain as to which of the characters is to be bound with which of the related concepts. For instance, characters denoting closely related concepts such as seasons or directions are often introduced in the same chapter of a textbook: the directional characters 東西南北 ("east/west/south/north") are introduced in Lesson 3 and the seasonal characters 春夏秋冬 ("spring/summer/autumn/winter") in Lesson 9 of Banno et al. (2009). In Banno et al. (1999), the four directions are

introduced in Lesson 6 and the seasons in Lesson 19, with the exception of 夏 (“summer”), which makes its first appearance in Lesson 15.

5.2.4. Circumstantial reading errors

A **circumstantial** reading error is a misapplication of (one of) the reading(s) of a circumstantially associated character, i.e. a character that is applicable based on **compound** constituents, **inflectional** endings or **contextual** congruence, as explained in section 2.5.

In order to make such a misapplication, one needs to be rather well versed in spoken Japanese: a sizeable vocabulary is required for compound-based substitutions, a decent knowledge of grammar for the inflection-based ones and adequate reading comprehension for the context-based ones. The implications of this type of error are general weakness in all three cognitive aspects – form, sound and meaning – of the knowledge of kanji and dependence on circum-character knowledge to compensate for such weakness. It also indicates a circumstantial approach to retrieval from the mental lexicon and kanji processing on a word basis rather than on a character or component basis.

5.3. Classification of kanji writing error types

Corresponding to the reading errors, writing errors were also classified into categories based on the concepts discussed in Chapter 2, namely, **phonological**, **orthographic**, **semantic** and **circumstantial**. In addition, writing error classification requires a fifth and writing-specific category of **pseudokanji**, i.e. production of a pseudocharacter configurationally similar to but deviating from existing kanji. These are common errors in writing, but by nature have no reading counterpart. Definitions, characteristics and subtypes (if any) of these categories will be described in subsections 5.3.1 to 5.3.5.

5.3.1. Phonological writing errors

There are a great number of homophonous characters. In fact, the top four most common kanji readings, namely *kō*, *shō*, *shi* and *kan*, which are represented by 64, 64, 48 and 45 characters respectively, amount to a total of 221 characters or over 11% of the 1,945 Jōyō Kanji characters. A **phonological** writing error is a substitution of the target character with a homophonous (phonologically identical) character. For example, the adjective *akarui* (“bright”) should be written as 明るい, using the character 明 which reads *aka* and means “bright” for the stem, and transcribing the

adjectival inflectional ending (*okurigana*) るい *rui* in *hiragana*. 赤るい is a typical example of a phonological writing error, with the homophonous character 赤 *aka* "red" replacing the target character 明 (赤るい cannot be a written representation of the adjective "red," which is *akai* and is transcribed 赤い, without る in the middle).

This type of error implies (i) a phonologically inclined retrieval approach (i.e. choosing a homophonous character), (ii) disregard for the orthographic and semantic contradiction between the retrieved character and the target character (赤 "red" and 明 "bright" are both orthographically and semantically dissimilar), and (iii) disregard for the circumstantial discrepancy (as an adjective, 赤 does not take るい for *okurigana*). In order to make this type of error, one must at least be equipped with phonological and orthographic knowledge of the mistaken character. Since the retrieval process is based on the sound-grapheme correspondence and there is no detectable attempt at component-based analogy or reliance on circumstantial clues, the unit of processing is regarded as character-based.

A necessary condition for this type of error is that one must have learned how to read and write characters that are homophonous with the target character. Therefore, the number of learned homophonous characters can affect the likelihood of generating this type of error. For instance, L1 learners learn only one character pronounced *kō* (校) in the first grade, but as many as eight (工, 公, 広, 行, 交, 光, 後 and 高) in the second grade. Consequently, it is highly unlikely for first graders to make a phonological writing error of the character 校 *kō* because they have not learned any homophonous characters, whereas second graders have learned eight characters with which they can get confused. In the case of a less common reading such as *roku*, one character (六) is learned in the first grade but none in the second grade. Therefore, there is very little risk for both first and second graders of making a phonological writing error with the target character 六 *roku*.

5.3.2. Orthographic writing errors

An **orthographic** writing error is a substitution of the target character with an orthographically similar character. For example, the word *jibun* ("oneself") should be written as 自分, using the character 自 which reads *ji* ([ʑi]) and means "self." An erroneous transcription of this word 白分, with the target character 自 replaced by the orthographically similar character 白 ("white", with the multiple readings *shiro*, *shira*, *haku* and *byaku*), would be treated as an orthographic writing error.

The criterion for orthographic similarity used in this study was as follows: a character was classified as orthographically similar to the target character

if the number of strokes composing the identical component is shared by at least half the number of strokes of the whole of the target character. For example, the character 客 *kyaku* (“guest”) was classified as orthographically similar to the target character 額 *gaku* (“price”), because 客 is identical with the 9-stroke left radical of the 18-stroke target character. On the other hand, the 6-stroke character 各 *kaku* (“each”), which is identical with the bottom component of the left radical of the target character, was judged as dissimilar, because the number of strokes of the identical part constitutes only one third of the 18 strokes of the target character. Likewise, the character 複 *fuku* (“multiple”), with the 5-stroke left radical and the 9-stroke right radical, is orthographically similar to the target character 復 *fuku* (“repeat”) with 3-stroke left and 9-stroke right radicals, since the 9-stroke identical component constitutes 75% of the target character. However, the character 彼 *kare* (“he”), with 3-stroke left and 5-stroke right radicals, is dissimilar to 復, because the identical component 彳 comprises no more than 25% of the target character.

It is of course quite possible to make a slip of the pen and unintentionally produce another character that is orthographically similar to the target character, rather than writing another existing character under the misconception that it is the correct one. For example, one could absent-mindedly add an extra stroke to or omit a necessary stroke from the correct form, for instance by adding an unnecessary horizontal stroke to 休 (“rest”) and unintentionally composing 体 (“body”), or by carelessly missing one of the inner strokes of 自 (“self”) and inadvertently producing 白 (“white”) instead. Nevertheless, as long as the product is an existing character orthographically similar to the target character, it was counted as substitution with an orthographically similar character; it is presumed that the incorrect addition/omission did not register as an error in the learner’s mind because the product is an actual character which exists somewhere in the mental lexicon; it must have passed through the primary mesh of kanji-or-not-kanji distinction and must have “looked right.”

This type of error implies an orthographic approach to retrieval and a disregard for the phonological and semantic contradiction between the retrieved character and the target character, as well as heedlessness of the circumstantial discrepancy. In order to make this type of error, one must at least be equipped with an orthographic knowledge of the mistaken character (e.g.: 白 for 自), unless the wrong character was composed as a result of a slip-of-the-pen. It is therefore highly likely that the mistaken character can be found among the characters the learner had already learned. Since the retrieval process is based on the character as a whole and there is no detectable attempt at component-based analysis or reliance on circumstantial clues, the unit of processing is regarded as character-based.

5.3.3. Semantic writing errors

A **semantic** writing error is a substitution of the target character with a semantically related character. For example, when the reading *aki* is given in a context which requires the target character 秋 (meaning "autumn"), the learner may write the antonymous character 春 meaning "spring" (pronounced *haru*) instead. Another example is substituting 燃 *nen* ("burn") in the word 燃料 *nenryō* ("fuel") with the synonymous 焼 *shō* ("burn, roast, grill, toast, bake or fry").

This type of error indicates a semantic approach to retrieval. Since semantic writing errors, unlike their reading counterparts, require orthographic knowledge of the semantically related characters, the candidate characters for this type of error are, in principle, found among the characters that have already been learned.

5.3.4. Circumstantial writing errors

A **circumstantial** writing error is a substitution with a circumstantially associated character, i.e. a character that is substitutable based on **compound** constituents, **inflectional** endings or **contextual** congruence, as explained earlier with regard to circumstantial reading errors in subsection 5.2.4.

As is the case with circumstantial reading errors, a good knowledge of spoken Japanese (vocabulary, grammar and reading comprehension) is required in order to make this type of error. Although circumstantial reading errors suggest weakness in all three cognitive aspects – form, sound and meaning – of the knowledge of the target characters and dependence on circum-character knowledge, circumstantial writing errors at least serve as proof of orthographic knowledge of the circumstantially associated characters, and therefore, such characters are generally found among the assembly of learned characters. The retrieval approach is circumstantial and the kanji appears to be processed on a word-basis.

5.3.5. Pseudokanji writing errors

The aforementioned four types of writing errors involve substitution with another existing kanji character, and have corresponding reading error types. **Pseudokanji** writing errors, however, are deviations from existing kanji, for example, with additional or missing strokes, additional, missing, replaced or switched components, or a disproportionate assemblage of intra-character components, etc. By its nature, this error type has no reading counterpart.

The pseudokanji writing error type has been further divided into the following four subtypes:

(1) **Component**: This is a subtype involving component-based alteration of the target character, which results in the production of a pseudokanji character. Alteration can involve replacement of a component with another component, switching of the positions of the components, the addition of an extra component or omission of a constituent component, or an incomplete assemblage of components. For example, erroneous omission of the top component 土 of the right radical 寺 of the character 持 ("to have/hold") creates a pseudokanji character which is a combination of 扌+寸. Likewise, replacement of the left radical 扌 of the same character 持 with 犭 produces another pseudokanji character with the illegitimate combination 犭+寺.

However, if alterations to an existing component resulted in producing an existing kanji, it was classified under the orthographic type, because the product is an existing character orthographically similar to the target character. For instance, the left radical replacement of the target character 持 ("to have/hold") with 彳 would merely transform the target character into another existing character 待 ("to wait"), which should be classified as an orthographic writing error.

This component error subtype results in a pseudokanji consisting of existing components in an illegitimate combination. It indicates that a maker of this subtype of error has developed configurational awareness at least on a component basis, but not on a whole character basis.

(2) **Stroke**: This subtype includes errors that are made through the addition of an extra stroke or the omission of a constituent stroke of the target character, creating a non-existent component and thereby causing the character to deviate from any existing character. A piercing stroke with an incomplete intersection was categorised under stroke omission, and a non-piercing stroke making an illegitimate intersection under stroke addition. There are cases where an addition/omission of a stroke results in the production of another existing component. For instance, a stroke addition to the left side component 日 of the character 明 ("bright") produces another existing component 目, but the resulting illegitimate combination 目

+月 is a pseudokanji. Likewise, omitting the topmost stroke of the left component 禾 in the character 秋 (“autumn”) produces 木, but the resulting illegitimate 木+火 combination creates a pseudokanji character. Such cases were categorised under the abovementioned “Component” subtype, because although production of another existing component was a result of an accidental addition/omission of a stroke, it does result in creating another existing component (not a non-existing pseudocomponent), which indicates that at least the screening function of the mental lexicon to eliminate pseudo-components was in action.

Corresponding to component alteration, stroke addition/omission resulting in the production of an existing kanji was classified under the orthographic type: e.g. with a stroke deletion, the character 自 (“self”) produces an existing and orthographically similar character 白 (“white”).

This stroke error subtype results in a pseudokanji consisting of non-existing components. It indicates that a maker of this subtype of error has not developed configurational awareness on a component basis.

(3) **Whole**: A pseudokanji character may deviate from the target character on a whole character basis rather than on a component or stroke basis. This subtype of error deviates more from the target character than the component or stroke subtypes do, and therefore they can be regarded as a result of less developed configurational awareness than the cases of these two subtypes.

(4) **Mirror**: This is a pseudokanji character that is a mirror image of the target character. There is no alteration on a component or stroke basis; the form is identical with the target character, only reversed in structure. This subtype was set up to check if mirror-writing errors that are common in the initial stage of learning *kana* or alphabet (e.g. the *hiragana* pair さ *sa* and ち *chi* or the alphabet pair *b* and *d*, etc.) are equally common in kanji. In both alphabetic (English) and moraic (Japanese *kana*) L1WS background, children are known to go through a stage of mirror writing (at least of certain, often reversible, letters as exemplified above) during the early stages of literacy development, mostly prior to the age of seven (Schott, 2007; Coernell, 1985; Tanaka, 1978). Although mirror writing may occur under different conditions such as immaturity, ageing, learning disabilities and varying forms of cognitive impairment, mirror writing observed during the developmental stage of literacy is likely to be a result of limited visuokinaesthetic coordination (conflicts

between visual perception and motor acts) (Schott, 2007; Corballis, 1983).

The retrieval approach for pseudokanji errors is most likely to be orthographic. The product of this type of error is a non-existing pseudokanji, but it nonetheless has orthographic similarity to the target character, unless it belongs to the "whole" subtype. Unlike other writing error types that show orthographic knowledge of a character related to the target character in one way or another, pseudokanji writing errors do not display correct orthographic knowledge of any character. Therefore, novice learners who have not yet developed good configurational awareness frequently make this type of error.

5.4. Error type classification summary and examples

This section is divided into three subsections, each of which will contrast reading and writing error types: subsection 5.4.1 will present summarised descriptions, subsection 5.4.2 examples, and subsection 5.4.3 characteristics.

5.4.1. Summarised descriptions of kanji reading and writing error types

Each error type described in sections 5.2 and 5.3 is summarised in Table 20, contrasting reading and writing errors. Although the classification principles for reading and writing errors are to be as equal as possible for the purpose of making comparisons between the two skills, the descriptions of the phonological type ought to differ to cover the differing aspects of each skill. Furthermore, the pseudokanji type, which by its nature occurs only in writing, is not included in the classification of reading errors. The words in boldface within the described type constitute subtypes.

Table 20 Summarised descriptions of kanji reading and writing error types

Types	Reading error descriptions	Writing error descriptions
Phonological	Application of an inappropriate/wrong reading of a character due to (a) **transcription error**, (b) misapplied **alternative reading**, or (c) incorrect **component-based analog**y.	Substitution with a phonologically identical character.
Orthographic	Misreading of the character as	Substitution with
	an orthographically similar character.	
Semantic	Misreading of the character as	Substitution with
	a semantically related character.	
Circumstantial	Misreading of the character as	Substitution with
	a circumstantially associated character, i.e. a character that is applicable based on (a) **contextual** congruence, (b) **compound** constituents or (c) **inflectional** endings	
Pseudokanji	N/A (for writing errors only).	Substitution with a pseudokanji character that deviates from any existing character due to alterations based on a (a) **component** (b) **stroke** (c) **whole** character, or being a (d) **mirror** image of the target character.
Others	Errors not falling into any of the above categories.	

A mixed-type error was counted not as an occurrence of single mixed-type error but as an occurrence of multiple error types in the statistics, because the aim of this study is not minute categorisation of each erroneous answer but to find out the frequency of error type, which would indicate how strong the inclination towards a certain retrieval approach is. For example, when the word しんせつな (親切な *shinsetsu-na* "kind") was erroneously written as 新切な, it was counted as two types of writing error occurrences: one as the phonological type and the other as the orthographic type, since the correct character 親 *shin* ("parent, intimate") was replaced with the homophonous and orthographically similar character 新 *shin* ("new").

5.4.2. Error types and examples according to skill and level

Examples of errors in each category according to skill and level are presented in Table 21. The target character/reading within the question words (Q-words) are underlined.

Table 21 Error types and examples according to skill and level

Types	Level	Reading		Writing	
		Q-Words (*correct answer*)	Errors (**subtype**: description)	Q-Words (*correct answer*)	Errors (**subtype**: description)
Phonological	LV240	*市長 "mayor" (しちょう shi-chō)*	ちちょう *chi-chō* <non-word> (**transcription error**: *sh*/*ch*)*	*あかるい "bright" (明るい aka-rui)*	赤るい *aka-rui* <non-word> (homophonous)
	LV800	*境目 "border" (さかいめ sakai-me)*	きょうめ *kyō-me* <non-word> (**alternative reading**)	*ふじんふく "women's clothing" (婦人服 fu-jin-fuku)*	夫人服 *fu-jin-fuku* <non-word> (homophonous)
		燃料 "fuel" (ねんりょう nen-ryō)	ぜんりょう *zen-ryō* <non-word> (**component-based analogy**) 火 (ひ) ＋ 然 (ぜん)		
Orthographic	LV240	*西口 "west exit" (にしぐち nishi-guchi)*	よんぐち *yon-guchi* <non-word> (四口)	*じぶん "oneself" (自分 ji-bun)*	白分 *haku-bun* <non-word>
	LV800	*衛星 "satellite" (えいせい ei-sei)*	まちせい *machi-sei* <non-word> (街星)	*むらがる "to flock" (群がる mura-garu)*	郡がる *gun-garu* <non-word>
Semantic	LV240	*首 "neck" (くび kubi)*	のど** *nodo* (喉 "throat") (synonymous)	*あき "autumn" (秋)*	春 *haru* "spring" (antonymous)
	LV800	*綿 "cotton" (わた wata)*	きぬ *kinu* (絹 "silk") (related)	*ねんりょう "fuel" (燃料 nen-ryō)*	焼料 *shō-ryō* <non-word> (synonymous)

(continued)

Table 21 *Continued*

Circum-stantial	LV240	*心 "heart" (こころ kokoro)*	へや *heya* "room" (部屋) (**context**: ～が広い) "… is big"	*げん気 "good health" (元気 gen-ki)*	天気 *ten-ki* "weather" (**compound**: ～気)
	LV800	*永久 "eternity" (えいきゅう ei-kyū)*	えいえん *ei-en* "permanence" (永遠) (**compound**: 永～)	*なれる "get used" (慣れる na-reru)*	通れる *too-reru* "can go through" (**inflectional**: ～れる)
Pseudo-kanji		N/A (for writing errors only)		Deviations from existing character (with additional or missing strokes; additional, missing, replaced or switched components; disproportionate assemblage; etc.)	
Others	LV240	*強い "strong" (つよい tsuyo-i)*	あらしい *arashii* <non-word> (unclassifiable)	*よる "night" (夜 yoru)*	**前** *mae* "before" (unclassifiable)
	LV800	*独立 "independence" (どくりつ doku-ritsu)*	いりつ *i-ritsu* <non-word> (unclassifiable)	*じゅうきょ "residence" (住居 jū-kyo)*	住谷 *jū-koku* <non-word> (unclassifiable)

* Japanese speech sounds *sh* ([ɕ]) as in し and *ch* ([tɕ]) as in ち are difficult to distinguish for native Swedish speakers.

** The Swedish word *hals* covers both 首 (*kubi* "neck") and 喉 (*nodo* "throat"), which intensifies the synonymy of the two characters and makes them even more difficult to distinguish from each other.

5.4.3. Summarised characteristics of kanji reading and writing error types

The characteristics of each error type mentioned in sections 5.2 and 5.3 are sorted according to the following points: (i) the absence of knowledge that could lead to such an error; (ii) the approach that must have been taken in the failed attempt to retrieve the target reading/character from the mental lexicon (i.e. the cognitive aspect of kanji that appears to have been activated or the features of kanji on which the learner is presumed to have been dependent); (iii) the knowledge of or ability regarding kanji that is necessary to produce that particular type of error; (iv) the causes of error-screening deficiency, namely, the aspects of kanji of which the learner must have been uncertain, the discrepancy disregarded by the learner, or the

learner's underdeveloped ability; and (v) the unit of kanji processing (either as the problematic part or in the retrieval attempt).

Table 22 is a list of summarised characteristics of kanji reading and writing error types. The following symbols and abbreviations are used in the list:

Symbols:
−: Lacking knowledge
A: Approach to retrieval
D: Displayed knowledge/ability/awareness
? : Uncertain aspects, disregarded discrepancy or underdeveloped ability
U: Unit of processing

Table 22 Summarised characteristics of kanji reading and writing error types

Types		Reading error characteristics		Writing error characteristics
Phonological	−	Target character reading		Target character orthography
	A	Phonological		
	D	(a) Trans.	(Target character reading)	Phonologically identical character orthography
		(b) Alt.	Alternative character reading	
		(c) Comp.	Character decomposition Component pronunciation	
	?	(a) Trans.	Phonological awareness *Kana* orthography	Target character meaning; Phonologically identical character meaning
		(b) Alt.	Circumstantial constraints	
		(c) Comp.	Circumstantial constraints Character-radical *on*-reading consistency	
	U	(a) Trans.	Phoneme/mora	Whole character
		(b) Alt.	Whole character	
		(c) Comp.	Component	
Orthographic	−	Target character reading		Target character orthography
	A	Orthographic		
	D	Orthographically similar character reading		Orthographically similar character orthography
	?	Target character meaning; orthographically similar character meaning		
		Orthographically similar character orthography		Orthographically similar character reading
	U	Whole character		

(continued)

Table 22 *Continued*

Semantic	−	Target character reading	Target character orthography
	A	Semantic	
	D	Semantically related character reading	Semantically related character orthography
	?	Target character meaning; semantically related character meaning	
		Semantically related character orthography	Semantically related character reading
	U	Whole character	
Circum-stantial	−	Target character reading	Target character orthography
	A	Circumstantial	
	D	Circumstantially associated character reading; Vocabulary; Grammar; Context	Circumstantially associated character orthography; Vocabulary; Grammar; Context
	?	Target character meaning; circumstantially associated character meaning	
		Circumstantially associated character orthography	Circumstantially associated character reading
	U	Word	

Types	Reading error characteristics	Cs	Writing error characteristics	
Pseudo-kanji	N/A (for writing errors only)	−	Target character orthography	
		A	Orthographic	
		D	(1) Component	Componential orthography
			(2) Stroke	Stroke assemblage pattern
			(3) Whole	---
			(4) Mirror	Reversed character orthography
		?	(1) Component	Whole character orthography
			(2) Stroke	Componential orthography
			(3) Whole	Basic configurational awareness
			(4) Mirror	(Visuokinaesthetic coordination)
		U	(1) Component	Component
			(2) Stroke	Stroke
			(3) Whole	Whole character
			(4) Mirror	Whole character

6. RESULTS AND ANALYSES

The reading and writing tests for LV240 and LV800 (attached as Appendices 3 to 6) were marked and the errors were classified into different types according to the classifications described in Chapter 5, and then entered into a database for analysis. A configural frequency analysis was conducted for each result compared between the two groups.

In this chapter, the results will be presented and analysed according to skill level and compared between the Swedish and Japanese participants at the same level as follows: reading at LV240, reading at LV800, writing at LV240, and writing at LV800. Comparison between the levels according to skill (LV240 vs. LV800 in reading; and LV240 vs. LV800 in writing) will also be made. Finally, a comparison between the skills (reading vs. writing) will be made. The subsequent sections, therefore, are structured as follows:

6.1. LV240 reading results

6.2. LV800 reading results

6.3. Inter-level comparison of reading results

6.4. LV240 writing results

6.5. LV800 writing results

6.6. Inter-level comparison of writing results

6.7. Comparison between reading and writing results

From the marked test results, consisting of correct, erroneous and blank answers, erroneous answers were classified into error types. It should be noted that the statistics show the frequencies and rates of **error types**, not of errors (erroneous answers). As stated in subsection 5.4.1, an erroneous answer falling into multiple categories was counted as multiple error type occurrences. The error cases included in the various error types (phonological, circumstantial, orthographic and semantic types for reading and the corresponding four types plus the pseudokanji type for writing) were statistically processed for analysis. Errors falling under the category "Others" were excluded from the statistics, since the purpose of this study is to explore approaches to retrieval and the difficulties therein by examining error type occurrence tendencies, and unclassified errors cannot provide the necessary data. The results will be analysed with regard to general occurrence tendencies and significant differences, with breakdowns into subtypes where appropriate. In each section/subsection the list of the

frequency and rates of error type will be presented, followed by a graph of the rates.

6.1. LV240 reading results

In this section, LV240 reading results will be compared between the Swedish participants (SJ240) and the Japanese participants (NJ240). The overall results will be presented and analysed in subsection 6.1.1, followed by breakdowns of phonological errors, transcription errors and circumstantial errors in subsections 6.1.2, 6.1.3 and 6.1.4 respectively. Finally, subsection 6.1.5 will summarise the error occurrence patterns observed in subsections 6.1.1 to 6.1.4.

6.1.1. Overall LV240 reading results

The two groups share the general distribution pattern of two predominant types (phonological and circumstantial) jointly comprising more than 85% of all occurrences, in contrast with the orthographic and semantic types, which account for less than 15% of errors together. The notable difference is the order of the top two types. While the phonological type is predominant over the circumstantial type for SJ240, NJ240's results are in the reverse order.

Table 23 is a list of the frequencies and rates of overall reading error types of SJ240 and NJ240. The rates are graphed in Figure 8.

Table 23 Frequencies and rates of LV240 overall reading error types

	Phonological	Circumstantial	Orthographic	Semantic	Total
SJ240	180 (55.2%)	101 (31.0%)	21 (6.4%)	24 (7.4%)	326 (100%)
NJ240	198 (34.4%)	327 (56.9%)	23 (4.0%)	27 (4.7%)	575 (100%)

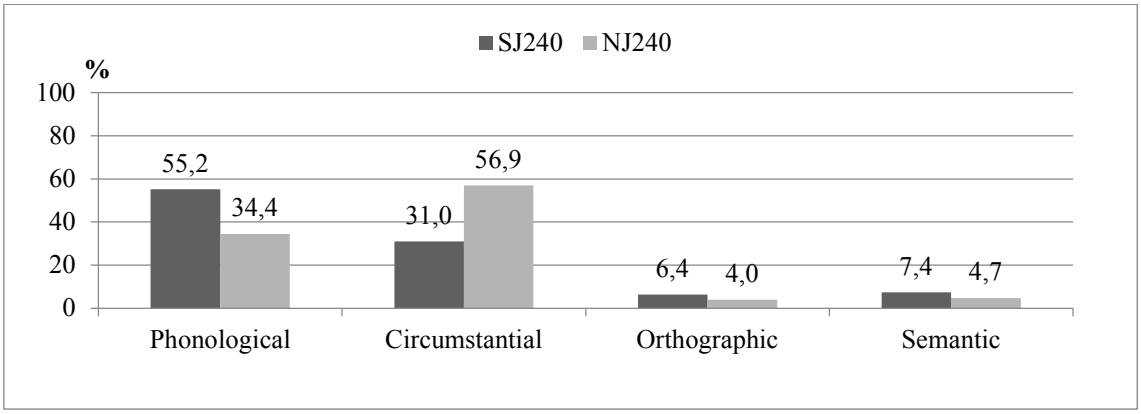

Figure 8 LV240 overall reading error type occurrence rates

The occurrence rate of each error type was compared between SJ240 and NJ240 through a configural frequency analysis. There was a significant difference ($df = 3$, $p < 0.0125$) between the two groups for two out of the four types, namely, the phonological type ($x^2 = 40.88$, $p = 6.9345E-09$) and the circumstantial type ($x^2 = 38.42$, $p = 2.3055E-08$). The differences for the orthographic type ($x^2 = 4.86$, $p = 0.1824$) and the semantic type ($x^2 = 4.93$, $p = 0.1765$) were at an insignificant level.

The large sum of the phonological and circumstantial types indicates the predominance of phonological and circumstantial approaches to retrieval in kanji reading. The tendency appears to be skill-specific rather than dependent on the characteristics of the L1 background, since it is shared by both groups. The differences in the distribution of the top two types (phonological and circumstantial) will be examined through subtype analyses in subsections 6.1.2 to 6.1.4.

6.1.2. Breakdown of LV240 phonological reading errors

As stated in subsection 5.2.1 and summarised in section 5.4, the phonological type is further divided into the three subtypes: (a) transcription errors, (b) alternative readings, and (c) component-based analogies.

The two groups of subjects share the general distribution pattern of absolute predominance of transcription errors and alternative readings, which jointly comprise practically 100% of all error occurrences, leaving component-based analogy at virtually nil. The notable difference is the order of the top two types. While the transcription subtype is predominant over the alternative reading type for SJ240, NJ240's results are in the reverse order.

The frequency and percentage of phonological errors falling into each subtype are listed in Table 24 and the percentage figures are graphed in Figure 9:

Table 24 Breakdown of LV240 phonological reading errors in frequency and rate

	Transcription errors	Alternative reading	Component-based analogy	Total
SJ240	83 (58.5%)	55 (38.7%)	4 (2.8%)	142 (100%)
NJ240	64 (41.0%)	92 (59.0%)	0 (0.0%)	156 (100%)

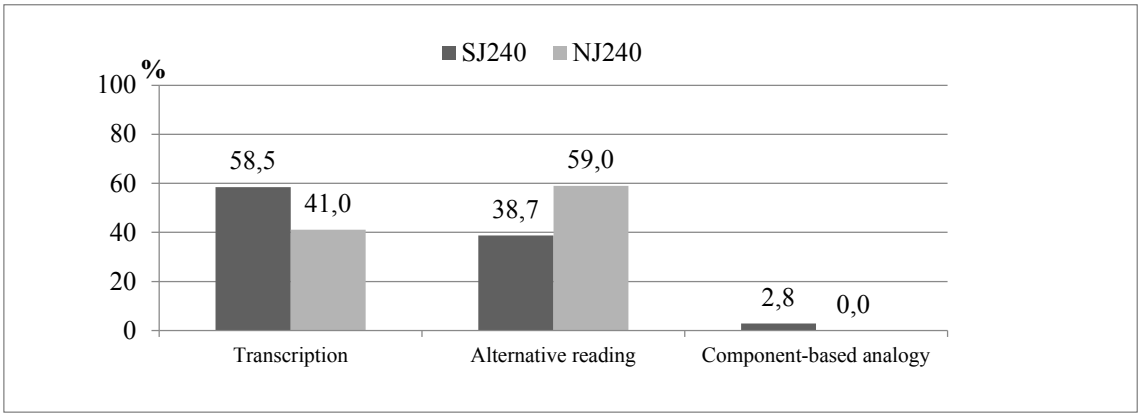

Figure 9 Breakdown of LV240 phonological reading errors by percentage

The occurrence rate of each subtype was compared between SJ240 and NJ240 through a configural frequency analysis. There was a significant difference ($df = 2$, $p < 0.0167$) between the two groups for all three subtypes, namely, the transcription errors ($x^2 = 10.51$, $p = 0.0052$), the alternative readings ($x^2 = 9.87$, $p = 0.0072$) and the component-based analogies ($x^2 = 0.00$, $p = 0.0000$).

Reasonable grounds for the virtual absence of component-based analogy at LV240 are that the learners' ability to decompose a character into components has not yet developed, and that they have not learned many phono-semantic composite pairs with character-radical *on*-reading consistency, and therefore do not have adequate phonological knowledge on which to base such analogies. This presumption is supported by implications from previous studies, such as the underdeveloped orthographic awareness observed among novice L2 learners with alphabetic L1WS (Matsumoto, 2013) and the late development of phonological radical awareness in L1 logographic children (Chan & Nunes, 1998).

The predominance of alternative readings over transcription errors for NJ240 can be explained with the following factors:

A. the order of kanji learning (vocabulary before character):

   a. As L1 speakers of Japanese, NJ240 participants have fairly large body of spoken words before learning kanji, and kanji are learned as the means of writing down words they already know.

   b. The sound and meaning (e.g. *ushi* = "cow") is already in the NJ240 participants' mental lexicon, and the grapheme (牛) is a

later addition. Therefore, the grapheme is likely to be loosely linked to the sound and meaning combination.

c. On the other hand, if kanji and vocabulary are learned more or less simultaneously, kanji graphemes are likely to be perceived as a part of the form-sound-meaning trinity of a word, conscious of the vocabulary-based circumstantial constraint.

d. Therefore, L1 learners may be less hesitant to apply a known reading of the target character regardless of the circumstantial constraints, such as misreading 牛肉 *gyū-niku* ("beef"), which is a combination of 牛 (*ushi* "cow") and 肉 (*niku* "meat"), as *ushi-niku*, misapplying the *kun*-reading *ushi* instead of the proper *on*-reading *gyū*.

B. There may be another reason for young L1 learners' difficulty in reading *on*-reading words, as indicated by BAASC (2007). When they cannot remember the target character's *on*-reading, which is often in a compound outside everyday vocabulary (e.g. 牛肉 *gyū-niku* "beef"), they choose to apply the *kun*-reading of the character, the only reading of the target character they can remember, because it tends to be a familiar single-character word (e.g. 牛 *ushi* "cow").

The higher percentage of transcription errors for SJ240 suggests that their mastery of *kana* orthography and their phonological awareness are poorer than those of NJ240. This aspect will be analysed more deeply in subsection 6.1.3 with a breakdown of the transcription errors.

6.1.3. Breakdown of LV240 transcription errors

Transcription errors are incorrect *hiragana* transcriptions of kanji readings involving orthographically less transparent syllables/morae and problematic phonemes. Examples of such syllables/morae/phonemes are as follows:

(1) palatalised syllables (拗音), transcribed in *hiragana* with additional reduced size: -や/-ゆ/-よ (*-ya/-yu/-yo* in *romaji*) (e.g. きゃ/きゅ/きょ *kya/kyu/kyo*, ひゃ/ひゅ/ひょ *hya/hyu/hyo*, etc., see Table 1 in section 2.1 for more examples);

So-called special morae (特殊拍), namely:

(2) moraic nasals (撥音), transcribed in *hiragana* with ん and with a word-final and pre-consonantal *n* (or *m* before *m*, *b* and *p*) in *romaji* (e.g. ほん *hon* "book," てんき *tenki* "weather," げんまい

gemmai “brown rice,” ぜんぶ *zembu* “all,” and しんぱい *shimpai* “worry”);

(3) geminate consonants (促音), transcribed in *hiragana* with -っ, representing all non-nasal geminate consonants (e.g. きっぷ *kippu* “ticket,” にっき *nikki* “diary,” etc.

(4) long vowels (長母音), transcribed with an additional vowel letter in *hiragana* and with *ā/ī/ū/ē/ō* (and in some variations with an additional vowel letter) in *romaji*, as /a:/ -あ, /i:/ -い, /u:/ -う, /e:/ -い / -え, /o:/ -う/ -お (e.g. おかあさん *okāsan/okaasan* “mother,” いい *ī/ii* “good,” くうき *kūki/kuuki*, “air,” せんせい *sensē/sensee/sensei* “teacher,” おねえさん *onēsan/oneesan* “elder sister,” おう *ō/ou/oo* “king,” and おおい *ōi/ooi* “many”)

Speech sound variants that are transcribed with the same *kana* with or without a diacritic, most of which are voiceless/voiced pairs (see Table 1 in section 2.1). The precise pronunciations are transcribed in the International Phonetic Alphabet (IPA) between the romanisation and the transcriptions in *hiragana*:

(5) *k* vs. *g*, as in:
- *ka, ki, ku, ke, ko* [ka, ki, kɯ, ke, ko] (か, き, く, け, こ)
- *ga, gi, gu, ge, go* [ga, gi, gɯ, ge, go] (が, ぎ, ぐ, げ, ご)

(6) *t* vs. *d*, as in:
- *ta, te, to* [ta, te, to] (た, て, と)
- *da, de, do* [da, de, do] (だ, で, ど)

(7) *h*/*f* vs. *b* vs. *p*, as in:
- *ha, hi, fu, he, ho* [ha, çi, ɸɯ, he, ho] (は, ひ, ふ, へ, ほ)
- *ba, bi, bu, be, bo* [ba, bi, bɯ, be, bo] (ば, び, ぶ, べ, ぼ)
- *pa, pi, pu, pe, po* [pa, pi, pɯ, pe, po]

(8) *s* vs. *z*, as in:
- *sa, su, se, so* [sa, sɯ, se, so] (さ, す, せ, そ)
- *za, zu, ze, zo* [za, zɯ, ze, zo] (ざ, ず, ぜ, ぞ)

The speech sound pairs that many Swedish learners have difficulty in distinguishing due to phonological L1 transfer. Again, the IPA transcriptions of the precise pronunciations are provided between the romanisation and the transcriptions in *hiragana*:

(9) *j* → *sh*, as in:
ja, ji, ju, je, jo [ʑa, ʑi, ʑɯ, ʑe, ʑo] (じゃ, じ, じゅ, じぇ, じょ)
sha, shi, shu, she, sho [ɕa, ɕi, ɕɯ, ɕe, ɕo] (しゃ, し, しゅ, しぇ, しょ)

(10) *ch* → *sh*, as in:
cha, chi, chu, che, cho [tɕa, tɕi, tɕɯ, tɕe, tɕo] (ちゃ, ち, ちゅ, ちぇ, ちょ)
sha, shi, shu, she, sho [ɕa, ɕi, ɕɯ, ɕe, ɕo] (しゃ, し, しゅ, しぇ, しょ)

Among the sound types described above, both L1 and L2 learners are known to have difficulty mastering the transcription rules for the sounds (1) to (8) due to opaqueness of the GSC (grapheme-sound correspondence) or need for extra attention to diacritics. In addition, many Swedish learners have particular difficulty in distinguishing and transcribing the sounds described in (4), (8), (9) and (10) due to phonological L1 transfer (see section 3.4).

Table 25 is a list of LV240 target characters whose readings involve the error-inducing sounds described above:

Table 25 LV240 target characters with transcription error inducing sounds

Phoneme type		Quantity, kanji and reading								
Long vowel	10	kanji	遠	牛	強	教	高	週	多	
		reading	tō	gyū	kyō	kyō	kō	shū	ō	
		kanji	長	東	曜					
		reading	chō	tō	yō					
Palatalised syllable	7	kanji	牛	強	教	社	週	茶	長	
		reading	gyū	kyō	kyō	sha	shū	cha	chō	
Geminate consonants	0	kanji								
		reading								
Moraic nasal	11	kanji	間	元	今	心	親	前	電	
		reading	kan	gen	kon	shin	shin	zen	den	
		kanji	半	分	万	門				
		reading	han	bun	man	mon				
With/ without diacritic *	*k*/*g* 22	kanji	画	外	楽	帰	牛		元	午
		reading	ga	gai	gaku	kae	gyū		gen	go
		kanji	後	語	国	国	黒	書	心	
		reading	go	go	kuni	koku	kuro	ka	kokoro	
		kanji	切	長	東	買	聞	米	方	
		reading	ki	naga	higashi	ka	ki	kome	gata	
	t/*d* 5	kanji	遠	作	体	鳥	電			
		reading	tō	tsuku	karada	tori	den			
	h/*f*/ *b*/*p* 5	kanji	首	走	太	分	話			
		reading	kubi	hashi	futo	bun	hanashi			
s/*z*, *sh*/*j***	4	kanji	自	時	前	風				
		reading	ji	ji	zen	kaze				
sh/*ch***	3	kanji	近	茶	長					
		reading	chika	cha	chō					

* The category "with/without diacritic" includes both naturally voiced (with diacritic) consonants as well as voiceless (without diacritic) consonants that can have voiced variants through sequential voicing (*rendaku*), such as the character 心 as in 心 ***kokoro*** ("heart") and 真心 ***magokoro*** ("sincerity").

** For the sounds *s* vs. *z*, *sh* vs. *j* and *sh* vs. *ch*, only the characters with the sounds *z*, *j*, *ch* are counted, as changes due to Swedish L1 influence occur from *z* to *s*, *j* to *sh*, and *ch* to *sh*, and normally not the other way around.

Table 26 is the list of a breakdown of LV240 transcription errors including frequency and percentage of characters involving relevant sounds. The L1

Swedish transfer inducing sounds (long vowel, *s*/*z, sh*/*j,* and *sh*/*ch*) are marked in bold:

Table 26 Breakdown of LV240 transcription error subtype

	Less transparent syllables/morae (-あ/-い/-う/-え/-お; -や/-ゆ/-よ; -つ & ん)				With/ without diacritic (*k*/*g*-, *t*/*d*- & *h*/*f*/*b*/*p*)	**Swedish L1 transfer consonants**		Total
	Long vowel	Palatal. syl.	Gem. con.	Mor. nasal		***s*/*z*, *sh*/*j***	***sh*/*ch***	
No. (%) of characters	**10 (14.9%)**	7 (10.4%)	0 (0%)	11 (16.4%)	32 (47.8%)	**4 (6.0%)**	**3 (4.5%)**	67 (100%)
	28 (41.7%)							
SJ240	**28 (33.7%)**	8 (9.6%)	3 (3.6%)	3 (3.6%)	24 (28.9%)	**10 (12.0%)**	**7 (8.4%)**	83 (100%)
						17 (20.5%)		
	42 (50.6%)				34 (41.0%)			
NJ240	**15 (23.4%)**	12 (18.7%)	1 (1.6%)	3 (4.7%)	27 (42.2%)	**6 (9.4%)**	**0 (0.0%)**	64 (100%)
						6 (9.4%)		
	31 (48.4%)				33 (51.6%)			

The percentage of each subtype is graphed in Figure 10:

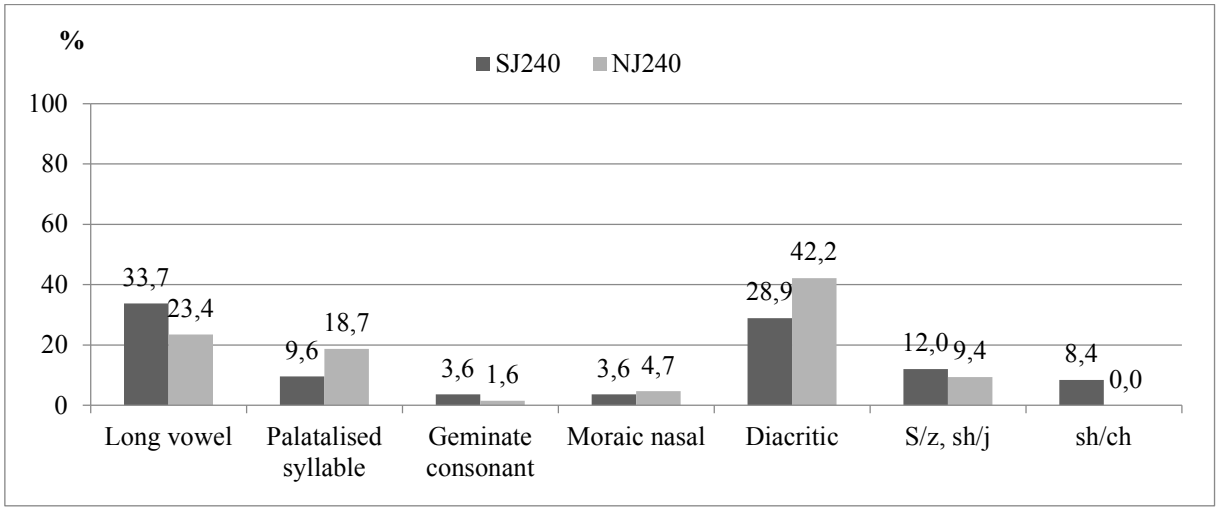

Figure 10 Breakdown of LV240 transcription error subtype

The occurrence rate of each subcategory was compared between SJ240 and NJ240 through a configural frequency analysis. There was a significant difference ($df = 6$, $p < 0.0071$) between the two groups for only one of the seven subcategories, namely, the sh/ch errors ($x^2 = 0.00$, $p = 0.0000$). The

differences for the other six subcategories were at an insignificant level: long vowel (x^2 = 3.76, p = 0.7098); palatalised syllable (x^2 = 3.67, p = 0.7206); geminate consonant (x^2 = 2.24, p = 0.8967); moraic nasal (x^2 = 0.20, p = 0.9998): diacritic (x^2 = 3.47, p = 0.7486); and *s*/*z* & *sh*/*j* (x^2 = 0.63, p = 0.9958).

The error subtype distribution patterns of both SJ240 and NJ240 roughly reflect the distribution pattern of sound formation of the characters included in transcription errors. As indicated in Table 26, the sound categories diacritic, long vowel and platalised syllable are among the highest for both SJ240 and NJ240 as well as in the sound formation. With regard to the Swedish L1 transfer inducing sound categories (long vowel, *s*/*z* & *sh*/*j* and *sh*/*ch*), SJ240 made significantly higher percentage of errors only with *sh*/*ch* than NJ800.

In order to investigate further the collective tendency of the Swedish L1 subjects to produce errors involving transfer inducing sounds, the seven sounds were divided into L1 transfer inducing and non-inducing groups. Table 27 is a list of a reorganised breakdown of LV240 transcription errors including frequency and percentage of characters involving collections of L1 Swedish transfer inducing sounds (long vowel, *s*/*z, sh*/*j, and sh*/*ch*) and non-inducing sounds (palatalised syllable, geminate consonant, moraic nasal and diacritic):

Table 27 Reorganised breakdown of LV240 transcription error subtype

	L1 transfer inducing sounds (long vowel, *s*/*z*, *sh*/*j*, *sh*/*ch*)	Non-L1 transfer inducing sounds (palatalised syllable, geminate consonant, moraic nasal, diacritic)	Total
SJ240	45 (54.2%)	38 (45.8%)	83 (100%)
NJ240	21 (32.8%)	43 (67.2%)	64 (100%)

The percentages in Table 27 are graphed in Figure 11:

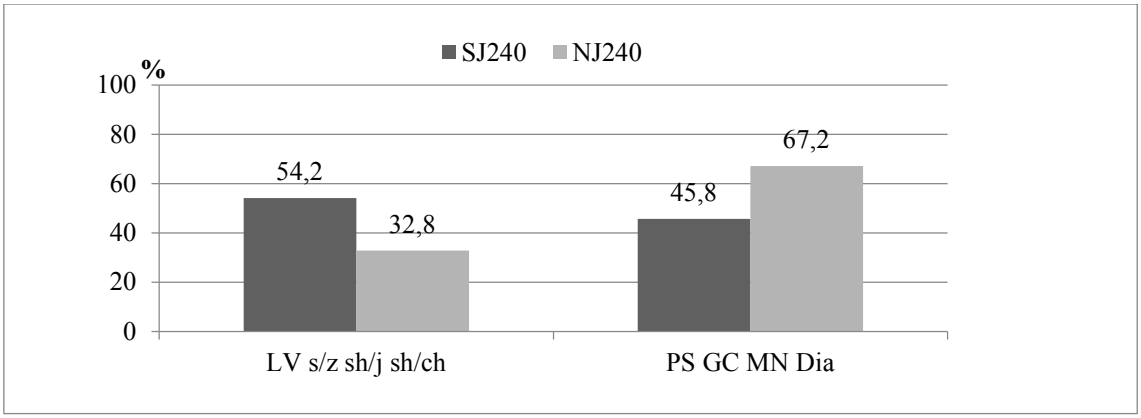

Figure 11 Reorganised breakdown of LV240 transcription error subtype

The occurrence rate of each subtype was compared between SJ240 and NJ240 through a configural frequency analysis. There was a significant difference ($df = 1$, $p < 0.025$) between the two groups for both L1 transfer inducing (LV, *s*/*z*, *sh*/*j*, *sh*/*ch*) and non-inducing sounds (PS, GC, MN, Dia): namely, $x^2 = 11.59$, $p = 0.0007$ for the inducing sounds and $x^2 = 5.66$, $p = 0.0174$ for the non-inducing.

Although the difference was invisible in a piecemeal comparison, the collective comparison indicated that SJ240 learners' transcription errors were influenced by L1 phonology.

6.1.4. Breakdown of LV240 circumstantial reading errors

Based on the criteria stated in subsection 5.2.4, the circumstantial reading error type is further divided into the three subtypes: (a) **context**, (b) **compound** and (c) **inflection** for further analysis.

The general tendency that is common for both groups is the predominance of the context and the compound over the inflection. The difference is the order of the top two subtypes: more compound than context for SJ240, while the order is reversed for NJ240.

The frequencies and rates of phonological errors falling into each subtype are listed in Table 28 and the percentage figures are graphed in Figure 12:

Table 28 Breakdown of LV240 circumstantial reading errors in frequency and percentage

	Context	Compound	Inflection	Total
SJ240	35 (34.7%)	47 (46.5%)	19 (18.8%)	101 (100%)
NJ240	160 (48.9%)	122 (37.3%)	45 (13.8%)	327 (100%)

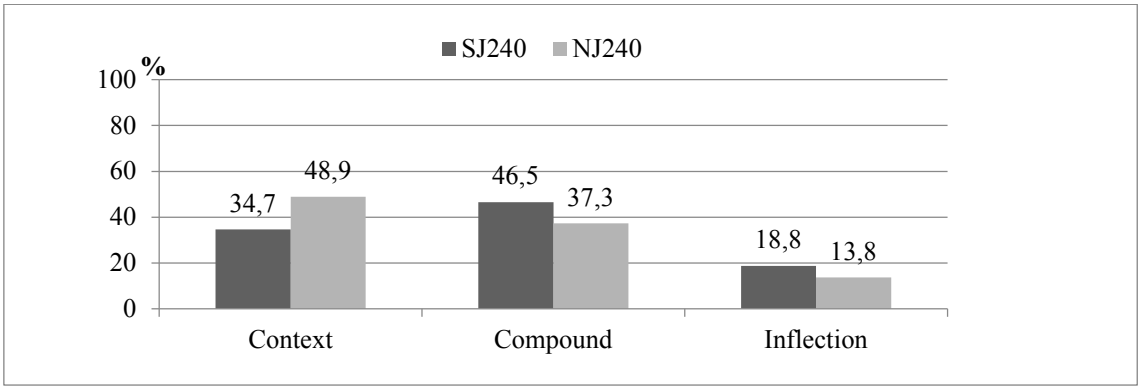

Figure 12 Breakdown of LV240 circumstantial reading errors by percentage

The occurrence rate of each subtype was compared between SJ240 and NJ240 through a configural frequency analysis. The result revealed no significant difference between the two groups for any of the subtypes (df = 2, $p < 0.017$; $p = 0.1220$ for context, 0.3160 for compound and 0.3922 for inflection).

The relatively small percentage of inflection errors is most likely due to the fact that only 27 out of the 113 target character-bearing words (23.9%) are accompanied by inflectional *okurigana*, i.e., no more than a quarter of the words provide *okurigana* on the basis of which which the participants can guess the reading of the character.

There are differences between the two groups' ability in spoken Japanese and the timing of their kanji learning: SJ240, as novice L2 learners, presumably have lower ability in spoken language, and learn vocabulary and the characters to write them with more or less simultaneously, whereas NJ240, being native speakers, had acquired the majority of words before they learned how to write them with kanji. As BAASC (2007) remarked, they tend to make fewer reading errors of *kun*-reading words, with which they are familiar in the spoken language, than less familiar *on*-reading words. It is therefore expected that SJ240 is less context dependent and more compound oriented than NJ240, due to SJ240's vocabulary-linked kanji knowledge and NJ240's dependence on their competence in spoken Japanese. Contrary to expectations, however, there are no significant differences regarding the frequencies of the subtypes context and compound.

6.1.5. Summary of LV240 reading results

This subsection summarises LV240 reading error occurrence patterns as analysed in subsections 6.1.1 to 6.1.4, and listed in Table 29:

Table 29 Summary of LV240 reading error occurrence patterns

Error occurrence patterns	Group	Similarities	Significant differences
6.1.1. Overall reading	SJ240	Phonological + Circumstantial > Orthographic + Semantic	Phonological > Circumstantial
	NJ240		Circumstantial > Phonological
6.1.2. Phonological breakdown	SJ240	Transcription + Alternative reading > Component-based analogy	Transcription > Alternative reading
	NJ240		Alternative reading > Transcription
6.1.3. Transcription breakdown	SJ240	Subtype distribution pattern mostly reflects distribution pattern of sound formation of characters in question	• *sh*/*ch* confusion • L1 transfer inducing sound > non- L1 transfer inducing sound
	NJ240		Non- L1 transfer inducing sound > L1 transfer inducing sound
6.1.4. Circumstantial breakdown	SJ240	Context + Compound predominance over Inflection	No significant difference
	NJ240		

6.2. LV800 reading results

In this section, LV800 reading results will be compared between the Swedish participants (SJ800) and the Japanese participants (NJ800). The overall results will be presented and analysed in subsection 6.2.1, followed by breakdowns of phonological errors, transcription errors and circumstantial errors in subsections 6.2.2, 6.2.3, and 6.2.4 respectively. Finally, subsection 6.2.5 will summarise the error occurrence patterns observed in subsections 6.2.1 to 6.2.4.

6.2.1. Overall LV800 reading results

The two groups share the general reading error type occurrence tendencies: the phonological and circumstantial types predominate over the orthographic and semantic types, the sum of the former two types being over 85%, and of the latter two under 15%. The difference is the order of the top two types. While the phonological type is predominant over the circumstantial type for SJ800, NJ800's results are in the reverse order.

Table 30 is a list of the frequencies and rates of overall error type occurrences of the reading results of SJ800 and NJ800. The occurrence rates are graphed in Figure 13.

Table 30 Frequencies and rates of LV800 reading error type occurrences

	Phonological	Circumstantial	Orthographic	Semantic	Total
SJ800	285 (47.7%)	237 (39.7%)	64 (10.7%)	11 (1.8%)	597 (100%)
NJ800	309 (29.4%)	682 (64.9%)	40 (3.8%)	20 (1.9%)	1058 (100%)

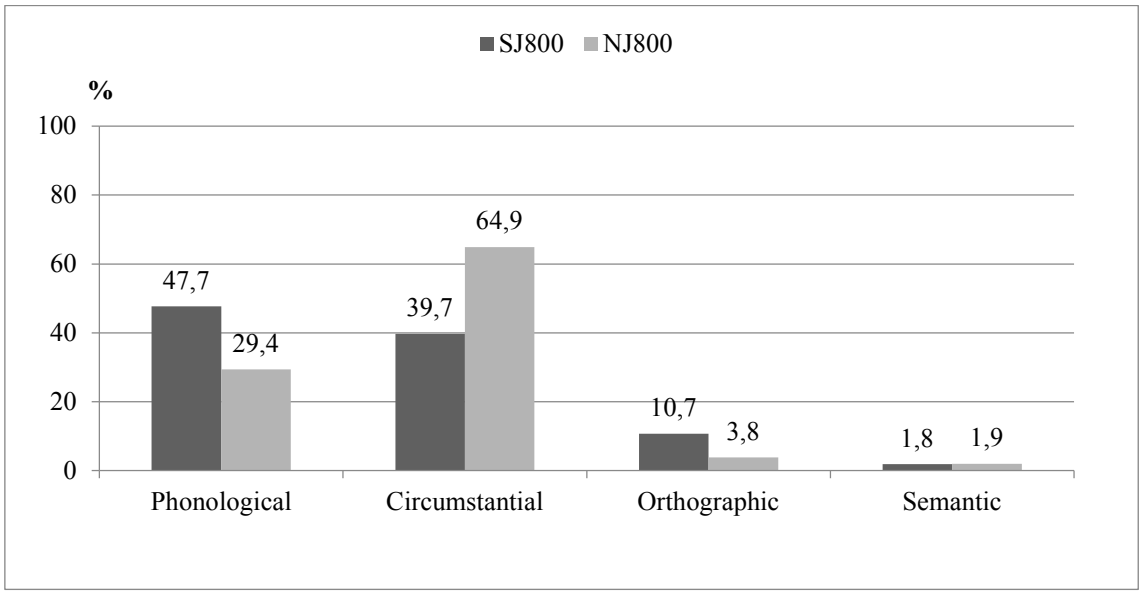

Figure 13 LV800 overall reading error type occurrence rates

The occurrence rate of each error type was compared between SJ800 and NJ800 through a configural frequency analysis. There was a significant difference ($df = 3$, $p < 0.0125$) between the two groups for three out of the four types, namely, the phonological type ($x^2 = 68.29$, $p = 9.9920\text{E-}15$), the circumstantial type ($x^2 = 58.39$, $p = 1.2990\text{E-}12$) and the orthographic type ($x^2 = 74.99$, $p = 0$). The difference for the semantic type ($x^2 = 0.01$, $p = 0.9997$) was at an insignificant level.

The phonological and circumstantial predominance which was observed in the corresponding LV240 result, was clearly the case for LV800 as well. This confirms that the tendency is skill-specific, regardless of proficiency. The differences in the distribution of the phonological and circumstantial types will be examined through subtype analyses in subsections 6.2.2 to 6.2.4.

6.2.2. Breakdown of LV800 phonological reading errors

As stated in subsection 5.2.1 and summarised in section 5.4, the phonological type is further divided into the three subtypes: (a) transcription errors, (b) alternative readings, and (c) component-based analogies.

The frequencies and rates of phonological errors falling into each subtype are listed in Table 31 and the percentage figures are graphed in Figure 14:

Table 31 Breakdown of LV800 phonological reading errors by frequency and rate

	Transcription errors	Alternative reading	Component-based analogy	Total
SJ800	56 (20.1%)	97 (34.9%)	125 (45.0%)	278 (100%)
NJ800	24 (8.6%)	223 (79.9%)	32 (11.5%)	279 (100%)

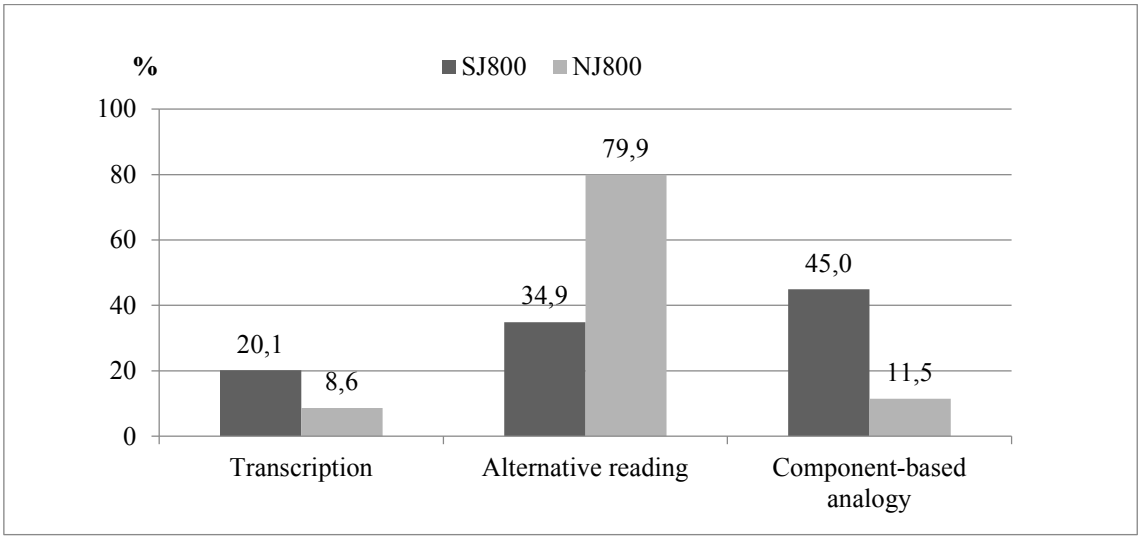

Figure 14 Breakdown of LV800 phonological reading errors by percentage

The occurrence rate of each subtype was compared between SJ800 and NJ800 through a configural frequency analysis. There was a significant difference (df = 2, p < 0.0167) between the two groups for all three subtypes: transcription errors (x^2 = 43.05, p = 4.484E-10), alternative readings (x^2 = 70.55, p = 0.0000), and component-based analogies (x^2 = 271.92, p = 0.0000).

The relatively low percentages of transcription errors (SJ800: 20.1%; NJ800: 8.6%) compared to the corresponding LV240 results (SJ240: 58.5%; NJ240: 41.0%) indicate an improvement in *hiragana* orthographic skills and phonological awareness at LV800. Nevertheless, SJ800's rate of

this type of error is as high as almost 20%, which suggests a deep-rooted L1 influence in this type of error.

The alternative reading was significantly more frequent in NJ800 than in SJ800 (SJ800: 34.9%; NJ800: 79.9%), which is in conformity with the corresponding results for LV240 (SJ240: 38.7%; NJ240: 59.0%). It is also supportive of the understanding stated in subsection 6.1.2 that this subtype is characteristic of L1 learners. They are apt to give priority to the form-sound connection of the character rather than extra-character conditions that determine the reading.

The NJ group's rate of alternative reading errors increases substantially between NJ240 (59.0%) and NJ800 (79.9%). This is most likely due to the difference in the size of the stock of multiple readings between the Japanese second and fifth graders. At the earlier stages of learning, a new character is often introduced with just one of its multiple readings (in many cases a *kun*-reading that constitutes basic vocabulary). For example, the character 古 ("old") is introduced at the second grade only with the reading *furu* as in the adjective "old" (古い *furu-i*), and the *on*-reading *ko* is not taught until later years, when it is introduced with less basic vocabulary such as 古代 *ko-dai* "ancient times" or 古典 *ko-ten* "the classics." Accordingly, higher-level learners have a larger stock of multiple readings that can be misapplied to a target character.

The most prominent difference from LV240 in this breakdown is the rise of component-based analogy errors, which can be interpreted on the basis that a higher number of learned characters is required to make this type of analogy. In fact, nearly half of all phonological reading errors by SJ800 are classified as belonging to this subtype. The high ratio of component-based analogy is supportive of the observation of Mori (1998) that pronounceable phonetic radicals have encouraging effects in kanji retrieval by L1 alphabetic learners.

6.2.3. Breakdown of LV800 transcription errors

In conformity with the corresponding analysis of breakdown of LV240 transcription errors in subsection 6.1.3, LV800 target characters whose reading involves the error inducing sounds are listed in Table 32:

Table 32 LV800 target characters with transcription error inducing sounds

Phoneme type		Quantity, kanji and reading											
Long vowel		33	kanji	衛	応	久	旧	境	経	効	耕	鉱	構	講
			reading	ē	ō	kyū	kyū	kyō	kē	kō	kō	kō	kō	kō
			kanji	雑	修	招	承	条	情	制	勢	精	製	税
			reading	zō	shū	shō	shō	jō	jō	sē	sē	sē	sē	zē
			kanji	総	造	程	導	能	評	報	豊	防	貿	領
			reading	sō	zō	tē	dō	nō	hyō	hō	hō	bō	bō	ryō
Palatalised syllable		11	kanji	久	旧	居	境	修	招	承	条	情	評	領
			reading	kyū	kyū	kyo	kyō	shū	shō	shō	jō	jō	hyō	ryō
Geminate consonants		1	kanji	設										
			reading	sett										
Moraic nasal		20	kanji	因	演	刊	均	禁	群	件	券	険	検	
			reading	in	en	kan	kin	kin	gun	ken	ken	ken	ken	
			kanji	限	混	賛	損	断	燃	犯	判	版	編	
			reading	gen	kon	san	son	dan	nen	han	ban	han	hen	
With/ without diacritic *	*k*/*g*	7	kanji	河	額	技	群	限	耕	貸				
			reading	ga	gaku	gi	gun	gen	tagaya	ka				
	t/*d*	8	kanji	妻	採	絶	断	導	独	務	留			
			reading	tsuma	to	ta	dan	dō	doku	tsuto	to			
	h/*f*/ *b*/*p*	16	kanji	破	犯	判	版	評	布	婦	富			
			reading	ha	han	ban	han	hyō	pu	fu	fu			
			kanji	復	複	編	報	豊	防	貿	暴			
			reading	fuku	fuku	hen	hō	hō	bō	bō	bō			
s/*z*, *sh*/*j***		12	kanji	財	罪	雑	雑	条						
			reading	zai	zai	zatsu	zō	jō						
			kanji	情	税	造	築	退						
			reading	jō	zē	zō	kizu	shirizo						
sh/*ch**		0	kanji											
			reading											

* The category "with/without diacritic" includes both naturally voiced (with diacritic) consonants as well as the voiceless (without diacritic) consonants that can have voiced variants by sequential voicing *rendaku* (連濁), such as the character 絶 as in 絶える ***ta***-*eru* ("cease") and 途絶える *to*-***da***-*eru* ("be interrupted").

** For the sounds *s* vs. *z*, *sh* vs. *j* and *sh* vs. *ch*, only the characters with the sounds *z*, *j*, *ch* are counted, as changes due to Swedish L1 influence occur only from *z* to *s*, *j* to *sh*, and *ch* to *sh*, and not normally the other way around.

Table 33 is the list of a breakdown of LV800's transcription errors including frequency and percentage of characters involving relevant sounds. The L1 Swedish transfer inducing sounds (long vowel, *s/z, sh/j,* and *sh/ch*) are marked in bold:

Table 33 Breakdown of LV800 transcription errors including frequency and rate of characters involving relevant sounds

	Less transparent syllables/morae (-あ/-い/-う/-え/-お; -や/-ゆ/-よ; っ & ん)				With/ without diacritic (*k/g, t/d, h/f/b/p*)	**Swedish L1 transfer consonants**		Total
	Long vowel	Palatal. syl.	Gem. con.	Mor. nasal		***s/z, sh/j***	***sh/ch***	
No. (%) of characters	33 (30.6%)	11 (10.2%)	1 (0.9%)	20 (18.5%)	31 (k/g:7, t/d:8, h/f/b/p:16 (28.7%)	12 (11.1%)	0 (0.0%)	108 (100%)
SJ800	12 (**21.4%**)	4 (7.1%)	3 (5.4%)	14 (25.0%)	6 (10.7%)	16 (28.6%)	1 (1.8%)	56 (100%)
						17 (**30.4%**)		
	33 (58.9%)				22 (39.3%)			
NJ800	8 (**33.3%**)	1 (4.2%)	1 (4.2%)	2 (8.3%)	7 (29.2%)	5 (20.8.%)	0 (0.0%)	24 (100%)
						5 (**20.8%**)		
	12 (50.0%)				12 (50%)			

The percentage of each subtype is graphed in Figure 15:

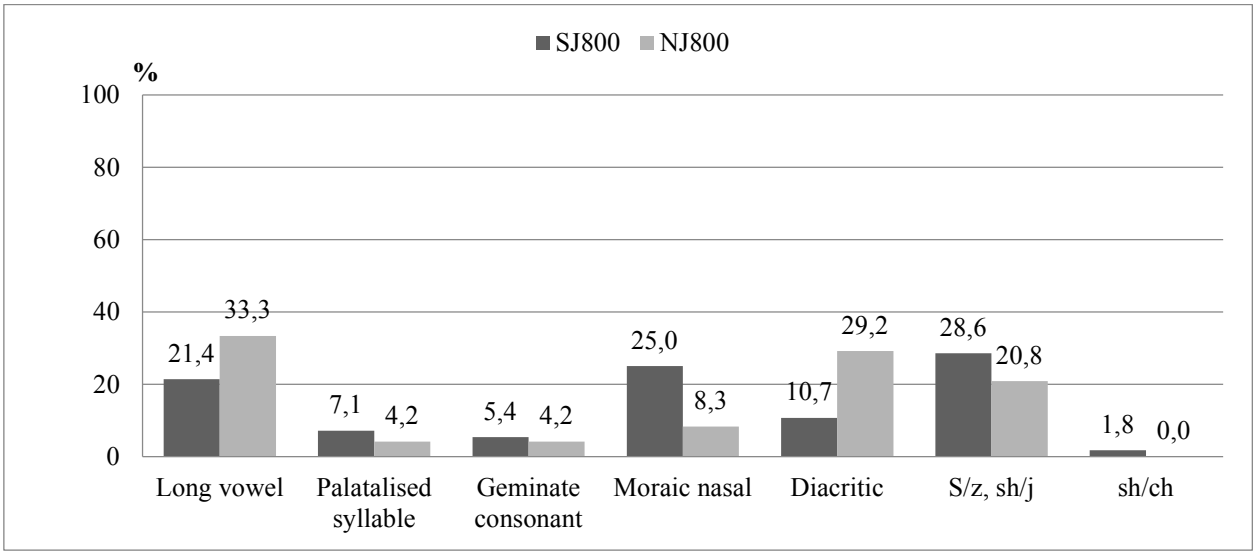

Figure 15 Breakdown of LV800 transcription reading error subtype

The occurrence rate of each subcategory was compared between SJ800 and NJ800 through a configural frequency analysis. There was a significant difference ($df = 6$, $p < 0.0071$) between the two groups for two of the seven subcategories, namely, the moraic nasal errors ($x^2 = 18.67$, $p = 0.0048$) and the *sh*/*ch* errors ($x^2 = 0.00$, $p = 0.0000$). The differences for the other six subcategories were at an insignificant level: long vowel ($x^2 = 2.38$, $p = 0.8815$); palatalised syllable ($x^2 = 1.19$, $p = 0.9774$); geminate consonant ($x^2 = 0.19$, $p = 0.9999$); diacritic ($x^2 = 6.54$, $p = 0.3657$); and *s*/*z*, *sh*/*j* ($x^2 = 1.61$, $p = 0.9519$).

The error subtype distribution patterns of both SJ800 and NJ800 mostly reflect the distribution pattern of sound formation of the characters included in transcription errors. As indicated in Table 33, the four highest-rate sound categories (long vowel, diacritic, moraic nasal and *s*/*z* and *sh*/*j*) in the sound formation are the four highest-rate error subtypes for both groups as well. Among the Swedish L1 transfer inducing sound categories (long vowel, *s*/*z* & *sh*/*j* and *sh*/*ch*), SJ800 made significantly higher percentage of errors with *sh*/*ch* than NJ800, but not with the other two.

In conformity with the corresponding analysis for LV240, the collective tendency to Swedish L1 transfer was investigated by reorganising the seven sounds into L1 transfer inducing and non-inducing groups. Table 34 is a list of a reorganised breakdown of LV800 transcription errors including frequency and percentage of characters involving collections of L1 Swedish transfer inducing sounds (long vowel, *s*/*z, sh*/*j, and sh*/*ch*) and non-inducing sounds (palatalised syllable, geminate consonant, moraic nasal, and diacritic):

Table 34 Reorganised breakdown of LV800 transcription error subtype

	L1 transfer inducing sounds (long vowel, *s*/*z*, *sh*/*j*, *sh*/*ch*)	Non-L1 transfer inducing sounds (palatalised syllable, geminate consonant, moraic nasal, diacritic)	Total
SJ800	29 (51.8%)	27 (48.2%)	56 (100%)
NJ800	13 (54.2%)	11 (45.8%)	24 (100%)

The percentages in Table 34 are graphed in Figure 16:

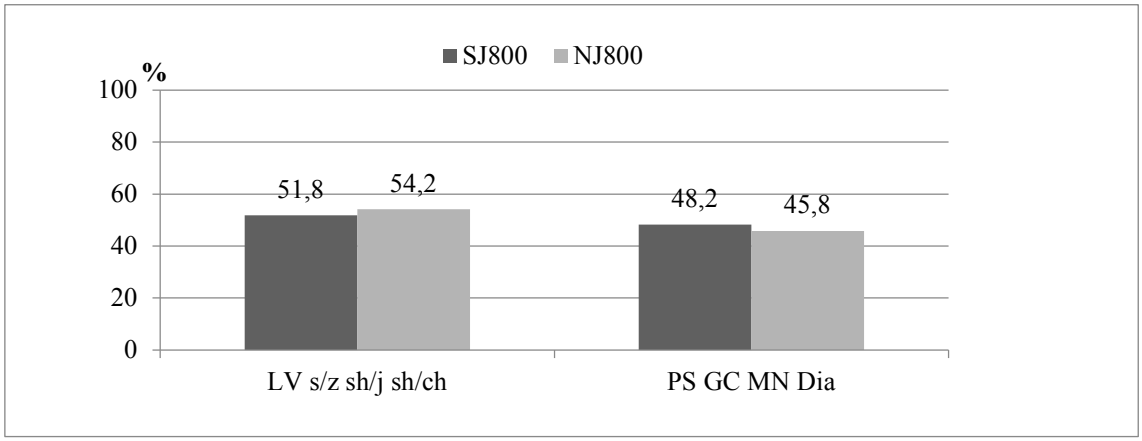

Figure 16 Reorganised breakdown of LV800 transcription reading error subtype

The occurrence rate of each subtype was compared between SJ800 and NJ800 through a configural frequency analysis. There was no significant difference ($df = 1$, $p < 0.025$) between the two groups for either L1 transfer inducing (LV, *s*/*z*, *sh*/*j*, *sh*/*ch*) ($x^2 = 0.06$, $p = 0.8087$) or non-inducing sounds (PS, GC, MN, Dia) ($x^2 = 0.07$, $p = 0.7924$).

The abovementioned absence of significant difference indicates that the apparent traits of L1 transfer observed at LV240 are fading at LV800.

6.2.4. Breakdown of LV800 circumstantial reading errors

Based on the criteria stated in subsection 5.2.4, the circumstantial reading error type is further divided into the three subtypes for further analysis: (a) **context**, (b) **compound**, and (c) **inflection**.

The frequency and rate of phonological errors falling into each subtype are listed in Table 35 and the percentage figures are graphed in Figure 17:

Table 35 Breakdown of LV800 circumstantial reading errors by frequency and rate

	Context	Compound	Inflection	Total
SJ800	68 (28.7%)	77 (32.5%)	92 (38.8%)	237 (100%)
NJ800	218 (32.0%)	243 (35.6%)	221 (32.4%)	682 (100%)

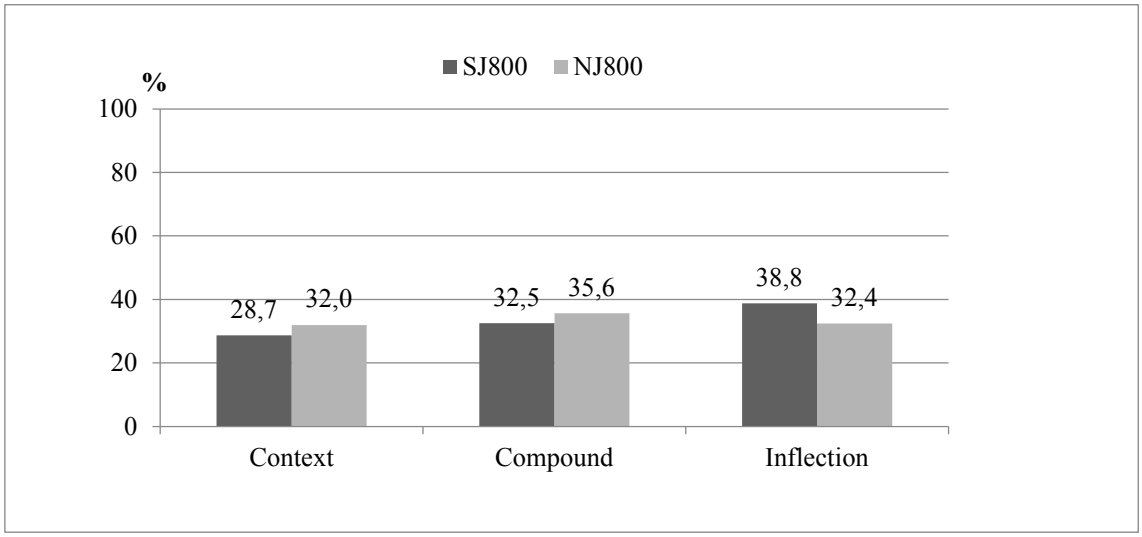

Figure 17 Breakdown of LV800 circumstantial reading errors by percentage

The occurrence rate of each subtype between SJ800 and NJ800 was compared through a configural frequency analysis. The result revealed no significant difference between the two groups for any of the subtypes (df = 2, $p < 0.017$; $p = 0.6722$ for context, 0.7203 for compound and 0.2222 for inflection).

As shown in subsection 6.2.1, NJ800 learners make circumstantial errors 1.6 times as often as SJ800 learners (64.9% vs. 39.7%). Interestingly though, the breakdown shows no significant difference in subtype rates between the two groups. Moreover, within each group, the three subtypes take nearly even shares. Although the target character accompanied by inflectional *okurigana* for LV800 is 25.2 % (36 out of 143), which is almost the same as LV240's 23.9 %, there is a substantial increase of this subtype for both SJ800 and NJ800. The prominent difference between the groups at LV240 is becoming indistinct at LV800, probably because the conditions for making circumstantial errors are being evened out. Being fifth graders, NJ800 learners are learning new advanced vocabulary and how to write it in kanji more simultaneously, rather than learning kanji to write the words they already know. On the other hand, SJ800 learners are more competent in spoken Japanese than SJ240 learners and they have become more dependent on spoken language in their kanji production.

6.2.5. Summary of LV800 reading results

This subsection summarises LV800 reading error occurrence pattern analyses in subsections 6.2.1 to 6.2.4, as listed in Table 36:

Table 36 Summary of LV800 reading error occurrence patterns

Error occurrence patterns	Group	Similarities	Significant differences
6.2.1. Overall reading	SJ800	Phonological + Circumstantial > Orthographic + Semantic	Phonological > Circumstantial
	NJ800		Circumstantial > Phonological
6.2.2. Phonological breakdown	SJ800	Transcription is least common	Component-based analogy > Alternative reading > Transcription
	NJ800		Alternative reading > Component-based analogy > Transcription
6.2.3. Transcription breakdown	SJ800	• Subtype distribution pattern mostly reflects distribution pattern of sound formation of characters in question • No significant difference between L1 transfer inducing and non-inducing sounds	• Moraic nasal • *sh*/*ch* confusion • No collective difference between L1 transfer inducing sound and non-L1 transfer inducing sound
	NJ800		
6.2.4. Circumstantial breakdown	SJ800	Even distribution pattern for context, compound and inflection subtypes	No significant difference
	NJ800		

6.3. Inter-level comparisons of reading results

This section will make inter-level comparisons of reading error occurrence patterns between the groups of the same L1 background (SJ240 vs. SJ800, and NJ240 vs. NJ800). Subsection 6.3.1 will compare overall reading error results, 6.3.2 will deal with the phonological error results, 6.3.3 the transcription error results, and 6.3.4 the circumstantial error results. Finally, subsection 6.3.5 will summarise the error occurrence patterns observed in subsections 6.3.1 to 6.3.4.

As mentioned in subsection 3.3.3, the use of different sets of target characters is likely to result in an uncontrolled balance of features (semantic/phonological transparency, complexity, familiarity, etc.) between the two sets of characters, which may influence error type occurrence patterns. Since LV240 and LV800 are tested for different sets of target characters, straightforward comparison is not quite appropriate. Therefore, LV240 and LV800 groups of the same L1 background will be compared

first, and then the inter-level shift in error making tendencies will be compared between SJ and NJ (SJ240-SJ800 shift vs. NJ240-NJ800 shift) to see how different (or similar) these shifts are.

6.3.1. Inter-level comparison of overall reading results

This subsection will make an inter-level comparison of overall reading error occurrence patterns. Figure 18 shows the comparison between SJ240 and SJ800 and Figure 19 between NJ240 and NJ800:

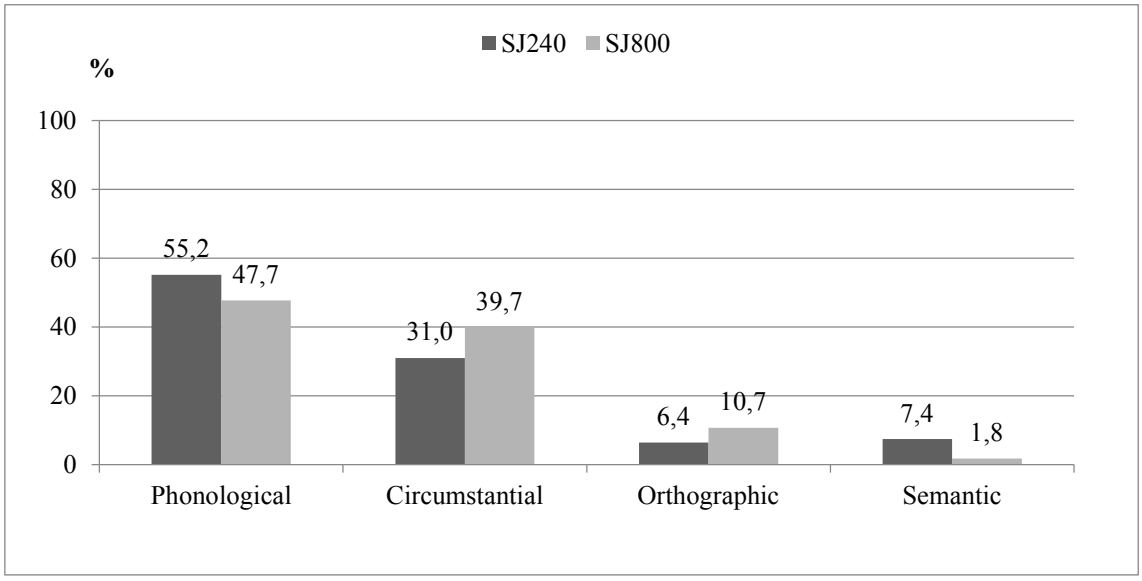

Figure 18 Inter-level comparisons of SJ groups' overall reading error type occurrence rates

The occurrence rate of each error type was compared between SJ240 and SJ800 through a configural frequency analysis. There was a significant difference ($df = 3$, $p < 0.0125$) between the two groups for only one out of the four types, namely, the semantic type ($x^2 = 53.90$, $p = 1.1788\text{E-}11$). The differences for the phonological type ($x^2 = 3.82$, $p = 0.2819$), the circumstantial type ($x^2 = 6.24$, $p = 0.1005$) and the orthographic type ($x^2 = 5.57$, $p = 0.1347$) were at an insignificant level.

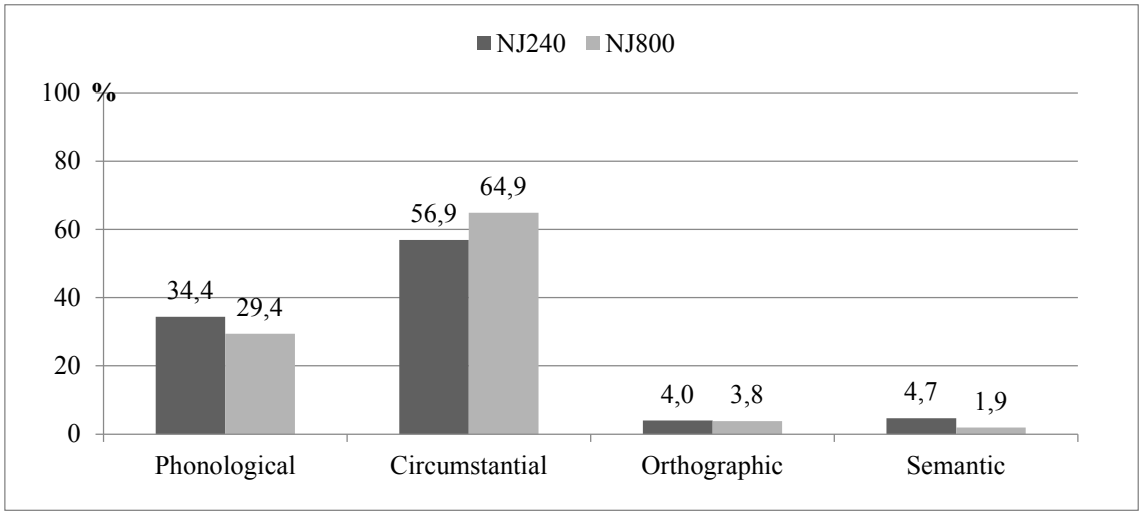

Figure 19 Inter-level comparisons of NJ groups' overall reading error type occurrence rates

The occurrence rate of each error type was compared between SJ240 and SJ800 through a configural frequency analysis. There was a significant difference ($df = 3$, $p < 0.0125$) between the two groups for only one out of the four types, namely the semantic type ($x^2 = 23.57$, $p = 3.0770\text{E-}05$). The differences for the phonological type ($x^2 = 4.96$, $p = 0.1750$), the circumstantial type ($x^2 = 5.70$, $p = 0.1005$) and the orthographic type ($x^2 = 0.06$, $p = 0.9964$) were at an insignificant level.

The semantic type was the only common error type that showed a significant inter-level difference for both SJ and NJ groups. Since the semantic type had a very low occurrence rate and was significantly higher at LV240 than at LV800 regardless of the L1 background, the grounds for this are likely to be either or both of the following:

(i) LV240 target characters include more characters belonging to the same semantic category, such as 春 "spring," 夏 "summer," 秋 "autumn" and 冬 "winter" or 東 "east, " 西 "west," 南 "south" and 北 "north," than LV800 target characters, which makes learners liable to produce semantic type errors ; and/or

(ii) Orthographic, phonological and semantic assemblies of such semantically related characters had not yet been well sorted within the mental lexicon of LV240 participants.

6.3.2. Inter-level comparison of phonological errors

This subsection will make an inter-level comparison of phonological error results. Figure 20 shows the comparison between SJ240 and SJ800 and Figure 21 between NJ240 and NJ800:

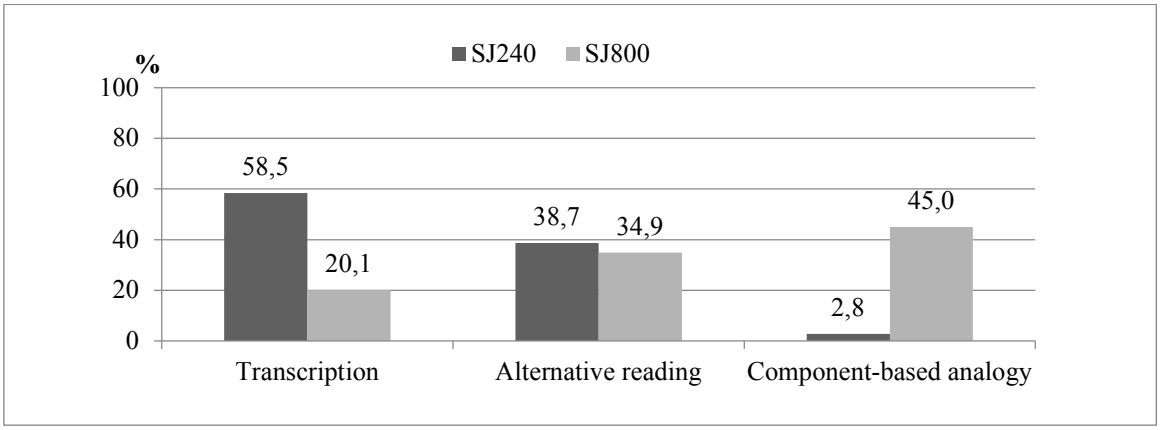

Figure 20 Inter-level comparison of SJ groups' phonological errors

The occurrence rate of each error type was compared between SJ240 and SJ800 through a configural frequency analysis. There was a significant difference ($df = 2$, $p < 0.0167$) between the two groups for two out of the three subtypes, namely, transcription errors ($x^2 = 103.44$, $p = 0.0000$) and component-based analogy errors ($x^2 = 56.10$, $p = 6.5792\text{E-}13$). The difference for the alternative reading subtype ($x^2 = 0.60$, $p = 0.7407$) was at an insignificant level.

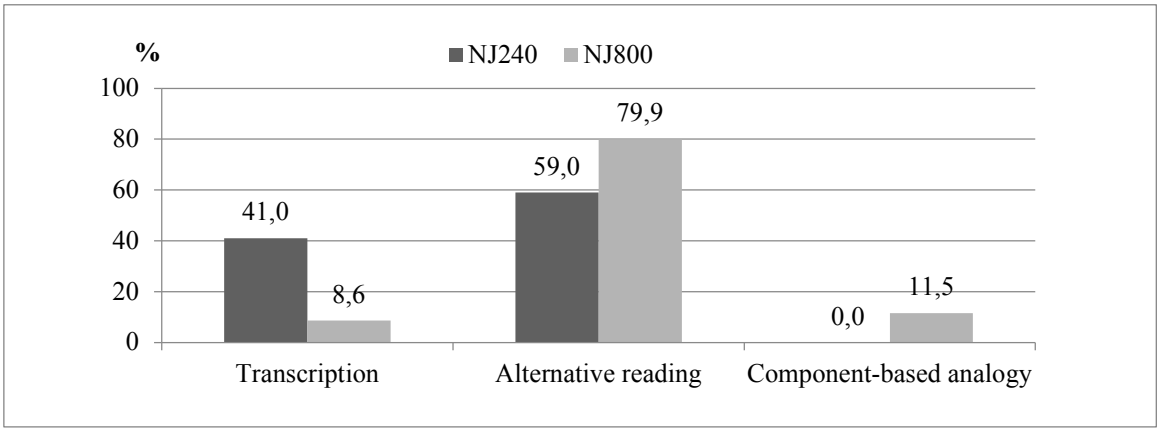

Figure 21 Inter-level comparison of NJ groups' phonological errors

The occurrence rate of each error type was compared between NJ240 and NJ800 through a configural frequency analysis. There was a significant difference ($df = 2$, $p < 0.0167$) between the two groups for all three subtypes, namely, transcription errors ($x^2 = 190.65$, $p = 0.0000$), alternative

reading errors ($x^2 = 8.57$, $p = 0.0138$) and component-based analogy errors ($x^2 = 17.89$, $p = 0.0001$).

In both SJ and NJ comparisons, LV800 groups' rates are significantly lower for transcription errors and higher for component-based analogy errors than LV240 groups. This should confirm the inferences that the transcription subtype is characteristic of LV240, and the component-based analogy of LV800. The alternative reading subtype, which is already high for NJ240, is significantly higher for NJ800. In comparing the SJ groups, however, there is no significant difference in alternative reading errors and the rate stays at approximately 35% for both SJ240 and SJ800.

The characteristics of inter-level shifting from LV240 to LV800 for phonological errors can be summarised as clear transition from the transcription subtype to the component-based analogy subtype for both SJ and NJ groups, and intensification of the predominance of the alternative reading subtype for NJ groups.

6.3.3. Inter-level comparison of transcription errors

This subsection will make an inter-level comparison between the L1 transfer inducing/non-inducing sounds within the transcription error results. Figure 22 shows the comparison between SJ240 and SJ800 and Figure 23 between NJ240 and NJ800:

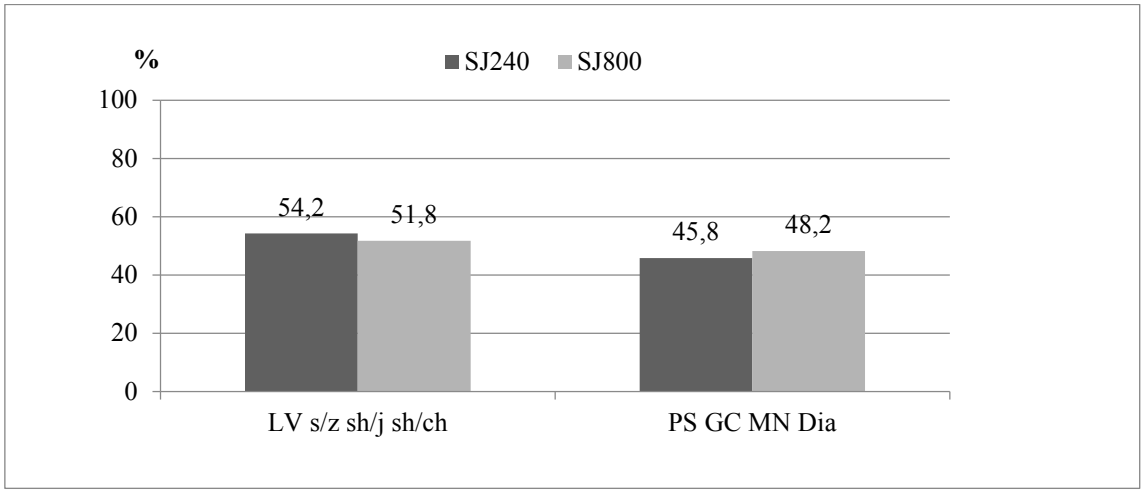

Figure 22 Inter-level comparison of SJ groups' transcription errors

The occurrence rate of each error type was compared between SJ240 and SJ800 through a configural frequency analysis ($df = 1$, $p < 0.025$). The differences for both L1 transfer inducing sounds (long vowels, s/z, sh/j and sh/ch) ($x^2 = 0.09$, $p = 0.7582$) and non-transfer inducing sounds (palatalised

syllables, geminate consonants, moraic nasals and diacritics) ($x^2 = 0.10$, $p = 0.7497$) were at an insignificant level.

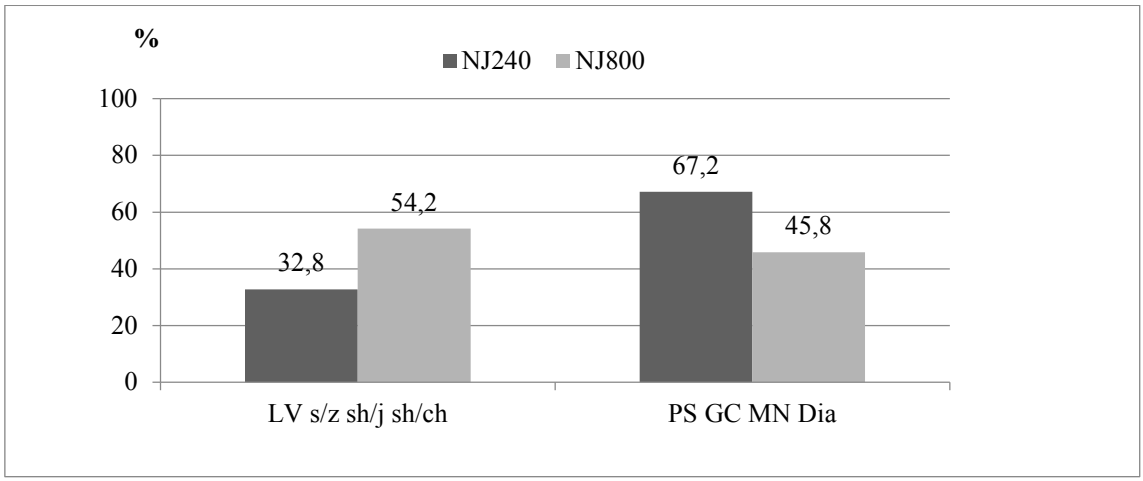

Figure 23 Inter-level comparison of NJ groups' transcription errors

The occurrence rate of each error type between NJ240 and NJ800 was compared through a configural frequency analysis. There was a significant difference ($df = 1$, $p < 0.025$) between the two groups for both categories: L1 transfer inducing sounds ($x^2 = 5.39$, $p = 0.0203$) and non-transfer inducing sounds ($x^2 = 6.37$, $p = 0.0116$).

Although SJ240 learners, who are at a beginner's level in learning Japanese, are expected to be under a stronger phonological influence from their L1 than SJ800 learners, the two groups' differences were at an insignificant level. When it comes to a comparison between the NJ groups, NJ240's percentage of Swedish L1 transfer inducing sounds is significantly lower than NJ800'sThis seemingly illogical result can be explained in a different light.

The sum of the numbers and rates of L1 transfer inducing/non-inducing sounds for both levels (as listed in Table 26 in subsection 6.1.3, and Table 33 in subsection 6.2.3) are listed in Table 37, together with corresponding error rates for each group:

Table 37 Rates of L1 transfer inducing/non-inducing readings and errors

	L1 transfer inducing sounds	L1 transfer non-inducing sounds
Total rates (numbers) of sounds within LV240 target readings	25.4% (17 out of 67)	74.6% (50 out of 67)
SJ240 error rates	54.2%	45.8%
NJ240 error rates	32.8%	67.2%
Total rates (numbers) of sounds within LV800 target readings	41.7% (45 out of 108)	58.3% (63 out of 108)
SJ800 error rates	51.8%	48.2%
NJ800 error rates	54.2%	45.8%

In comparison with the total rates of the sound categories, the error rates for each group are quite close to the corresponding rates of the sound categories, except for those of SJ240. The error rates for SJ240 deviate significantly from the corresponding rates within LV240 target readings, with a strong inclination toward L1 transfer inducing sounds (54.2% error rate against 25.4%). This seems to indicate that SJ240, far more than the other groups, is under the phonological influence of the L1.

6.3.4. Inter-level comparison of circumstantial errors

This subsection will make an inter-level comparison of circumstantial error results. Figure 24 shows the comparison between SJ240 and SJ800 and Figure 25 between NJ240 and NJ800:

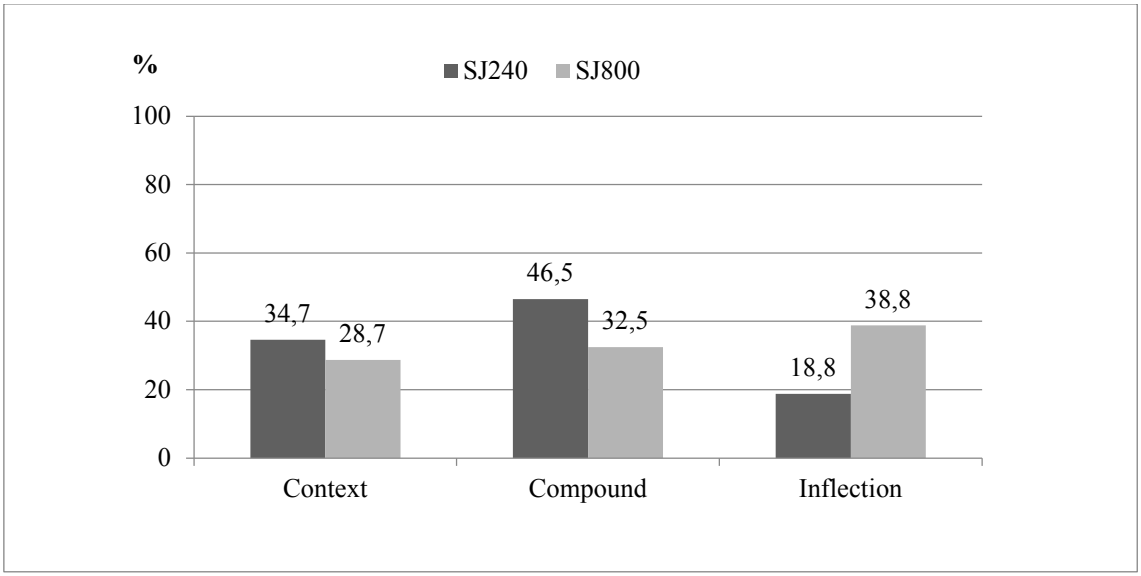

Figure 24 Inter-level comparison of the SJ groups' circumstantial errors

The occurrence rate of each error subtype was compared between SJ240 and SJ800 through a configural frequency analysis. There was a significant difference ($df = 2$, $p < 0.0167$) between the two groups for only one out of the three subtypes, namely the inflection subtype ($x^2 = 10.41$, $p = 0.0055$). The differences for the context subtype ($x^2 = 1.25$, $p = 0.5350$) and the compound subtype ($x^2 = 6.13$, $p = 0.0466$) were at an insignificant level.

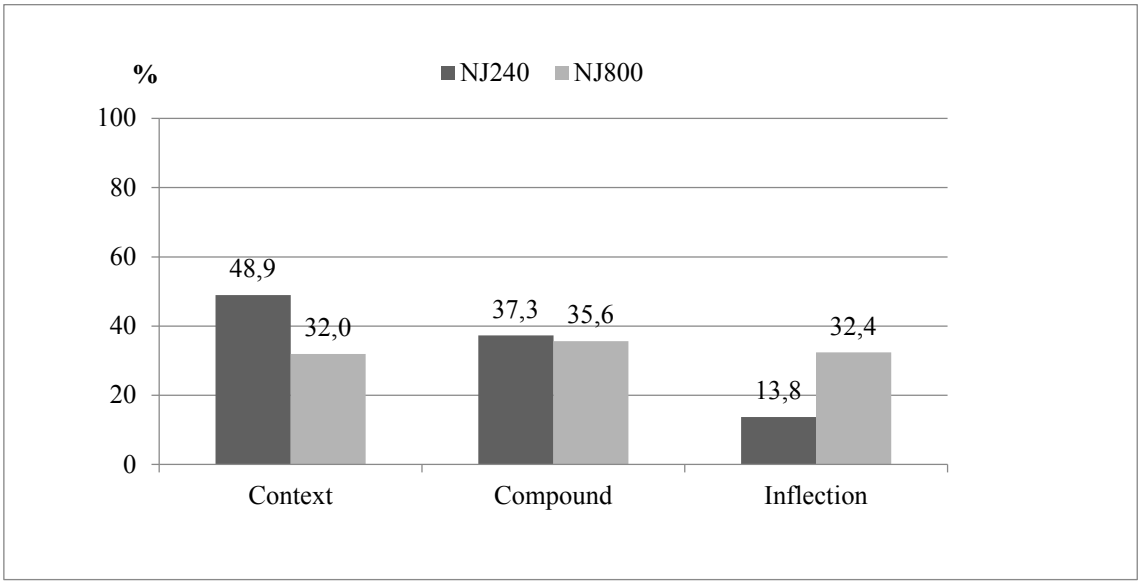

Figure 25 Inter-level comparison of the NJ groups' circumstantial errors

The occurrence rate of each error subtype was compared between NJ240 and NJ800 through a configural frequency analysis. There was a significant difference ($df = 2$, $p < 0.0167$) between the two groups for two out of the three subtypes, namely the context subtype ($x^2 = 29.44$, $p = 4.0423\text{E-}07$) and the inflection subtype ($x^2 = 35.08$, $p = 2.4201\text{E-}08$). The difference for the compound subtype ($x^2 = 0.26$, $p = 0.8787$) was at an insignificant level.

Although SJ240 showed an inclination towards the compound subtype and NJ240 towards the context subtype, SJ800 and NJ800 shared a more or less even distribution pattern of the three error subtypes. The inflection subtype, which was low for the LV240 groups, was significantly higher for the LV800 groups. The inflection subtype of error can occur only with the readings with *okurigana* or inflectional endings such as 強い *tsuyo-i* ("strong"), 会う *a-u* ("meet") and 群れ *mu-re* ("flock, herd") and the percentage of such readings is approximately 25% of all target readings for each level, as listed in Table 38 (see Appendices 3 and 4 for target readings). Therefore, the higher rate for LV800 was not influenced by the percentage of candidate readings but was a genuine increase in inclination towards the inflection subtype.

Table 38 Rates of target readings with *okurigana* and inflection subtype

Total rates (numbers) of readings with *okurigana* within LV240 target readings	25.7% (29 out of 113)
SJ240 rate of inflection subtype	18.8%
NJ240 rate of inflection subtype	13.8%
Total rates (numbers) of readings with *okurigana* within LV800 target readings	25.9% (37 out of 143)
SJ800 rate of inflection subtype	38.8%
NJ800 rate of inflection subtype	32.4%

6.3.5. Summary of inter-level comparison of reading error results

This subsection summarises inter-level comparison of reading error results in subsections 6.3.1 to 6.3.4, as listed in Table 39:

Table 39 Summary of inter-level comparison of reading error results

Error occurrence patterns	Group	SJ/NJ Similarities	SJ/NJ differences
6.3.1. Overall reading	SJ240 SJ800 NJ240 NJ800	LV240 Semantic > LV800 Semantic (influenced by the ratio of candidate readings)	SJ/NJ difference > LV240/LV800 difference
6.3.2. Phonological breakdown	SJ240 SJ800 NJ240 NJ800	• LV240 Transcription > LV800 Transcription • LV240 Component-based analogy < • LV800 Component-based analogy	Predominance shifting from transcription to component-based analogy Consistent predominance of alternative reading
6.3.3. Transcription breakdown	SJ240 SJ800 NJ240 NJ800	• No particular similarities	• Only SJ240 shows disproportionately high error rate for L1 transfer inducing sounds (under L1 phonological influence) • Greater SJ/NJ difference at LV240
6.3.4. Circumstantial breakdown	SJ240 SJ800 NJ240 NJ800	L800's inclination towards inflection subtype	Greater SJ/NJ difference at LV240

As indicated above, inter-level comparison of two L1-matched groups illuminate characteristics and tendencies according to L1 background.

6.4. LV240 writing results

In this section, LV240 writing results will be compared between the Swedish participants (SJ240) and the Japanese participants (NJ240). The overall results will be presented and analysed in subsection 6.4.1, followed by breakdowns of pseudokanji errors in subsection 6.4.2. Finally, subsection 6.4.3 will summarise the error occurrence patterns observed in subsections 6.4.1 and 6.4.2.

6.4.1. Overall LV240 writing results

The two groups showed the same general occurrence pattern of writing error types: the pseudokanji type was in the overwhelming majority, leaving the second and third most common orthographic and phonological types far behind, while the circumstantial and semantic types occupied insignificantly small percentages. The high rate of pseudokanji type errors indicates that handwriting kanji is difficult for all learners at this early stage of learning, presumably due to their underdeveloped configurational sense of kanji components.

Among the error types consisting of substitution with another existing character (orthographic, phonological, circumstantial and semantic), the orthographic type is most common. This suggests an orthographic approach to retrieval in this task and indicates that there is no close link between the task of character writing and the extra-character (circumstantial) conditions; it also suggests little activation of the phonological and semantic assemblies within the mental lexicon. The predominance of the orthographic type over the phonological type can be ascribed to the limited stock of kanji homophones within the mental lexicon and to an orthographic approach to kanji retrieval at this stage.

Table 40 is a list of the frequencies and rates of overall error type occurrences of LV240 writing results, comparing SJ240 and NJ240. The occurrence rates are graphed in Figure 26.

Table 40 Frequencies and rates of LV240 overall writing error type occurrences

	Phonological	Circum-stantial	Ortho-graphic	Semantic	Pseudokanji	Total
SJ240	38 (4.4%)	24 (2.7%)	77 (8.8%)	24 (2.7%)	710 (81.3%)	873 (100.0%)
NJ240	116 (7.2%)	69 (4.3%)	260 (16.2%)	53 (3.3%)	1106 (69.0%)	1604 (100.0%)

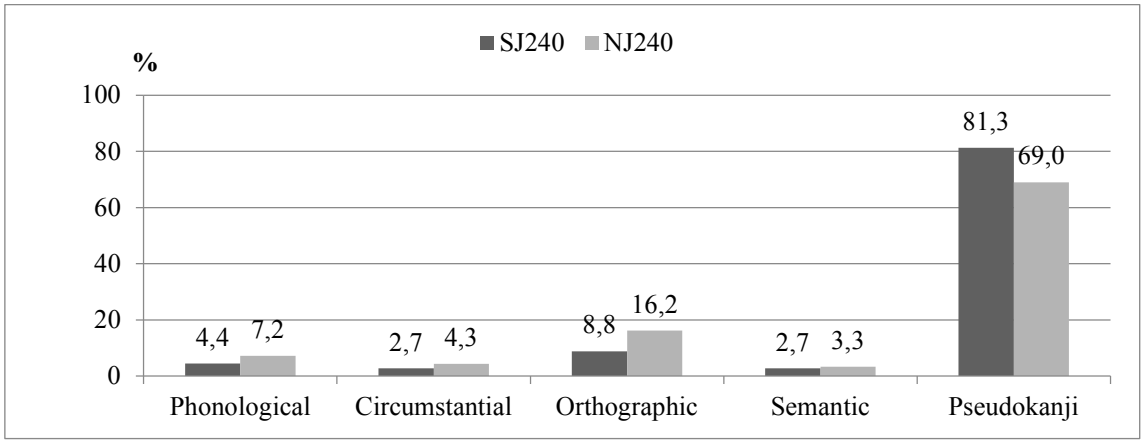

Figure 26 LV240 overall writing error type occurrence rates

The occurrence rate of each error type was compared between SJ240 and NJ240 through a configural frequency analysis. There was a significant difference ($df = 4$, $p < 0.01$) between the two groups for two out of the five types, namely, the orthographic type ($x^2 = 29.41$, $p = 6.4611E\text{-}06$) and the pseudokanji type ($x^2 = 19.39$, $p = 0.0007$). The differences for the phonological type ($x^2 = 10.01$, $p = 0.0403$), the circumstantial type ($x^2 = 4.89$, $p = 0.2986$) and the semantic type ($x^2 = 0.81$, $p = 0.9365$) were at an insignificant level.

SJ240 made a significantly higher percentage of pseudokanji type errors and a significantly lower percentage of the orthographic type than NJ240. This indicates that SJ240 has a less developed sense of the basic configuration of kanji than NJ240 and that they are less orthographically inclined in their retrieval of characters than NJ240.

The pseudokanji type will be further examined through subtype analysis in subsection 6.4.2 in order to explore the degrees of development of the two groups' awareness of kanji configuration.

6.4.2. Breakdown of LV240 pseudokanji writing errors

As stated in subsection 5.3.5, pseudokanji errors have been further divided into the following four subtypes, according to the manners in which the pseudokanji character deviates from the target character:

(1) **Component**: pseudokanji character consisting only of existing kanji components, made by altering a component of the target character;

(2) **Stroke**: pseudokanji character involving a non-existing component, produced by the addition of an extra stroke or omission of a constituent stroke of the target character;

(3) **Whole**: pseudokanji character that deviates from the target character on a whole character-basis rather than a component- or stroke-basis; and

(4) **Mirror**: a mirror image of the target character.

The two groups showed the same general occurrence pattern of pseudokanji error types: the component subtype was in the great majority, the second and third most common were the stroke and whole subtypes, accounting for a much smaller percentage, and the mirror subtype was practically non-existent. In terms of the developmental degree of configurational awareness implied by the occurrence of each of the top three subtypes, the component subtype is more developed, the stroke subtype less developed and the whole subtype least developed, as mentioned in subsection 5.3.5. Although the high ratio of the pseudokanji type does imply an underdeveloped configurational awareness for both LV240 groups, the high percentage of the component subtype indicates that the learners have at least developed configurational awareness of kanji components. A feasible explanation for the nearly non-existent mirror subtype is that the learners are no longer at the primary stage to make such a basic error. In addition, kanji is more orthographically complicated than *hiragana/katakana* or the Roman alphabet, which works as an obstruction to making the complete structural reversal necessary to produce a mirror image.

The frequencies and rates of pseudokanji errors falling into each subtype are listed in Table 41 and the percentage figures are graphed in Figure 27:

Table 41 Breakdown of LV240 pseudokanji writing errors in frequency and rate

	Component	Stroke	Whole	Mirror	Total
SJ240	387 (54.5%)	162 (22.8%)	159 (22.4%)	2 (0.3%)	710 (100%)
NJ240	788 (71.2%)	241 (21.8%)	75 (6.8%)	2 (0.2%)	1106 (100%)

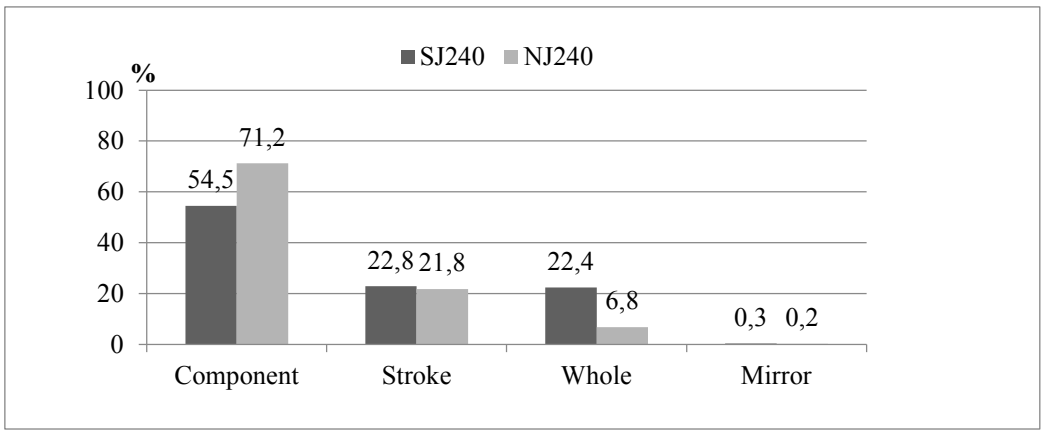

Figure 27 Breakdown of LV240 pseudokanji writing errors by percentage

The occurrence rate of each error type was compared between SJ240 and NJ240 through a configural frequency analysis. There was a significant difference (df = 3, p < 0.0125) between the two groups for two out of the four types, namely the component type (x^2 = 27.93, p = 3.7613E-06) and the whole type (x^2 = 255.23, p = 0.0000). The differences for the stroke type (x^2 = 0.34, p = 0.9404) and the mirror type (x^2 = 0.40, p = 0.9403) were at an insignificant level.

SJ240 made significantly fewer errors of the component type and had a significantly higher rate for the whole subtype. Since configurational awareness is expected to be better developed for making the component subtype and less so for making the whole subtype, it appears that SJ240 learners displayed less well-developed configurational awareness than NJ240 learners.

6.4.3. Summary of LV240 writing results

This subsection summarises LV240 writing error occurrence pattern analyses in subsections 6.4.1 and 6.4.2, as listed in Table 42:

Table 42 Summary of LV240 writing error occurrence patterns

Error occurrence patterns	Group	Similarities	Differences
6.4.1. Overall writing	SJ240 NJ240	Pseudokanji (overwhelming majority) > Orthographic > Phonological > Circumstantial > Semantic	• Pseudokanji: SJ240 > NJ240 • Orthographic: NJ240 > SJ240
6.4.2. Pseudokanji breakdown	SJ240 NJ240	Component > Stroke > Whole > Mirror	• Component: NJ240 > SJ240 • Whole: SJ240 > NJ240

6.5. LV800 writing results

In this section, LV800 writing results will be compared between the Swedish participants (SJ800) and the Japanese participants (NJ800). The overall results will be presented and analysed in subsection 6.5.1, followed by breakdowns of pseudokanji errors in subsection 6.5.2. Finally, subsection 6.5.3 will summarise the error occurrence patterns observed in subsections 6.5.1 and 6.5.2.

6.5.1. Overall LV800 writing results

The two groups showed similar general occurrence patterns of writing error types: the pseudokanji type was by far the most common, the second and third most common phonological and orthographic types accounted for a much lower rate, and the circumstantial and semantic types occupied insignificantly small percentages. The rate of the pseudokanji type was considerably smaller than in LV240's corresponding results, which can be interpreted as showing that the LV800 learners' configurational awareness of kanji components is better developed than that of LV240 learners. The phonological type accounted for over 20% for both groups, in contrast with a much lower figure in LV240's corresponding results. The most feasible explanation for this is a substantially larger stock of kanji homophones within the LV800 learners' mental lexicon. In other words, they know more characters with the same reading that can therefore get confused.

Table 43 is a list of the frequencies and rates of overall error type occurrences of LV800 writing results in comparison between SJ800 and NJ800. The occurrence rates are graphed as Figure 28.

Table 43 Frequencies and rates of LV800 overall writing error type occurrences

	Phonological	Circum-stantial	Ortho-graphic	Semantic	Pseudokanji	Total
SJ800	176 (22.9%)	25 (3.3%)	63 (8.2%)	11 (1.4%)	493 (64.2%)	768 (100%)
NJ800	537 (22.6%)	102 (4.3%)	579 (24.3%)	47 (2.0%)	1113 (46.8%)	2378 (100%)

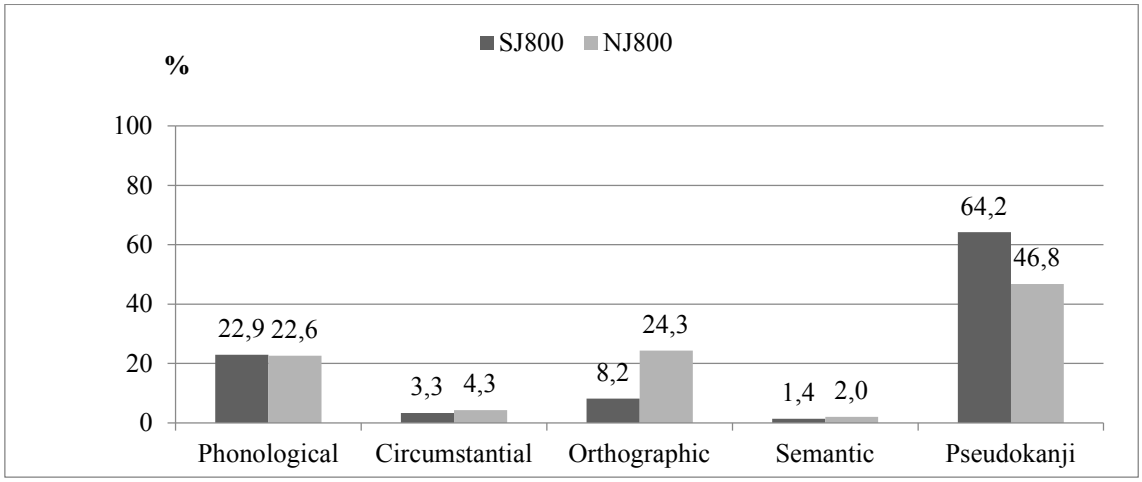

Figure 28 LV800 overall writing error type occurrence rates

The occurrence rate of each error type was compared between SJ800 and NJ800 through a configural frequency analysis. There was a significant difference ($df = 4$, $p < 0.01$) between the two groups for two out of the five types, namely the orthographic type ($x^2 = 82.22$, $p = 0.00$) and the pseudokanji type ($x^2 = 49.61$, $p = 4.3455E-10$). The differences for the phonological type ($x^2 = 0.04$, $p = 0.9998$), the circumstantial type ($x^2 = 1.91$, $p = 0.7514$) and the semantic type ($x^2 = 1.15$, $p = 0.8862$) were at an insignificant level.

LV800's writing results show differences that are parallel to LV240's, i.e. the higher percentage of the pseudokanji type for SJ learners and the lower percentage of the orthographic type. They indicate, again in parallel to LV240, that SJ800 has a less developed sense of the basic configuration of kanji than NJ800 and that they are less orthographically oriented in their approach to retrieving the characters than SJ800.

The notable difference of the pseudokanji types between the two groups will be examined through subtype analysis in subsection 6.5.2.

6.5.2. Breakdown of LV800 pseudokanji writing errors

The two groups showed the same general occurrence pattern of pseudokanji error types: the component subtype was in the overwhelming majority, the stroke and whole subtypes, which were the second and third most common, accounted for much smaller percentages, and the mirror subtype was practically non-existent.

The frequencies and rates of pseudokanji errors falling into each subtype (component, stroke, whole or mirror, as stated in subsection 5.3.5) are listed in Table 44 and the percentage figures are graphed in Figure 29:

Table 44 Breakdown of LV800 pseudokanji writing errors in frequency and rate

	Component	Stroke	Whole	Mirror	Total
SJ800	360 (73.0%)	87 (17.7%)	44 (8.9%)	2 (0.4%)	493 (100%)
NJ800	941 (84.5%)	83 (7.5%)	89 (8.0%)	0 (0.0%)	1113 (100%)

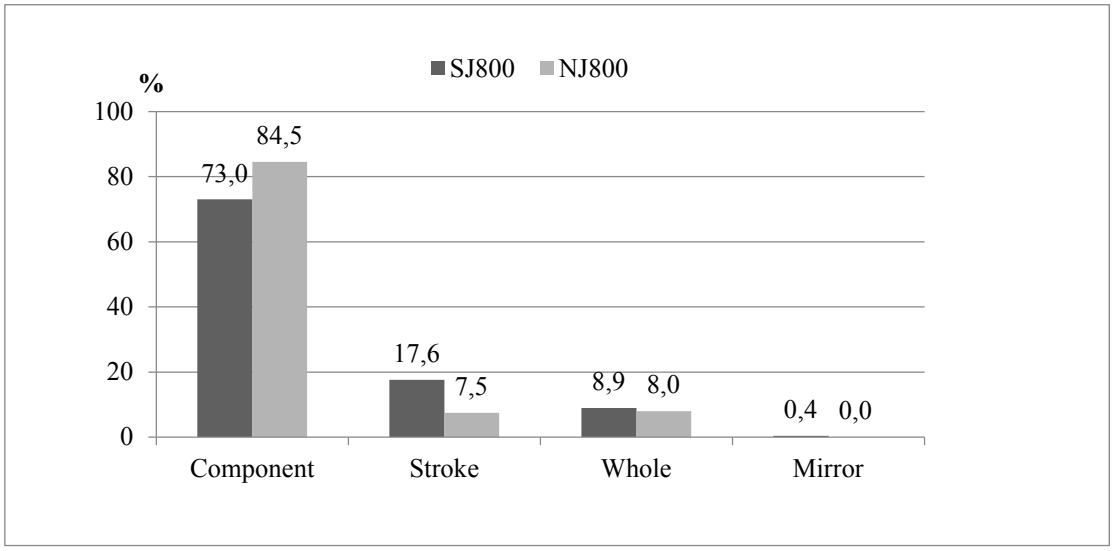

Figure 29 Breakdown of LV800 pseudokanji writing errors by percentage

The occurrence rate of each error type was compared between SJ800 and NJ800 through a configural frequency analysis. There was a significant difference ($df = 3$, $p < 0.0125$) between the two groups for two out of the four types, namely the stroke type ($x^2 = 68.64$, $p = 8.3267\text{E-}15$) and the mirror type ($x^2 = 9.03\text{E+}12$, $p = 0.0000$). The differences for the component type ($x^2 = 7.74$, $p = 0.0516$) and the whole type ($x^2 = 0.53$, $p = 0.9119$) were at an insignificant level.

The rather high ratio of the pseudokanji type does imply that LV800 learners' configurational awareness is not fully developed yet. Nevertheless, the percentages of their component subtype errors (73.0% for SJ800 and 84.5% for NJ800) are markedly higher than those of the LV240 learners (54.5% for SJ240 and 71.2% for NJ240). Since the component subtype implies a higher developmental degree of configurational awareness than the stroke and whole subtypes, as mentioned in subsections 5.3.5 and 6.4.2, the predominance of the component subtype indicates that the LV800 learners have a more developed configurational awareness than LV240 learners.

Similarly to the corresponding results for LV240, the mirror subtype was practically non-existent at LV800. Since the difference in level did not affect the results, the explanation suggested for LV240 can be repeated for LV800: errors of the mirror subtype are too basic for these learners to be making, and/or the kanji is too orthographically complex to reverse the structure to produce a mirror image.

SJ800 learners' significantly higher percentage of the stroke type suggests that they are more oblivious to the non-existent components involved in the pseudokanji character produced, which indicates that their mental screening

for non-existent components malfunctions more often than NJ800 learners'. As for SJ800's statistically higher percentage of the mirror type, this may well be due to the difference in the number of participants with mirror writing inducing conditions such as learning disability or some form of cognitive impairment, rather than a general tendency as a group, since (i) there were only two cases of the mirror subtype; and (ii) among the conditions under which mirror writing may occur, immaturity and ageing can be excluded based on SJ800's age range (20 to 30).

6.5.3. Summary of LV800 writing results

This subsection summarises LV800 writing error occurrence pattern analyses in subsections 6.5.1 and 6.5.2, as listed in Table 45:

Table 45 Summary of LV800 writing error occurrence patterns

Error occurrence patterns	Group	Similarities	Differences
6.5.1. Overall writing	SJ800	Pseudokanji > Orthographic > Phonological > Circumstantial > Semantic	• Pseudokanji: SJ800 > NJ800 • Orthographic: • NJ800 > SJ800
6.5.2. Pseudokanji breakdown	SJ800 NJ800	Component > Stroke > Whole > Mirror	• Stroke: SJ800 > NJ800 • Mirror: • SJ800 > NJ800

6.6. Inter-level comparison of writing results

This section will make inter-level comparisons of writing error occurrence patterns between groups of the same L1 background (SJ240 vs. SJ800, and NJ240 vs. NJ800). Subsection 6.6.1 will compare overall writing error results, subsection 6.6.2 will deal with the pseudokanji breakdown results, and subsection 6.6.3 will summarise the error occurrence patterns observed in subsections 6.6.1 and 6.6.2.

For the same reasons stated in section 6.3, LV240 and LV800 groups of the same L1 background will be compared first, and then the inter-level shift of error making tendencies will be compared between SJ and NJ (SJ240-SJ800 shift vs. NJ240-NJ800 shift) to examine the pattern of shift.

6.6.1. Inter-level comparison of overall writing error results

This subsection will make an inter-level comparison of overall writing error occurrence patterns. Figure 30 shows the comparison between SJ240 and SJ800 and Figure 31 between NJ240 and NJ800:

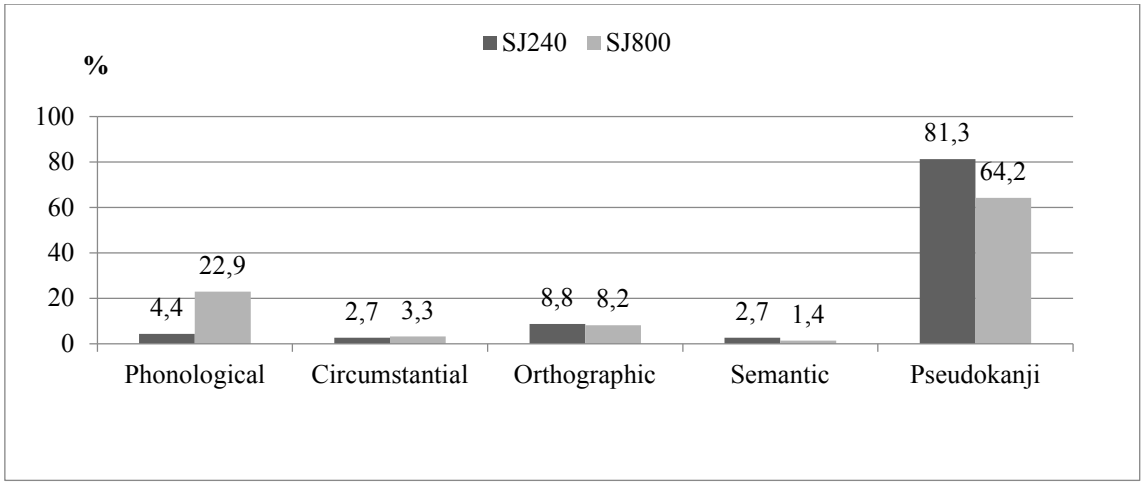

Figure 30 Inter-level comparisons of overall writing error rates

The occurrence rate of each error type between SJ240 and SJ800 was compared through a configural frequency analysis. There was a significant difference ($df = 4$, $p < 0.01$) between the two groups for two out of the five types, namely the pseudokanji type ($x^2 = 39.93$, $p = 4.4653E\text{-}08$) and the phonological type ($x^2 = 131.28$, $p = 0.0000$). The differences for the circumstantial type ($x^2 = 0.69$, $p = 0.9529$), the orthographic type ($x^2 = 0.41$, $p = 0.9821$) and the semantic type ($x^2 = 10.57$, $p = 0.0319$) were at an insignificant level.

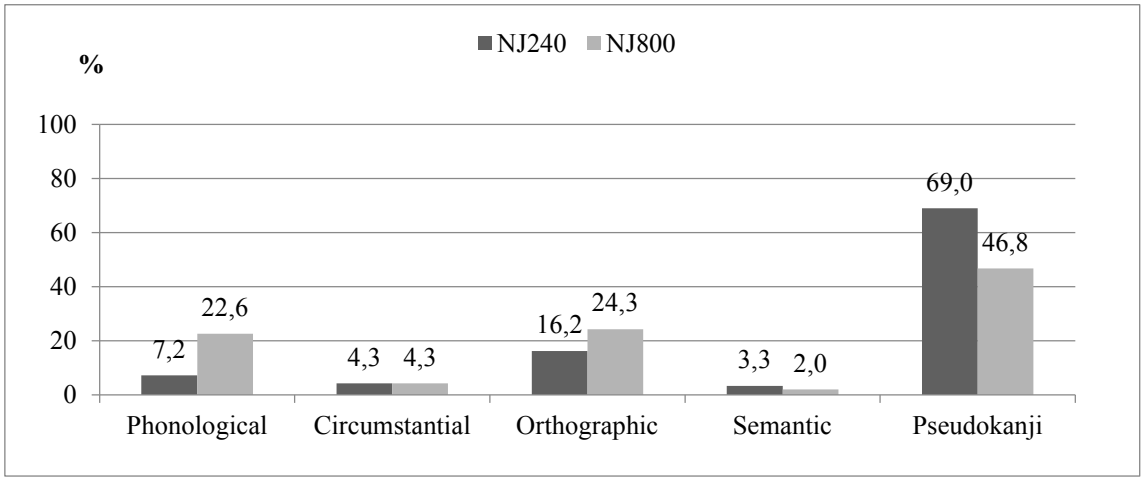

Figure 31 Inter-level comparisons of overall writing error rates

The occurrence rate of each error type was compared between NJ240 and NJ800 through a configural frequency analysis. There was a significant difference ($df = 4$, $p < 0.01$) between the two groups for four out of the five types, namely the pseudokanji type ($x^2 = 168.12$, $p = 0.0000$), the phonological type ($x^2 = 167.36$, $p = 0.0000$), the orthographic type ($x^2 = 43.64$, $p = 7.6339\text{-E}09$) and the semantic type ($x^2 = 14.31$, $p = 0.0064$). The difference for the circumstantial type ($x^2 = 0.00$, $p = 1.0000$) was at an insignificant level.

Since there were four types with significant differences between the two NJ groups and only two significantly different types for the SJ groups, inter-level difference is greater for NJ groups than SJ groups. Although both groups share a decrease in the pseudokanji type at LV800, the shift in percentage is from the pseudokanji type to the phonological type for SJ groups, while the shift is divided into the phonological type and the orthographic type for NJ groups.

6.6.2. Inter-level comparison of pseudokanji errors

This subsection will make an inter-level comparison of pseudokanji error subtype results. Figure 32 shows the comparison between SJ240 and SJ800 and Figure 33 between NJ240 and NJ800:

Figure 32 Inter-level comparison of SJ groups' pseudokanji error subtypes

The occurrence rate of each error type was compared between SJ240 and SJ800 through a configural frequency analysis. There was a significant difference ($df = 3$, $p < 0.0125$) between the two groups for two out of the four types, namely, the component type ($x^2 = 33.33$, $p = 2.7408\text{E-}07$) and the whole type ($x^2 = 144.33$, $p = 0.0000$). The differences for the stroke type ($x^2 = 10.75$, $p = 0.0131$) and the mirror type ($x^2 = 0.27$, $p = 0.9657$) were at an insignificant level.

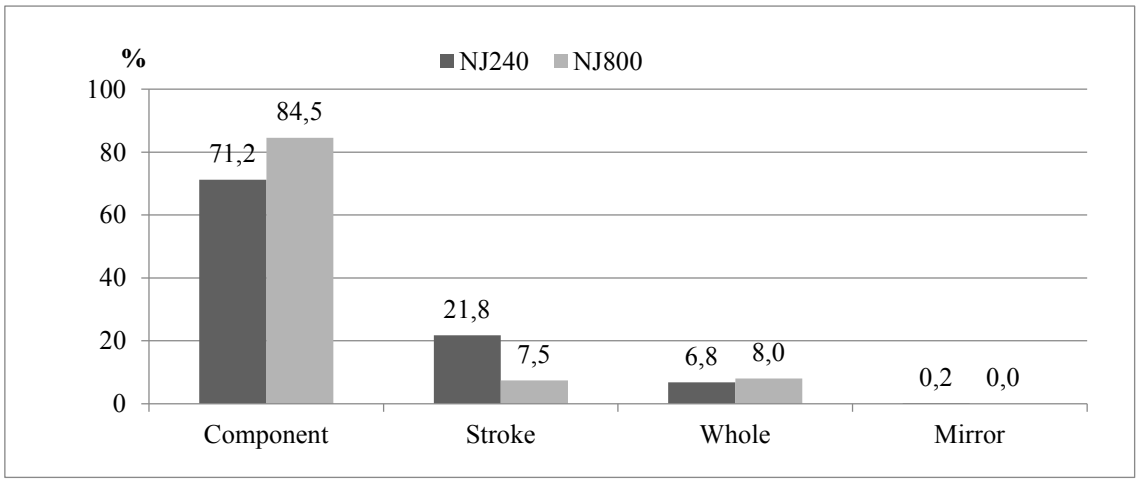

Figure 33 Inter-level comparisons of NJ groups' pseudokanji error subtypes

The occurrence rate of each error type was compared between NJ240 and NJ800 through a configural frequency analysis. There was a significant difference ($df = 3$, $p < 0.0125$) between the two groups for three out of the four types, namely the component type ($x^2 = 23.13$, $p = 3.7851\text{E-}05$), the stroke type ($x^2 = 304.68$, $p = 0.0000$) and the mirror type ($x^2 = 4.03\text{E+}12$, p

= 0.0000). The difference for the whole type ($x^2 = 2.04$, $p = 0.5636$) was at an insignificant level.

Both SJ800 and NJ800 show a higher rate of the component subtype and a lower rate of the whole and/or the stroke subtype than their respective L1-matched counterpart, which indicates that the advanced groups have developed better configurational awareness than their novice counterparts. Since there were three subtypes with significant differences for the NJ groups and only two for the SJ groups, the inter-level difference is greater for the NJ groups than for the SJ groups.

6.6.3. Summary of inter-level comparison of writing error results

This subsection summarises the inter-level comparison of writing error results in subsections 6.6.1 and 6.6.2, as listed in Table 46:

Table 46 Summary of inter-level comparison of writing error results

Error occurrence patterns	Group	SJ/NJ Similarities	SJ/NJ differences
6.6.1. Overall writing	SJ240 SJ800	• LV240 Pseudokanji > LV800 Pseudokanji • LV800 Phonological > LV240 Phonological	Greater LV240/LV800 difference for NJ
	NJ240 NJ800		
6.6.2. Pseudokanji breakdown	SJ240 SJ800	• LV800 Component > LV240 Component	• SJ240 Whole > SJ800 Whole
	NJ240 NJ800	•	• NJ240 Stroke > NJ800 Stroke

6.7. Comparison between reading and writing results

In this section, differences between the reading and writing error type occurrence patterns will be investigated by comparing the reading and writing results of each group, namely, SJ240 reading vs. SJ240 writing, NJ240 reading vs. NJ240 writing, SJ800 reading vs. SJ800 writing, and NJ800 reading vs. NJ800 writing. In order to make a proper interskill comparison, the writing error type occurrence rates have been recalculated without the pseudokanji type, which is not included in the reading results, as listed in Table 47. In other words, comparisons will be made only for the types of errors in which the target kanji character or reading is substituted with an existing non-target character or reading.

Table 47 Frequencies and rates of LV240 and LV800 reading and writing error type occurrence rates (without pseudokanji type)

	Phonological	Circumstantial	Orthographic	Semantic	Total
SJ240 Reading	180 (55.2%)	101 (31.0%)	21 (6.4%)	24 (7.4%)	326 (100%)
SJ240 Writing	38 (23.3%)	24 (14.7%)	77 (47.2%)	24 (14.7%)	163 (100%)
NJ240 Reading	198 (34.4%)	327 (56.9%)	23 (4.0%)	27 (4.7%)	575 (100%)
NJ240 Writing	116 (23.3%)	69 (13.9%)	260 (52.2%)	53 (10.6%)	498 (100%)
SJ800 Reading	285 (47.7%)	237 (39.7%)	64 (10.7%)	11 (1.8%)	597 (100%)
SJ800 Writing	176 (64.0%)	25 (9.1%)	63 (22.9%)	11 (4.0%)	275 (100%)
NJ800 Reading	309 (29.4%)	682 (64.9%)	40 (3.8%)	20 (1.9%)	1058 (100%)
NJ800 Writing	537 (42.5%)	102 (8.1%)	579 (45.8%)	47 (3.7%)	1265 (100%)

Figure 34 shows various in-group comparisons of reading and writing error type occurrence rates:

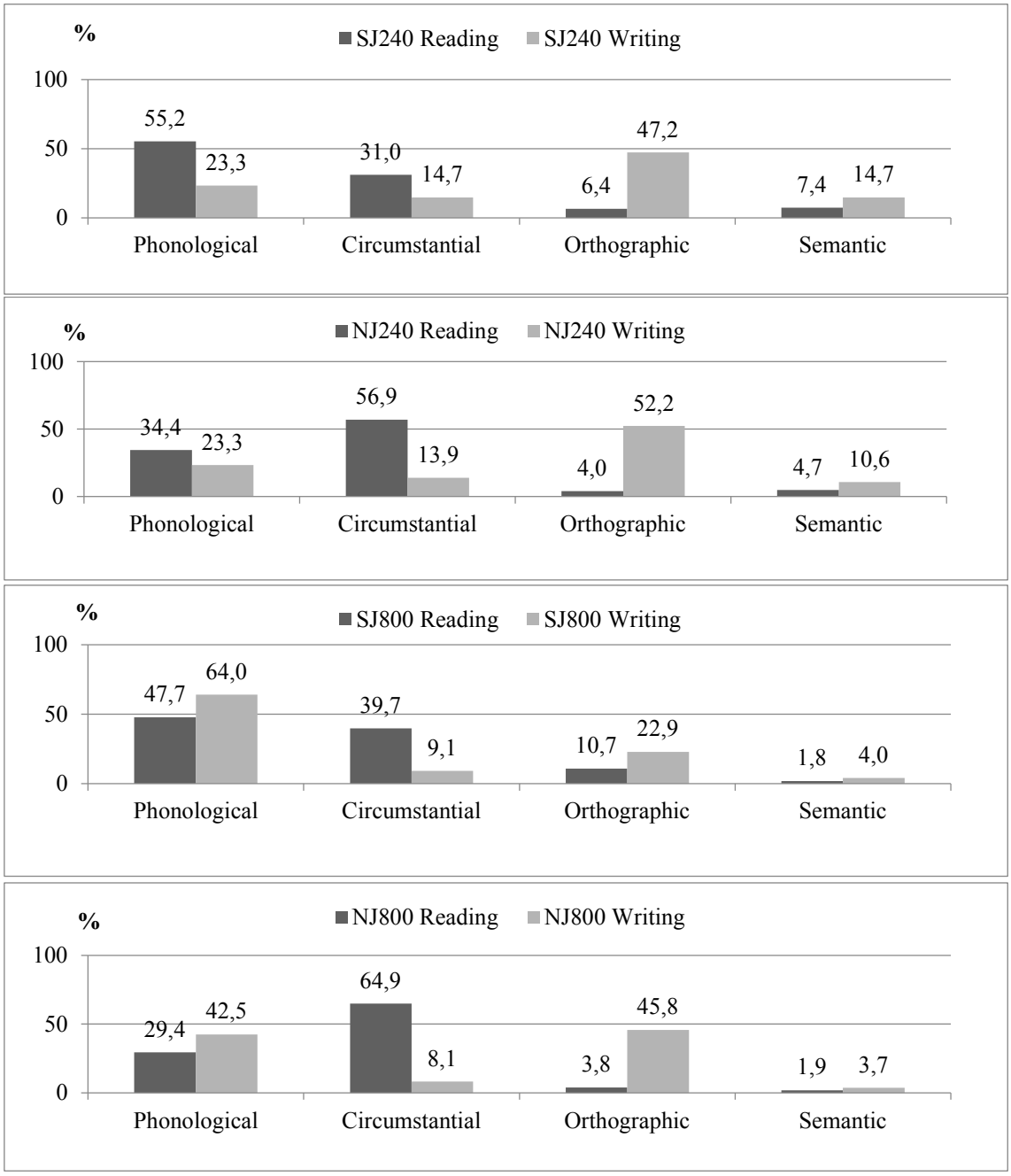

Figure 34 In-group comparisons of overall reading and writing error rates

The occurrence rate of each error type between reading and writing for each group was compared through a configural frequency analysis. A significant difference ($df = 3$, $p < 0.0125$) between the two skills was found for all four types for SJ240 and NJ240 and for three types (with the exception of the semantic type) for SJ800 and NJ800, as listed in Table 48:

Table 48 Error types with significant difference between reading and writing

SJ240	Phonological	Circumstantial	Orthographic	Semantic
p-value	0	1.2168E-12	0	0.0074
NJ240	Phonological	Circumstantial	Orthographic	Semantic
p-value	1.01051E-06	0	0	0.000259749
SJ800	Phonological	Circumstantial	Orthographic	Semantic
p-value	1.81302E-05	0	1.99358E-08	0.073606331
NJ800	Phonological	Circumstantial	Orthographic	Semantic
p-value	3.70321E-09	0	0	0.025643849

As shown in Figure 34, there are clear leanings towards the phonological and circumstantial types in all four groups' reading results. It can therefore be concluded that this tendency is characteristic of the reading skill rather than of the level or L1 background. On the other hand, the predominance of the other two error types (phonological for SJ learners and circumstantial for NJ learners) is decided by the group's L1 background.

In writing, the leaning is towards orthographic and phonological error types and the level is the deciding factor in predominance: the orthographic type had a clear majority at LV240, but at LV800 it becomes less distinct and the rate of the phonological type becomes significantly higher than its reading equivalent.

Characteristics of each error type can be summarised as follows:

The phonological type is common in both reading and writing for all groups. Its rate of occurrence shifts according to the level, i.e. it is more common in reading for LV240 but in writing for LV800. The circumstantial type is common in reading, and the rate is higher for NJ groups than for SJ groups. The orthographic type is definitely more common in writing, although this tendency is less extreme for SJ800. The semantic type is significantly higher in writing than reading for LV240, while there is no such interskill difference for LV800. This suggests that the semantic aspect of kanji is of greater importance in writing than in reading only for novice learners.

7. DISCUSSION

This chapter will discuss and analyse results described in Chapter 6 and will clarify the characteristics of Swedish learners of Japanese with regard to their kanji interlanguage in the developmental process of kanji learning. Section 7.1 will present hypotheses to the research questions raised in section 1.4, and will consider them in the light of the results. Section 7.2 will discuss the aspects of comparison that are not covered in section 7.1. Section 7.3 is a comprehensive discussion of the characteristics of Swedish learners' kanji learning process. Section 7.4 will refer to limitations of this study and suggest future courses of research.

7.1. Research hypotheses and their verifications

Research hypotheses were formulated based on the observations and implications of the previous studies stated in Chapter 3. Subsections 7.1.1, 7.1.3 and 7.1.5 will present hypotheses regarding the research questions A, B and C respectively, and each of subsections 7.1.2, 7.1.4 and 7.1.6 will verify the hypothesis presented in the immediately preceding subsection.

7.1.1. Hypothesis A on general error type occurrence patterns

Research Question A:

What are the kanji error type occurrence patterns and their similarities and differences according to skill (reading/writing), level (LV240/LV800) and L1 (SJ/NJ)?

Hypothesis A:

(1) Predominant reading error types: phonological for the SJ groups, and phonological-circumstantial for the NJ groups.

(2) Predominant writing error types: pseudokanji for SJ240 and NJ240, phonological-pseudokanji for SJ800 and phonological-orthographic for NJ800.

(3) General similarities and minor differences in error type occurrence tendencies between the level-matched groups (similar in reading between L1WS-matched groups and in writing between level-matched groups).

For LV240 reading, the phonological type is expected to predominate due to the learner groups' insufficient knowledge of *kana* orthography (Koike et

al., 2003), although the circumstantial type may also be strong for NJ240 learners because of their dependence on non-kanji linguistic skills (BAASC, 2007). On the other hand, the pseudokanji type is expected to account for the majority of LV240 writing errors, having conditions similar to those of the Australian novice learners in Hatta et al. (2002). The underdeveloped orthographic awareness (Matsumoto, 2013) of novice L2 learners may affect SJ240's results negatively and therefore they may show a higher rate of the pseudokanji type than the NJ240 learners.

An expected common feature of the LV800 groups is a predominance of phonological errors among their reading errors due to the amount of kanji homophones learned by this time (BAASC, 2007). Nevertheless, L2 learners' dependence on phonological information and preference for a phonological approach to decoding (rather than whole-character processing) may boost SJ800's phonological inclination, whereas L1 learners' dependence on non-kanji linguistic skills (BAASC 2007) may shift the inclination of NJ800 learners towards the circumstantial error type. As for LV800 writing, the large stock of kanji homophones available (BAASC, 2007) predicts an inclination to phonological errors for both groups. An expected SJ/NJ difference is SJ's higher percentage of the pseudokanji type due to L2 learners' limited configurational awareness.

Table 49 is a list of error types expected to be predominant (in bold type) and their grounds (in normal type) according to skill, level and L1WS background:

Table 49 Expected inter-group differences in error type occurrence patterns

	Reading	Writing
SJ240	**Phonological** • Insufficient knowledge of *kana* orthography (Koike et al., 2003)	**Pseudokanji** • Non-kanji predominance for novice group (Hatta et al., 2002) • Underdeveloped orthographic awareness (Matsumoto, 2013)
NJ240	**Phonological** • Insufficient knowledge of *kana* orthography (Koike et al., 2003) **Circumstantial** • Dependence on non-kanji linguistic skills (BAASC, 2007)	**Pseudokanji** • Non-kanji error predominance for novice group (Hatta et al., 2002)
SJ800	**Phonological** • Large enough kanji homophone stock at 640 character level (BAASC, 2007) • Dependent on phonological information; orthographic and phonological decoding • (Leong & Tamaoka, 1995; Tamaoka, 1992; Komori, 2009)	**Phonological** • Large enough kanji homophone stock at 640 character level (BAASC, 2007) **Pseudokanji** • Limited orthographic awareness (Komori, 2009) • Large gap between recognition and production skills (Chikamatsu, 2005) • Underdeveloped configurational awareness (Okita, 2001; Hatta, et al., 1998; 2002)
NJ800	**Phonological** • Large enough kanji homophone stock at 640 character level (BAASC, 2007) **Circumstantial** • Dependence on non-kanji linguistic skills (BAASC 2007)	**Phonological** • Large enough kanji homophone stock at 640 character level (BAASC, 2007) • L1 adults' phonological inclination for character retrieval • (Hatta et al., 1998; 2002) **Orthographic** • Later development (still limited at Grade 7) of radical awareness (Koike et al., 2003; Hatta et al., 2002)

The level-matched groups will show similar general error patterns with minor differences depending on the L1WS. In other words, the similarities represent error occurrence patterns that are characteristic of the level, and the differences indicate the characteristics of the groups' L1WS backgrounds.

Each of the learner groups is expected to show general error patterns that are more similar to those of level-matched counterparts than to those of L1WS-matched counterparts (e.g. SJ240 will show results that are more similar to those of NJ240 than to SJ800's). The grounds for this are:

- the difference in number and in the intrinsic features and extrinsic factors of kanji that have been learned at each level (e.g. more simple-structured kanji with basic and concrete meanings for novice learners, ratio of homophonous characters, etc.), which can affect the ratio of each error type occurrence, as pointed out by BAASC (2007); and

- the findings of Hatta et al. (2002) that different levels of learner groups showed different error type patterns.

Nevertheless, the level-matched SJ/NJ groups are expected to show minor differences in tendencies from each other. As pointed out by Cook & Bassetti (2005), L2 learners have been reported to behave differently from L1 learners in L2WS research in general. Furthermore, the different degrees of development of phonological and orthographic awareness of learner groups observed in previous kanji research, such as Komori (2009) and Matsumoto (2013), are likely to contribute to L1WS background based differences.

7.1.2. Verification of Hypothesis A

In this subsection, Hypothesis A will be verified based on the results set out in Chapter 6.

Hypothesis A (1):

Predominant reading error types: phonological for the SJ groups, and phonological-circumstantial for the NJ groups.

Table 50 is a comparative summary of overall reading results from subsections 6.1.1, 6.2.1 and 6.3.1. Figures 8, 13, 18 and 19 from these subsections are combined in sequence as Figure 35, so they will be readily comparable.

As shown in Figure 35, the phonological and circumstantial types were predominant over orthographic and semantic types for all four groups, which suggests this error occurrence pattern is characteristic of the skill rather than level- or L1WS-specific.

The orders of the top two error types were phonological-circumstantial for the SJ groups, and circumstantial-phonological for the NJ groups. Therefore, Hypothesis A (1) is verified in principle, although the circumstantial type gained higher rates than expected, and as a result, the predominant pattern of reading error type turned out to be phonological-circumstantial for SJ groups and circumstantial-phonological for NJ groups.

Table 50 Comparative summary of LV240/LV800 overall reading error occurrence patterns

Error occurrence patterns	Group	Similarities	Significant differences
6.1.1. LV240 Overall reading	SJ240	Phonological + Circumstantial > Orthographic + Semantic	• Phonological SJ240 > NJ240 • Circumstantial NJ240 > SJ240
	NJ240		
6.2.1. LV800 Overall reading	SJ800	Phonological + Circumstantial > Orthographic + Semantic	• Phonological SJ800 > NJ800 • Circumstantial NJ800 > SJ800 • Orthographic NJ800 > SJ800
	NJ800		
6.3.1. Inter-level comparison of overall reading results	SJ240	Phonological > Circumstantial	Semantic: SJ240 > SJ800
	SJ800		
	NJ240	Circumstantial > Phonological	Semantic: NJ240 > NJ800
	NJ800		

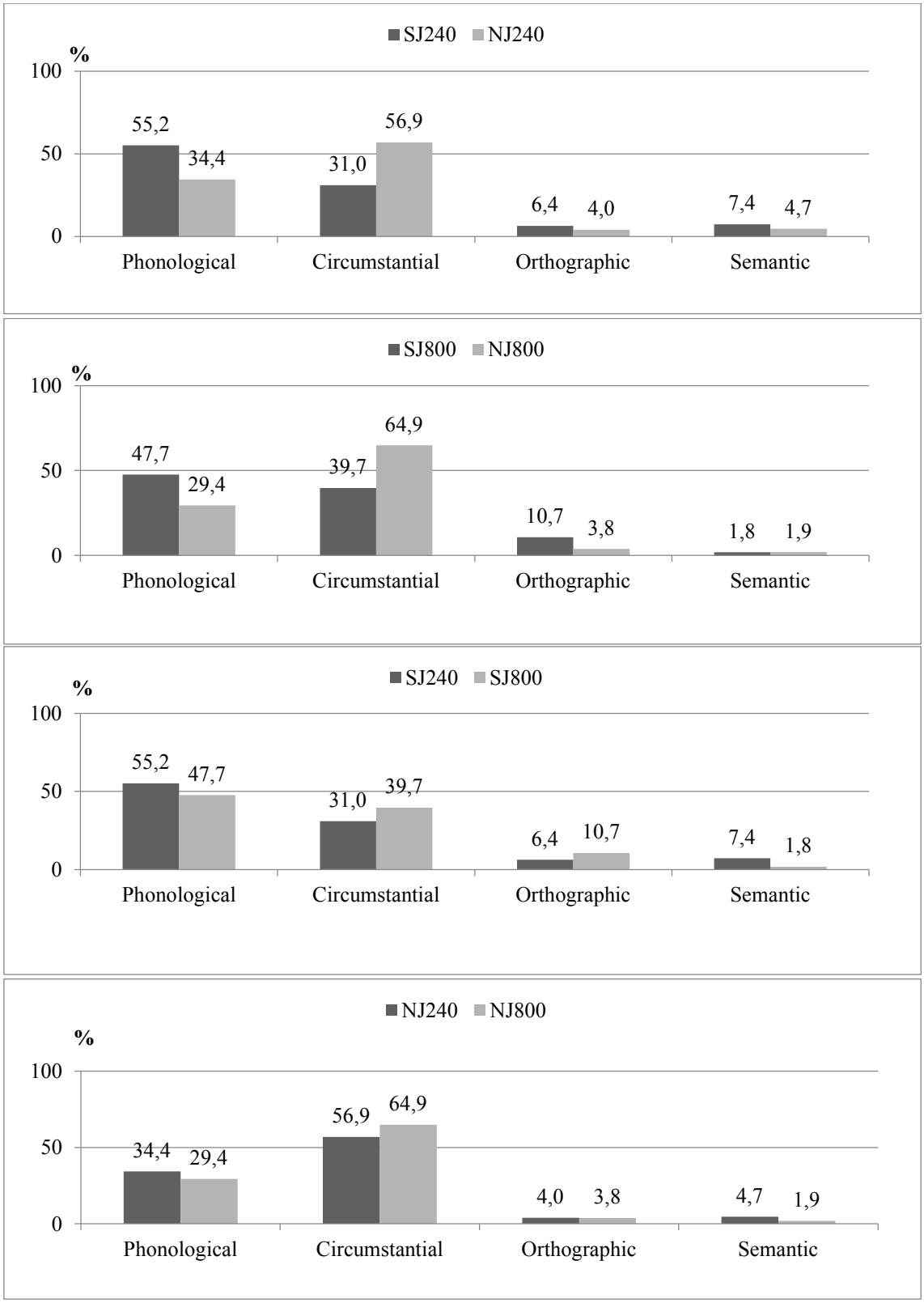

Figure 35 Collected overall reading error rate figures (Figures 8, 13, 18 & 19)

Hypothesis A (2):

Predominant writing error types: pseudokanji for SJ240 and NJ240, phonological-pseudokanji for SJ800 and phonological-orthographic for NJ800.

Table 51 is a comparative summary of the overall writing error results from subsections 6.4.1, 6.5.1 and 6.6.1. Figures 26, 28, 30 and 31 from these subsections are combined in sequence as Figure 36, so they will be readily comparable.

As seen in Table 51 and Figure 36, the predominance of pseudokanji errors for LV240 groups, as predicted by the hypothesis, is verified. However, the pseudokanji type was strong even in the LV800 groups and the predominant types for them turned out to be pseudokanji-phonological for SJ800 and pseudokanji-orthographic-phonological for NJ800.

Table 51 Comparative summary of LV240/LV800 overall writing error occurrence patterns

Error occurrence patterns	Group	Similarities	Significant differences
6.4.1. LV240 Overall writing	SJ240	Pseudokanji > Orthographic + Phonological + Circumstantial + Semantic	• Pseudokanji SJ240 > NJ240 • Orthographic NJ240 > SJ240
	NJ240		
6.5.1. LV800 Overall writing	SJ800		• Pseudokanji SJ800 > NJ800 • Orthographic NJ800 > SJ800
	NJ800		
6.6.1. Inter-level comparison of overall writing results	SJ240	Pseudokanji > Orthographic Phonological + Circumstantial + Semantic	• Pseudokanji: SJ240 > SJ800 • Phonological SJ800 > SJ240
	SJ800		
	NJ240	Pseudokanji predominance	• Pseudokanji & Semantic: NJ240 > NJ800 • Phonological & Orthographic NJ800 > NJ240
	NJ800		

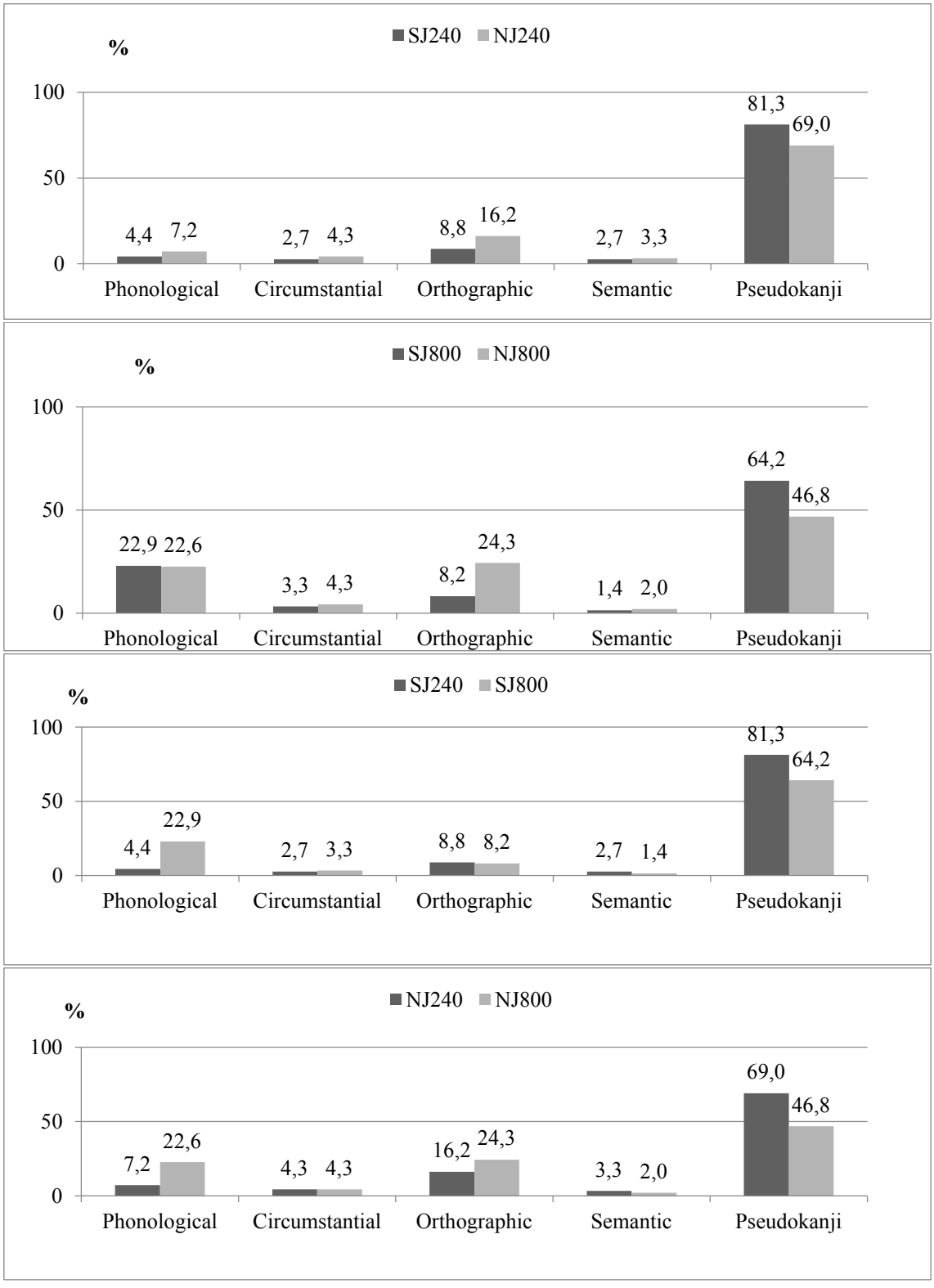

Figure 36 Collected overall writing error rate figures (Figures 26, 28, 30 & 31)

Hypothesis A (3):

General similarities and minor differences in error type occurrence tendencies between the level-matched groups (similar in reading between L1WS-matched groups and in writing between level-matched groups).

As shown in Table 50 and Figure 35, when the level-matched SJ and NJ groups were compared, the difference was a reversal of the two major error types for both comparisons (phonological-circumstantial for SJ and circumstantial-phonological for NJ). On the other hand, in the inter-level comparisons of the L1WS-matched groups, the difference was a reversal of the two minor error types for the SJ groups (semantic-orthographic for SJ240 and orthographic-semantic for SJ800). It may therefore be concluded that the L1WS-matched groups show greater similarity in reading error type occurrence patterns than the level-matched groups, which verifies the hypothesis (similarity in reading between L1WS-matched groups).

On the other hand, as shown in Table 51 and Figure 36, the overall analyses of the writing error results show significant differences for two error types in the inter-level comparisons (pseudokanji/phonological or pseudokanji/semantic) and for one in the L1WS-matched comparisons (pseudokanji or orthographic). This verifies the hypothesis of the level-matched groups' similarity in writing error type occurrence patterns.

7.1.3. Hypothesis B on processing unit, retrieval and difficulties in recognition and production

Research Question B:

What are the units of processing, approaches to retrieval and difficulties in reading and writing of kanji of SJ and NJ in their respective developmental stages?

Hypothesis B:

(1) Unit of processing is mostly component-based for SJ and character-based for NJ.

(2) Approaches to retrieval are phonologically inclined for the SJ groups and circumstantially inclined for the NJ groups.

(3) All groups have difficulty in writing configurationally correct characters.

(4) The LV240 groups have difficulty in mastery of *kana* orthography.

With respect to the unit of processing, SJ240 is expected to be component-based in reading on the grounds of the L2 learners' inclinations for phonological decoding (Komori, 2009) and stroke-based in writing on the grounds of the high ratio of non-kanji type errors observed in Hatta et al. (1998, 2002). NJ240 is expected to be character-based in both reading and writing because of underdeveloped radical awareness (Chan & Nunes, 1998; Koike et al., 2003; and Hatta et al. 2002).

Based on Komori's observation of dependence on phonological information (2009) and Chikamatsu's analysis of L1 orthographic transfer and its effects on the L2 memory mechanism and retrieval strategy (2005), both SJ240 and SJ800 are expected to show phonological inclinations in their retrieval of kanji from the mental lexicon in reading tasks. Japanese schoolchildren's dependence on non-linguistic skills (BAASC, 2007) suggests that the NJ groups would have circumstantial retrieval approaches.

The observations of underdevelopment and later development of configurational awareness made by Chikamatsu (2005), Hatta et al. (1998, 2002), Koike et al. (2003) and Okita (2001) indicate that it is on the whole difficult for any of the four groups to write configurationally correct characters. The LV240 groups are expected to have additional difficulty in transcribing the reading with *kana* due to their incomplete mastery of *kana* orthography and/or underdeveloped phonological awareness. The phonological L1 transfer pointed out by Toda (2003) and Inoue (2008) serves as an additional difficulty for SJ240 in mastering *kana* orthography.

Table 52 is a list of hypothesised units of processing, inclinations of retrieval and difficulties in reading and writing of kanji of SJ and NJ in their respective developmental stages as well as the grounds for each point:

Table 52 Inclinations of retrieval and reading/writing difficulties for each group

	Reading	Writing
SJ240	**Unit of processing** • Component-based (Komori, 2009)	**Unit of processing** • Stroke-based (Hatta et al., 1998; 2002)
	Inclinations of retrieval • Phonological (Komori, 2009)	**Inclinations of retrieval** • Phonological • L1 orthographic transfer effects on L2 memory mechanism and retrieval strategy (Chikamatsu, 2005)
	Difficulties • Phonological awareness and *kana* orthography (Koike et al., 2003; Toda, 2003; Inoue, 2008)	**Difficulties** • Configurational correctness of character/component • (Chikamatsu, 2005; Okita, 2001; Hatta et al., 1998; 2002) • Character decompositional ability (Chikamatsu, 2005)
NJ240	**Unit of processing** • Character-based (Chan & Nunes, 1998)	**Unit of processing** Character-based (Koike et al., 2003; Hatta et al. 2002)
	Inclinations of retrieval • Circumstantial (dependence on non-kanji linguistic skills) (BAASC, 2007)	**Inclinations of retrieval** • Circumstantial (dependence on non-kanji linguistic skills) (BAASC, 2007)
	Difficulties • Special mora orthography (Koike et al., 2003)	**Difficulties** • Later development (still limited at Grade 7) of radical awareness (Koike et al., 2003; Hatta et al. 2002)
SJ800	**Unit of processing** Componential (Komori, 2009)	**Unit of processing** Componential (Chikamatsu, 2005)
	Inclinations of retrieval • Phonological (Komori, 2009)	**Inclinations of retrieval** • Phonological • L1 orthographic transfer effects on L2 memory mechanism and retrieval strategy (Chikamatsu, 2005)

(continued)

Table 52 *Continued*

	Difficulties • Complex/abstract words (Komori, 2009)	**Difficulties** • Configurational correctness gap between recognition and production (Chikamatsu, 2005) • Underdeveloped configurational awareness (Okita, 2001; Hatta et al., 1998; 2002)
NJ800	**Unit of processing** Character-based (Tamaoka, 1992)	**Unit of processing** Character-based (Hatta et al., 2002)
	Inclinations of retrieval • Circumstantial (dependence on non-kanji linguistic skills) (BAASC, 2007)	**Inclinations of retrieval** • Circumstantial (dependence on non-kanji linguistic skills) (BAASC, 2007)
	Difficulties • *On*-reading words (unfamiliar vocabulary) (BAASC, 2007)	**Difficulties** • Orthographic correctness, later development (still limited at Grade 7) of radical awareness (Koike et al., 2003; Hatta et al., 2002)

7.1.4. Verification of Hypothesis B

In this subsection, Hypothesis B will be verified based on the results presented in Chapter 6.

Hypothesis B (1):

Unit of processing is mostly component-based for SJ and character-based for NJ.

In the reading results, the SJ groups had high rates of the phonological reading error subtypes "transcription" and "component-based analogy." The unit of processing suggested by these subtypes are phoneme/mora-based for the former and component-based for the latter. The NJ groups had high rates of the circumstantial error type and the phonological reading error subtype "alternative reading", the former of which suggests their inclination towards word-based processing and the latter towards character-based processing (see subsections 5.2.1, 5.2.4 and 5.4.3 for clarifications and summary regarding the unit of processing).

Table 53 is a comparative summary of phonological reading error breakdowns from subsections 6.1.2, 6.2.2 and 6.3.2. Figures 9, 14, 20 and 21 from these subsections are presented in sequence as Figure 37, so they will be readily comparable.

As indicated in Table 53 and Figure 37, SJ240 learners, with limited configurational awareness, produced considerably higher rates of the transcription and alternative reading subtypes than of the component-based analogy subtype. Therefore, their unit of kanji reading processing is regarded as more mora-/phoneme- or character-based than component-based. SJ800, with a higher rate of component-based analogy, is deemed more component-based. On the other hand, the quite high rates of alternative reading errors for NJ240 and NJ800 indicate that they consistently process kanji on a whole character basis when reading.

Table 53 Comparative summary of LV240/LV800 phonological breakdowns

Error occurrence patterns	Group	Similarities	Significant differences
6.1.2. Phonological breakdown	SJ240 NJ240	Transcription + Alternative reading > Component-based analogy	• Transcription SJ240 > NJ240 • Alternative reading NJ240 > SJ240 • Component-based analogy SJ240 > NJ240
6.2.2. Phonological breakdown	SJ800 NJ800	Component-based analogy + Alternative reading > Transcription	• Transcription SJ800 > NJ800 • Alternative reading NJ800 > SJ800 • Component-based analogy SJ800 > NJ800
6.3.2. Inter-level comparison of phonological breakdowns	SJ240 SJ800	A little over 1/3 is alternative reading	• Transcription SJ240 > SJ800 • Component-based analogy SJ800 > SJ240
	NJ240 NJ800	Transcription + Alternative reading > Component-based analogy	• Transcription NJ240 > NJ800 • Alternative reading & component-based analogy NJ800 > NJ240

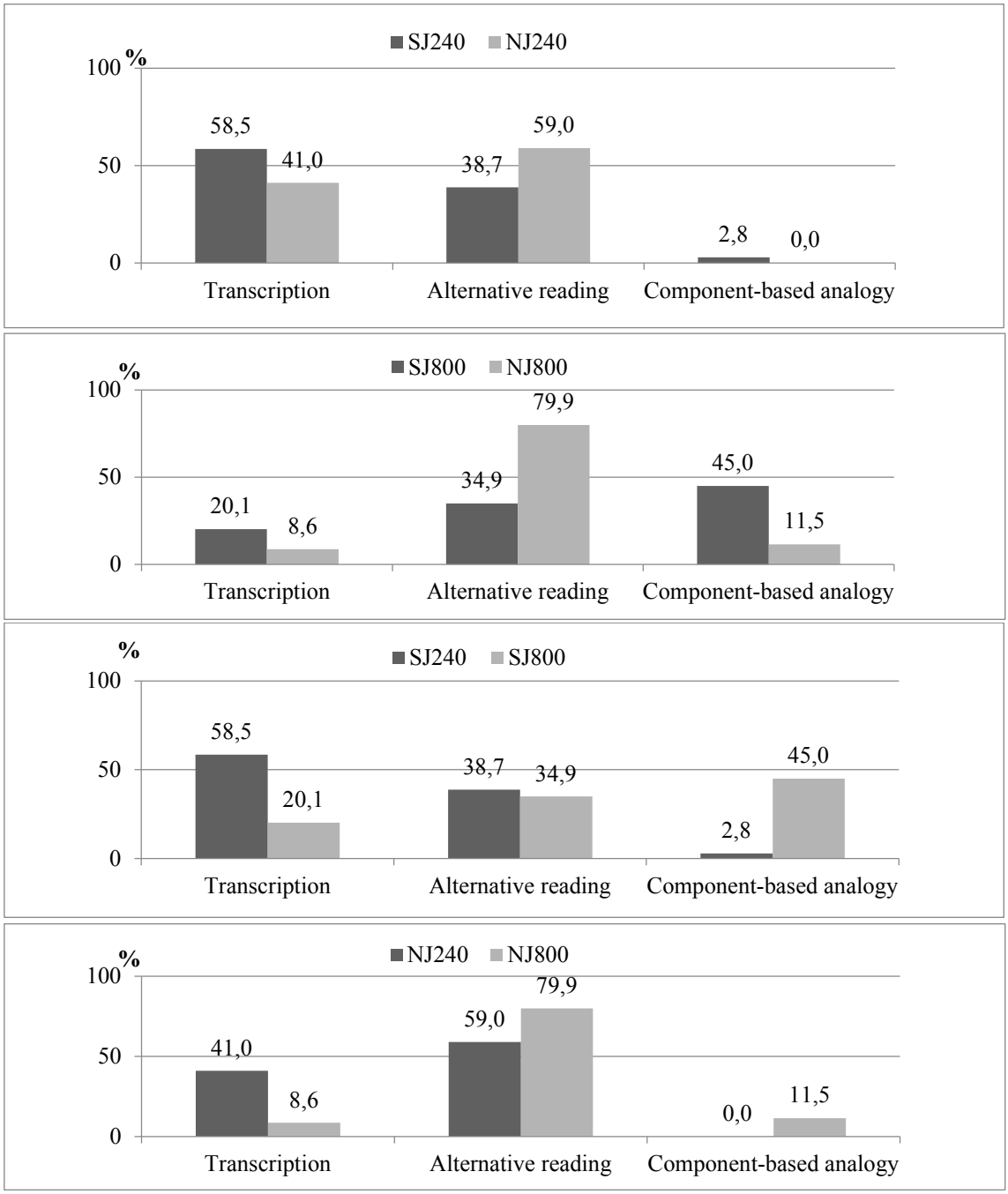

Figure 37 Collected figures of phonological breakdown (Figures 9, 14, 20 & 21)

In writing, the pseudokanji subtype "component" is associated with component-based processing, "stroke" and "whole" with stroke-based processing and "mirror" with character-based processing, as described in subsection 5.4.3.

Table 54 is a comparative summary of pseudokanji writing error breakdowns from subsections 6.4.2, 6.5.2 and 6.6.2. Figures 27, 29, 32 and

33 from these subsections are presented in sequence as Figure 38, so they will be readily comparable.

As indicated in Table 54 and Figure 38, the component subtype is predominant for all four groups, and therefore all of them basically use component-based units of processing in writing. In terms of relative differences, the SJ groups have leanings towards the stroke subtype in comparison with their level-matched NJ counterparts. In the inter-level comparisons, the LV240 groups are more stroke-based than their respective LV800 counterparts.

Table 54 Comparative summary of LV240/LV800 pseudokanji breakdowns

Error occurrence patterns	Group	Similarities	Significant differences
6.4.2. Pseudokanji breakdown	SJ240	Predominance of component Extremely low rates of mirror	Component: NJ240 > SJ240 Whole: SJ240 > NJ240
	NJ240		
6.5.2. Pseudokanji breakdown	SJ800	Absolute predominance of component, extremely low rates of mirror	Stroke and mirror: SJ800 > NJ800
	NJ800		
6.6.2. Inter-level comparison of pseudokanji breakdowns	SJ240	Component > Stroke > Whole > Mirror	• Component: SJ800 > SJ240 • Whole: SJ240 > SJ800
	SJ800		
	NJ240	Absolute predominance of component, extremely low rates of mirror	• Component: NJ800 > NJ240 • Stroke: NJ240 > NJ800
	NJ800		

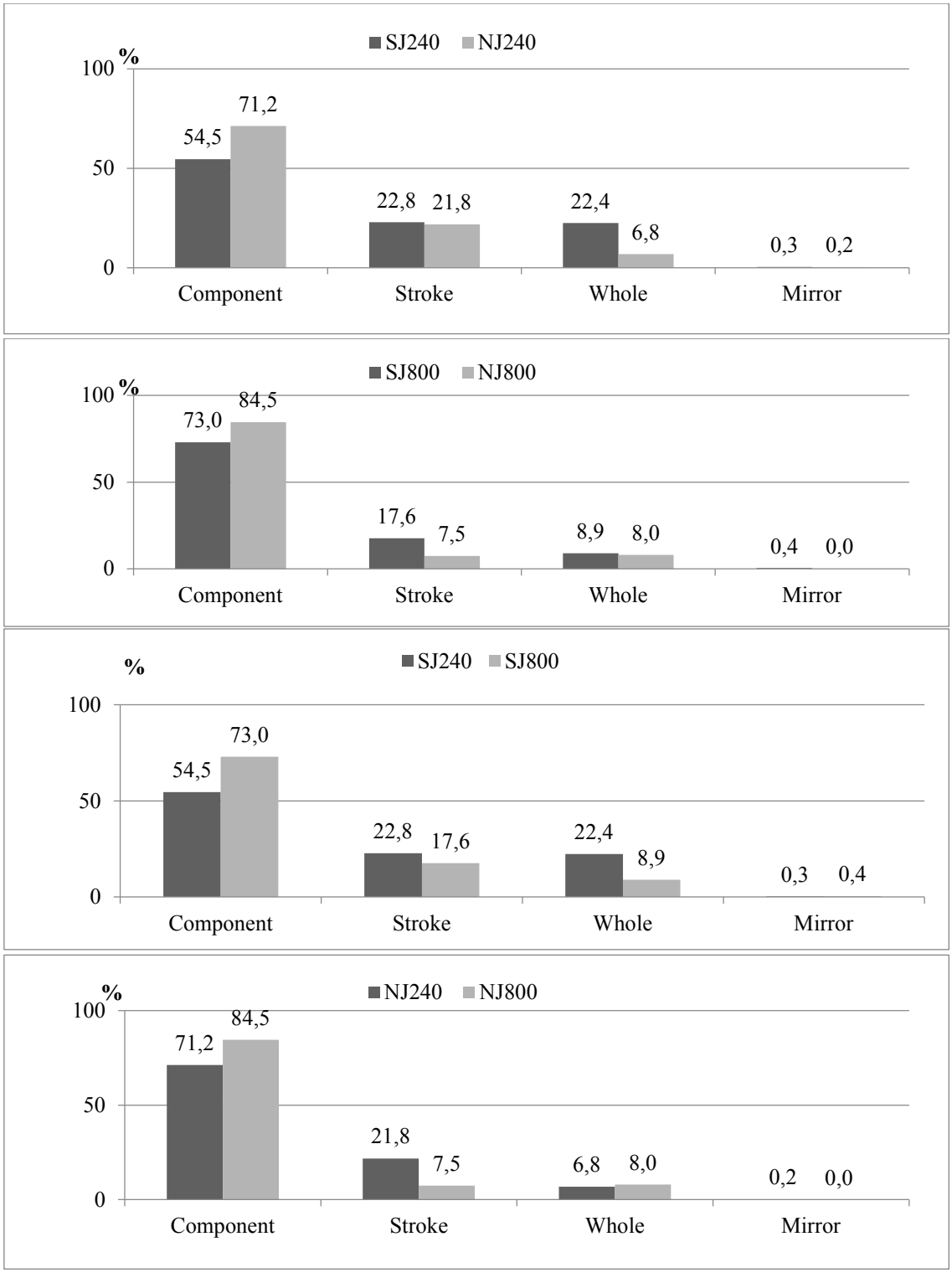

Figure 38 Collected figures of pseudokanji breakdown (Figures 27, 29, 32 & 33)

Hypothesis B (2):

Retrieval is phonologically inclined for the SJ groups and circumstantially inclined for the NJ groups.

This hypothesis is verified for the retrieval of the readings of kanji by the predominance of phonological reading errors for SJ240 and SJ800 and of circumstantial errors for NJ240 and NJ800 (see Table 50 and Figure 35 in subsection 7.1.2).

In terms of writing, LV240 groups show subtle signs of orthographic inclinations, with the orthographic type being rather low but second highest after the pseudokanji, as indicated by Table 51 and Figure 36 in subsection 7.1.2. Since SJ240's rate of orthographic errors is significantly lower than NJ240's, their orthographic inclination is only marginal. SJ800, whose phonological errors were the second most common after the pseudokanji type, are phonologically inclined. NJ800, with equally common phonological and orthographic errors, show both phonological and orthographic inclinations. Regarding the circumstantial type, writing errors of this type are quite infrequent in the results of all of the groups.

Hypothesis B (3):

All groups have difficulty in writing configurationally correct characters.

This hypothesis turned out to be quite true, since the pseudokanji type gained absolute predominance in all four groups' writing results, as indicated in Table 51 and Figure 36. Even for NJ800, whose pseudokanji rate was the lowest of the four, nearly half of the error type occurrences were of this type.

Hypothesis B (4):

The LV240 groups have difficulty in mastery of *kana* orthography.

As indicated in Table 53 and Figure 37, Hypothesis B (4) is verified by the LV240 groups' significantly higher rate of the transcription subtype, which is a sign of incomplete mastery of *kana* orthography in comparison with their respective LV800 counterparts.

7.1.5. Hypothesis C on the characteristics of the kanji learning process of SJ

Research Question C:

What are the characteristics of the kanji learning process of SJ?

Taking Hypotheses A and B into consideration, Swedish learners are expected to show following characteristics:

Hypothesis C:

SJ groups have components as the base unit of processing and phonological inclinations of retrieval. Their configurational awareness is less developed than that of level-matched NJ groups and shows a lesser degree of inter-level development than NJ groups.

In comparison with their respective level-matched NJ groups, the SJ groups are likely to be more phonologically inclined in retrieval and less dependent on circumstantial clues. They also have less developed configurational awareness.

In terms of the level differences, the SJ groups are likely to show similar general shifting patterns to those of the NJ groups. The reasons for the common phonological reading errors shift from poor skills in the *kana* orthography to an increased stock of homophonous characters in the mental lexicon, while the shift of the common writing error type from the pseudokanji to the phonological type is due to the development of configurational awareness and the increase in homophonous characters. However, SJ800 is expected to have different dominant writing error type combinations from NJ800 due to their lesser degree of development of configurational awareness: the likely combination is phonological-pseudokanji for SJ800, and phonological-orthographic for NJ800, since the pseudokanji type is a result of less developed awareness and the orthographic type a result of more developed awareness.

7.1.6. Verification of Hypothesis C

Based on the verifications in subsections 7.1.2 and 7.1.4 (the component as the base unit of processing for Hypothesis B (1), phonological inclinations towards retrieval for B (2), and configurational awareness for B (3)), Hypothesis C is deemed verified.

7.2. Other aspects of comparison

This subsection will discuss the aspects that have not been covered by the verification of Hypotheses A to C, namely, the breakdowns of transcription errors and of circumstantial errors.

7.2.1. Inter-level difference: semantic type reading errors: LV240 > LV800

As indicated in Table 50, the semantic type was significantly more common for LV240 groups than for their respective LV800 counterparts in reading. Since this error type was low in occurrence rate, it is unlikely that the LV240 groups are simply more semantically oriented than the LV800 groups. Instead, the grounds for this is likely to be either or both of the hypotheses below, because the rate was significantly higher for LV240 than LV800 regardless of the L1WS background, and also because of the use of different sets of the target characters for LV240 and LV800:

- LV240 target characters include more characters belonging to the same semantic category, such as the four seasons and the four directions, than LV800 target characters, which leads to a liability to semantic type errors; and/or

- Orthographic, phonological and semantic assemblies of such semantically related characters have not yet been well sorted within the mental lexicon of LV240 learners.

7.2.2. Transcription breakdown

There was a clear difference according to the participant groups' L1WS in their tendencies towards the three subtypes of phonological errors: transcription, alternative reading and component-based analogy (see Table 53 and Figure 37 in subsection 7.1.4). However, the transcription breakdowns did not show a similar result in terms of evidence of L1 transfer in the SJ groups.

Table 55 is a comparative summary of the transcription reading error breakdowns from subsections 6.1.3, 6.2.3 and 6.3.3. Figures 11, 16, 22 and 23 from these subsections are presented in sequence as Figure 39, so they will be readily comparable.

The error subtypes involving L1 transfer-inducing sounds were significantly more common for SJ240 than NJ240, which suggests that SJ240 are under the influence of their L1 phonology. On the other hand, SJ800 may be under less influence from their L1 phonology because there

was no significant difference in the L1 transfer-inducing sounds between the LV800 groups. In the inter-level comparison, there was no significant difference between SJ240 and SJ800. However, as indicated in Table 37, SJ800's error rates were quite close to the corresponding rates of the sound categories (51.8% error rate against 41.7% sound category rate), while the error rates for SJ240 deviated significantly from the corresponding rates, with a strong inclination toward L1 transfer inducing sounds (54.2% error rate against 25.4% sound category rate). This seems to indicate that SJ240, far more than SJ800, is under the phonological influence of the L1.

Table 55 Comparative summary of LV240/LV800 transcription error breakdowns

Error occurrence patterns	Group	Similarities	Significant differences
6.1.3. Transcription breakdown	SJ240 NJ240	• Subtype distribution pattern roughly reflects distribution pattern of sound formation of characters in question	• *sh/ch* confusion for SJ240 • L1 transfer inducing sound SJ240 > NJ240 • Non-L1 transfer inducing sound NJ240 > SJ240
6.2.3. Transcription breakdown	SJ800 NJ800	• Subtype distribution pattern mostly reflects distribution pattern of sound formation of characters in question • No significant difference for L1 transfer-inducing and non-inducing sounds	• Moraic nasal: SJ800 > NJ800 • *sh*/*ch* confusion: SJ800 > NJ800
6.3.3. Inter-level comparison of transcription breakdowns	SJ240 SJ800	Even distribution of L1 transfer-inducing and non-inducing sounds	SJ240 deviates from
	NJ240 NJ800	No particular similarity	• L1 transfer-inducing sound NJ800 > NJ240 • Non-L1 transfer-inducing sound NJ240 > NJ800

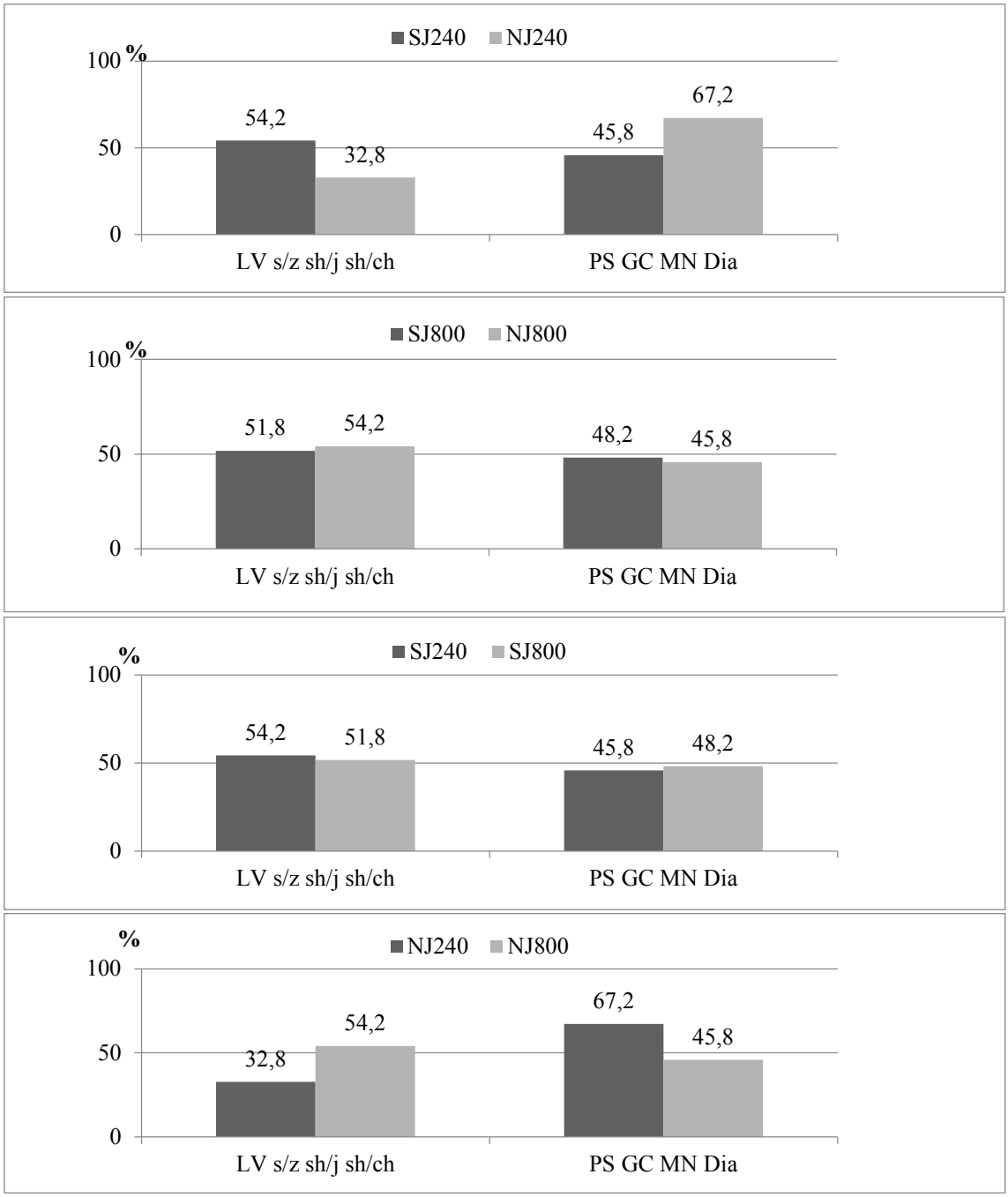

Figure 39 Collected figures of transcription breakdowns (Figures 11, 16, 22 & 23)

7.2.3. Circumstantial breakdown

Although there was a clear difference in tendencies in terms of the participant groups' L1WS in the occurrence rates of the circumstantial type of error, its breakdown showed no significant differences based on L1WS background. The inter-level comparisons showed differences in the inflection subtype for both SJ and NJ groups and in the context subtype for NJ groups. Since it was a common characteristic for the LV800 groups to have a significantly higher rate of the inflection error subtype, and the rate of readings that are presented with inflectional endings in the tests is approximately 25% of all target readings for each level, this may suggest that development of grammatical awareness requires both a higher knowledge level of the language (SJ800 in comparison with SJ240) and a higher maturity of the mental faculty (NJ800 in comparison with NJ240).

Table 56 is a comparative summary of the circumstantial reading error breakdowns from subsections 6.1.4, 6.2.4 and 6.3.4. Figures 12, 17, 24 and 25 from these subsections are combined in sequence as Figure 40, so they will be readily comparable.

Table 56 Comparative summary of LV240/LV800 circumstantial reading error breakdowns

Error occurrence patterns	Group	Similarities	Significant differences
6.1.4. Circumstantial breakdown	SJ240 NJ240	Context + Compound predominance over inflection	No significant difference
6.2.4. Circumstantial breakdown	SJ800 NJ800	Even distribution pattern for context, compound and inflection subtypes	No significant difference
6.3.4. Inter-level comparison of circumstantial breakdowns	SJ240 SJ800	Percentage of context (approximately 30% for each group)	Inflection SJ800 > SJ240
	NJ240 NJ800	Percentage of compound (approximately 35% for each group)	Context NJ240 > NJ800 Inflection NJ800 > NJ240

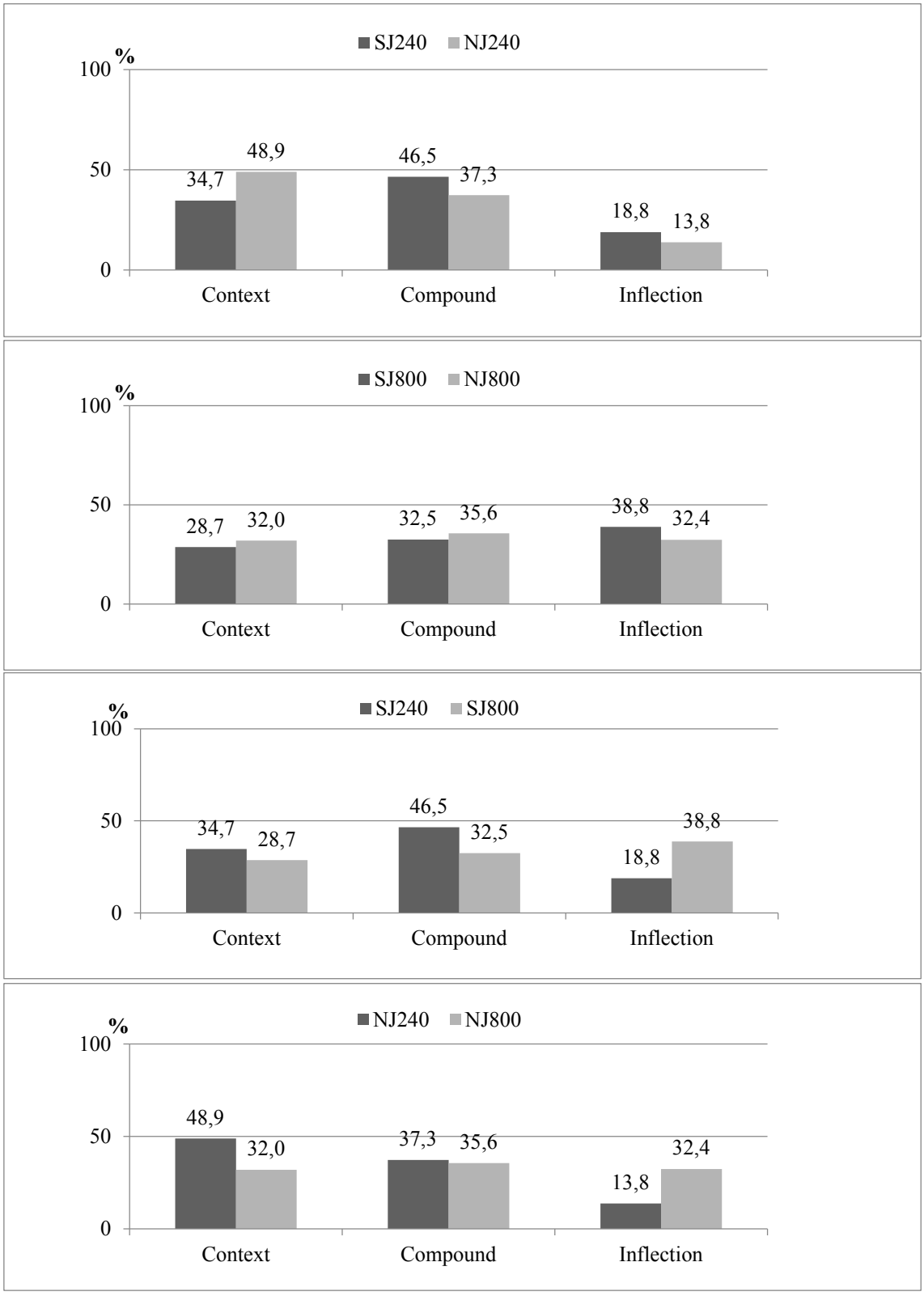

Figure 40 Collected figures of circumstantial breakdowns (Figures 12, 17, 24 & 25)

7.2.4. Reading vs. writing

As the first kanji research that has compared both reading and writing errors of L1 and L2 learners under the same conditions, this study presents a few interesting findings. Firstly, there was a significant difference in 14 out of the 16 in-group reading-writing comparisons of error type occurrences (see Table 48 in section 6.7). This indicates that different skills have different approaches towards kanji processing and retrieval and further suggests activation of different assemblies in the mental kanji lexicon in reading versus writing tasks.

As mentioned in section 6.7, the leanings are towards the phonological and circumstantial types in reading and the orthographic and phonological types in writing. This suggests that the retrieval is phonologically and/or circumstantially inclined in reading and orthographically and/or phonologically inclined in writing. The phonological inclination is understood as a sign of activation of the phonological assembly in the mental lexicon. It seems only natural that there is phonological activation involved in the retrieval of kanji sound, and that there is orthographic activation in the retrieval of kanji graphemes. However, why should phonological activation be involved in the writing task?

In the free recall writing experiment in Kano et al. (1989), few cases of phonological association were observed, presumably because the task of free recall writing is performed without phonological mediation. The writing task in this study, on the other hand, involved a substantial number of phonological clues (the pronunciations of the non-target parts of the material were provided in the form of *hiragana*). This phonological mediation must have caused phonological activation in the mental lexicon, which should explain the phonological inclination in writing.

With respect to the circumstantial type, the ratios are high in reading and low in writing. This suggests that retrieval of kanji in reading is circumstantially inclined but that this is not the case with retrieval of kanji orthography. In other words, circumstantial clues are more useful in reading than in writing.

7.3. Comprehensive discussion

In this section, the characteristics of the SJ learners' kanji learning process will be discussed comprehensively, according to skill and aspects and in comparison with that of NJ learners. Careful consideration will be given to possible causes of such characteristics and how they affect the transition from the novice to advanced level. Furthermore, pedagogical implications

will be discussed. Subsections 7.3.1 will deal with reading, 7.3.2 with writing and 7.3.3 with the universality of the characteristics.

7.3.1. Reading

(1) Inclinations of retrieval and unit of processing

In terms of their inclinations towards retrieval of kanji from the mental lexicon in reading tasks, the SJ groups are more phonologically inclined and the NJ groups are more circumstantially inclined. The NJ groups are consistently circumstance-reliant in decoding kanji in reading, presumably due to their linguistic competence as native speakers. SJ learners, however, do not show such a tendency even at the advanced level, at which they have developed a better command of Japanese. As advanced level learners with a sizable stock of multiple readings and kanji homophones, the NJ learners tend to increase the rate of their alternative reading errors from high to even higher, whereas the SJ learners, combining their enhanced phonological awareness with their now more developed configurational awareness, start to favour component-based analogy.

This can be explained by the difference in the basic processing unit of kanji reading. The SJ learners' unit of processing for reading kanji shifts from phoneme-/mora-based to component-based, whereas the NJ learners' is word- and character-based at both novice and advanced levels. This contrast can be compared to the characteristics of Korean and Chinese learners of English in Wang, Koda & Perfetti (2003), mentioned in subsection 3.1.2. L1 alphabetic Korean learners are used to taking the phonological route and use more sublexical strategies when reading English, whereas L1 logographic Chinese learners, who are more dependent on the lexical route in their L1WS, read English using a lexical strategy and visual recognition of the whole word rather than phonological decoding.

As illustrated in Figure 3 in subsection 3.1.1, Swedish is a phoneme-based orthography of medium transparency, whereas kanji is a coarse-grained, highly opaque orthography. Although both SJ and NJ learners had learned medium-grained and extremely transparent *kana* before learning kanji, the *kana* orthography does not seem to affect their reading strategy, probably because *kana* can be learned in a short time and has little influence on the SJ learners' already established L1WS-based sublexical strategy, while the NJ learners learn both *kana* and kanji as their L1WS. Just like the English learners of Chinese in Nelson et al. (2005), who were unable to phonemically decode Chinese characters and had to develop a L2WS lexical strategy, SJ learners had to learn to decode kanji on a component basis.

(2) Difficulties and pedagogical implications

In terms of difficulty in reading, the transcription error subtype is characteristic of LV240. Between the two LV240 groups, SJ240 are more prone to this error subtype than NJ240, presumably because of phonological L1 transfer, as the transcription breakdown results suggest. The inter-level comparison of SJ groups' error rates and the rates of the corresponding sound category indicated that SJ800 is under less influence of their L1 phonology than SJ240.

Component-based analogy errors are specific to LV800, at which stage the learners have stored enough kanji homophones in their mental lexicon and have acquired the ability to decompose a character into components. The rate of this subtype is notably higher for SJ800 than for NJ800. The reason why SJ800 learners tend to try and read kanji using this strategy can be explained in the same way that Cook (2004) explained why L2WS learners of English were quicker than L1WS English children in mastery of written past tense morphology (see subsection 3.1.2). It was possibly due to the difference in age and their own L1WS literacy skills, but most probably because SJ learners had received explicit instructions on this strategy, which was apparently absent from the NJ groups' learning process.

In fact, there are explicit instructions on how to try and read unfamiliar characters using this component-based analogy in one of the course books used by the SJ learners. It recommends "Whenever you come across an unfamiliar kanji, try to pinpoint its sound part. You might be able to figure out the character's reading if you know the reading of the other kanji with the same part" (Bann, 2009:[25]). In addition, students are reminded or encouraged to use this strategy from time to time in class by teachers.

The high rate of the component-based analogy subtype implies that they do use this strategy whenever they come across an unfamiliar kanji, but at the same time, it also demonstrates that many of their attempts have been unsuccessful, either due to failure in pinpointing the sound part, to incorrect memory of the reading of the shared component, or to being oblivious of the possibility of character-radical *on*-reading inconsistency. As described in section 2.7, 66% of the Joyo Kanji characters are PSCs (phono-semantic composites) and 57.6% of them have complete character-radical *on*-reading consistency. That means that only 38% of the 1,945 the Joyo Kanji characters are preconditioned for a successful component-based analogy. Furthermore, at a lower level of kanji learning, the rate of PSCs within the list of learned kanji is substantially lower than 57.6%. Therefore, there are only a few PSCs

within the target language to start with, and it is highly likely that they have not learned the *on*-readings that are required for this strategy.

However, since the strategy of component-based analogy is suitable for SJ learners' characteristics of phonological inclination of retrieval and component-based processing, it is advisable for course books and teachers to give instructions that will increase the strategy's success rate (e.g. presentation of a list of PSCs with different levels of character-radical *on*-reading consistency), rather than trying to encourage them to adopt another strategy such as alternative reading, which is more suitable for an orthographic inclination of retrieval and character-based processing.

(3) Greater SJ/NJ reading error differences for LV800 than LV240

There was a significant difference for two error types (phonological and circumstantial) between SJ240 and NJ240, whereas three error types (phonological, circumstantial and orthographic) showed a significant difference between SJ800 and NJ800. In addition, the *p*-values for the phonological and circumstantial types are lower for LV800, as stated in subsections 6.1.1 and 6.1.2 (LV240 phonological p = 6.9345E-09 vs. LV800 phonological p = 9.9920E-15; LV240 circumstantial p = 2.3055E-08 vs. LV800 circumstantial p = 1.2990E-12). This indicates that LV800 has a greater difference between phonological and circumstantial error types than LV240. It can therefore be inferred that the differences between SJ and NJ groups are greater at LV800 than at LV240.

This finding was in fact unanticipated. The LV800 groups were expected to be more similar to each other than the LV240 groups were, because SJ800 have an increased knowledge of kanji and better command of the Japanese language and therefore should be under less L1/L1WS influence than SJ240. On the contrary, the difference between the groups is greater at LV800. This means that SJ and NJ learners' error type occurrence patterns do not converge at a higher level. Instead, they keep their respective initial partiality in processing and retrieval, and develop differently from their level-matched counterparts.

The previous L2 kanji reading studies have shown that L2 learners behave differently from L1 users and that such differences are ascribable to transfer from the learners' L1WS. Nevertheless, the previous comparisons involved level differences between L1 users and L2 users/learners (e.g. Tamaoka, 1992; Komori, 2009). Komori (2009) compared advanced/intermediate L2 learners and advanced L1 users and found greater similarities between the advanced L2 learners and L1 users than between the intermediate L2 learners and L1 users, although the advanced L2 learners still behaved differently from the L1 users. This study compared strictly level-matched

L1 and L2 learner groups at both novice and advanced levels and revealed that L1 and L2 learners behave differently at both levels and that the difference is in fact greater at the advanced level than at the novice level.

7.3.2. Writing

(1) Inclinations of retrieval and unit of processing

As discussed in subsection 7.1.4, both SJ and NJ learners use components as the basic unit of processing kanji as a grapheme. In terms of relative differences, the SJ groups are more inclined towards stroke-based processing in comparison with NJ groups, and this is also true of LV240 groups in comparison with LV800 groups.

In terms of their inclinations towards kanji retrieval in writing tasks, SJ groups show marginal orthographic inclinations at the novice level. This orthographic inclination is influenced by the nature of the skill (the act of writing triggers orthographic activation) and the low stock of homophonous characters in the mental lexicon, which limits the possibility of making phonological errors. At the advanced level, at which they have acquired a good stock of homophonous characters, their L1WS-influenced partiality for the phonological route in writing manifests itself and shows a clear sign of phonological inclinations towards retrieval of kanji.

(2) Pedagogical implications of configurational awareness

With respect to configurational awareness, the experiments showed results that consistently indicated that SJ learners are less developed than their NJ counterparts. The SJ groups' rates of the pseudokanji type of error were significantly higher than those of the respective level-matched NJ groups. Although the LV800 groups' rates were significantly lower than those of the LV240 groups, SJ800's rate was nearly as high as NJ240's. In addition, the *p*-values for the pseudokanji type were higher in the inter-level comparison for the SJ groups, as stated in 6.6.2 (p = 4.4653E-08 in the SJ comparison vs. p = 0 in the NJ comparison). This indicates that the degree of development from SJ240 to SJ800 is less than that from NJ240 to NJ800.

The pseudokanji breakdown results further confirm this trait. Swedish groups have higher rates of the subtypes that indicate less developed configurational awareness than their respective level-matched counterparts (Component: NJ240 > SJ240; Whole: SJ240 > NJ240; and Stroke SJ800 > NJ800, as summarised in Table 54). This was possibly due to the difference in age, duration of study, number of class hours and their study environments (e.g. if they are surrounded by and make daily use of kanji or not).

Alternatively, as Chikamatsu (2005) pointed out, it may be due to their L1WS-influenced overconfidence about their kanji production ability (they feel they can write the character when the pronunciation is given, but fail to produce a correct form, being oblivious to the configurational details of the kanji). However, the most substantial cause is probably the difference in the amount of handwriting practice.

The effect of kinaesthetic facilitation in retrieval of the orthography of kanji has been reported in a number of neurolinguistic and psychological linguistic studies. This effect can be explained either by the union hypothesis (learning kanji by rote creates a union between the motor action of handwriting and kanji recognition) (e.g. Sasaki & Watanabe, 1983; Sasaki, 1984; and Sasaki & Watanabe, 1984) or the cognitive function hypothesis (the motor action itself has a cognitive effect that supports certain aspects of grapheme recognition) (Murakami, 1991; Sumiyoshi, 1996). Furthermore, in her study on the cerebral mechanisms relating to this effect, Matsuo (2004) hypothesised that the correspondence between grapheme and sound is mediated by the motor action.

Handwriting by rote is a traditional and still very common way of learning kanji in Japan. Daily handwriting practice of newly introduced kanji in a drill book is typical homework for Japanese schoolchildren, although they do receive instructions on recognising the intra-character structures of kanji and do decompositional exercises as well. There are opportunities for handwriting practice of the same type for SJ learners as well, but the frequency and duration is much more limited than those available to NJ learners. After learning approximately 500 characters by the end of the novice to early intermediate level, SJ learners' opportunities for handwriting decrease markedly, because then they tend to shift to computer word-processing for Japanese essay writing, and many of the kanji exams, including the Japanese Language Proficiency Test, have multiple choice questions.

However, it would be neither practical nor necessary to make SJ learners go through the same intensity of rote training as NJ learners. The workload and speed of the full time course for a degree in Japanese at Swedish universities are usually so demanding that the learners cannot be expected to spend much time on rote learning of kanji. It would be ideal if they could achieve the level of L1 adult users' handwriting accuracy, but most L2 learners in the twenty-first century are not in very much need of more than basic handwriting ability. Decompositional ability for efficient and accurate recognition, on the other hand, is required for both reading and computer word-processing.

Therefore, it would be useful for L2 learners to go through adequate handwriting practice of the novice level characters to establish the basic sense of kanji configurations. Kanji characters are structurally more dense and complicated than alphabetic characters, and therefore the minute details of orthographic differences in kanji can be easily overlooked (i.e. two orthographically similar characters can easily be confused) (Tollini, 1992). By reproducing each character with a pen, learners would be forced to pay attention to details and the motor action of handwriting would facilitate the process of character recognition.

At the same time as handwriting practice, it is also important to have visual training to recognise intracharacter components in order to develop decompositional ability as a basis for successful component-based analogy, which is the strategy the SJ learners favour. Special consideration should be given to the assembly patterns and character pairs that are known to be difficult for learners with an alphabetic L1WS to recognise or distinguish (Tollini, 1994).

7.3.3. Universality of the characteristics

Apart from the phonological L1 transfer in reading, there is no evidence that the abovementioned characteristics are specific to Swedish learners. It is most likely that these characteristics are shared by the majority of adult university level learners with an alphabetic background learning Japanese outside Japan. The grounds for this presumption are (i) most languages with the alphabetic writing system have the same writing system granularity and similar orthographic transparency (see Figure 3 in subsection 3.1.1); and (ii) such L2 learners' study environments and conditions are similar to those of the Swedish learners in this study.

7.4. Limitations and future courses of research

This study compared participant groups with different age ranges and educational backgrounds (schoolchildren vs. university students), since level matching has priority over all other conditions in the experiments. Although such conditions are normally matched in L1/L2 comparison studies in general, these unmatched conditions in this study do not constitute limitations per se, since L1/L2 comparison aims, after all, to compare learners who learn(ed) the target language as their first language in their childhood and those who learned it as a second language in their youth or as adults. On the other hand, matching the knowledge levels of the two groups inevitably unmatches the age ranges and educational backgrounds. It is a choice between level matching and matching age/educational background, and this study has merely chosen the former. Furthermore, the varying mental faculties of the L2 and L1 groups due to

the difference in age and educational background were taken into consideration in designing the task types and the method of data presentation, as stated in Chapter 4.

On the other hand, at the time of sorting the error types, there were a varying number of errors that did not fit any of the definitions of the error classifications and were categorised as "others". For instance, in an example presented in Table 21 in section 5.4.2, there was a writing error in which よる *yoru* ("night"), which should be written as 夜, was erroneously written as 前 まえ "before." This error can be inferred to be a confusion between the two characters due to a combination of minor similarities: they are both used in time-related vocabulary (夜 "night" and 午前 "a.m.") and share a radical assemblage pattern of "top, bottom-left and bottom-right." However, this error was categorised as "other" because it does not match any of the writing error criteria of this study (the semantic association is not on a character basis but on a vocabulary basis, and they only share an assemblage pattern and no radical is shared).

Although some of the previous studies mention the influence of frequency and/or familiarity of characters/words in their experiments (BAASC, 2007; Matsumoto-Sturt, 2004; Tamaoka, 1992), consideration of such features was not included in the analyses in this study, since a familiarity/frequency scale that is appropriate for all of the compared groups was practically impossible due to the different study conditions and environment between the groups (e.g. the character 店 *mise* ("shop") might have been introduced equally recently for both groups, but it is a new word learned in the textbook and used only in the classroom for SJ learners, whereas NJ learners have known the word for several years in advance and have seen it daily on shop signs, etc.).

As a future course of research, the following two issues may be both meaningful and of interest. First, the same experiments as those in this study can be conducted using L2 learners with other L1/L1WS backgrounds (such as English, Chinese or Korean) as participants and the results can be compared with the corresponding L1 results to explore possible differences depending on the L2 learners' L1/L1WS background. Another possible area of research is to investigate possible correlations between the kanji reading and writing error occurrence patterns and the individual participants' cognitive characteristics. The participants can be tested not only on kanji reading and writing, but also on cognitive abilities such as visual memory and phonological awareness. Subsequently, the kanji test results and the cognitive test results can be compared and correlations between the two scores can be explored. For example, visual memory and orthographic writing error type occurrence rate may be correlated in some way, as may phonological awareness and phonological

reading error type occurrence rate. If such correlations can be found, it may be possible to establish whether learners' characteristics in kanji learning are more closely connected to their cognitive characteristics or to their L1/L1WS background.

8. CONCLUSION

This chapter will provide a summary of this study in section 8.1 and present the conclusion in section 8.2.

8.1. Summary

The purpose of this study was to investigate the characteristics of Swedish learners of Japanese (SJ) concerning their kanji interlanguage in the developmental process of kanji learning, especially the processing unit, preferences of retrieval methods and difficulties in reading and writing kanji in comparison with native Japanese (NJ) learners. In order to explore the said characteristics, the following research questions were asked:

A. What are the kanji error type occurrence patterns and their similarities and differences according to skill (reading/writing), level (LV240/LV800) and L1 (SJ/NJ)?

B. What are the units of processing, retrieval methods and difficulties in reading and writing kanji for SJ and NJ in their respective developmental stages?

C. What are the characteristics of the kanji learning process for SJ?

These questions were examined through a set of experiments analysing kanji reading and writing errors. Participants' levels were set based on the number of kanji characters that they had learned so that the same material could be used. In order to explore the developmental process of kanji learning, errors were collected from two levels of Swedish learners (SJ240 and SJ800, who had learned approximately 240 and at least 800 characters, respectively), to be compared with their respective level-matched Japanese learner groups (NJ240 and NJ800). The Swedish groups consisted of novice and advanced learners who were studying Japanese at Swedish universities, and the Japanese groups of second and fifth grade primary school pupils.

Errors were collected from kanji reading and writing tests. The same data collection method was employed in all of the experiments, and identical sets of target characters and test materials were used for the level-matched groups. The target characters were selected from the characters commonly included in the lists of kanji recently learned by both of the level-matched groups. Written tests of reading and writing were prepared as parallel tasks, presenting the target characters/readings embedded in simple

phrases/sentences. The tests were taken and marked, and the errors collected were entered into a database. Reading errors were classified into phonological, circumstantial, orthographic and semantic error types, and writing errors into the same four types plus an additional pseudokanji type. A mixed-type error was counted as multiple error type occurrences in the statistical analysis.

The results were presented and analysed for each skill and compared between the level-matched Swedish and Japanese learner groups, namely, reading at the 240-character level (LV240), reading at the 800-character level (LV800), writing at LV240 and writing at LV800. In addition, inter-level differences were examined according to skill (LV240 vs. LV800 in reading and LV240 vs. LV800 in writing). Furthermore, reading and writing results within the same learner groups were compared as well in order to investigate the characteristics according to skill.

In the analyses, the overall error type occurrence patterns were examined first, and then breakdowns were made for the error types in which a significant difference was found between the compared groups. Phonological reading errors were divided into three subtypes, namely, transcription errors, alternative reading and component-based analogy. Transcription errors were further divided into seven different error-prone sounds. Circumstantial reading errors were divided into three subtypes, namely, context, compound and inflection, and the pseudokanji writing errors into four subtypes, namely, component, stroke, whole and mirror image.

Hypotheses for the research questions were formed based on the observations and implications of the previous studies, and they were verified based on the analysed results of the experiments. In the following, the findings related to each of the research questions are presented.

Findings regarding Research Question A
(General error type occurrence patterns, similarities and differences)

a. The general occurrence patterns of kanji reading and writing error types reflect the nature of each skill: the reading skill was linked to the phonological and circumstantial types and the writing skill to the orthographic type.

b. The predominant reading error types were phonological-circumstantial for SJ groups and circumstantial-phonological for NJ groups.

c. The predominant writing error types were pseudokanji for LV240 groups, pseudokanji-phonological for SJ800 and pseudokanji-orthographic-phonological for NJ800.

d. The L1WS-matched groups showed greater similarities in reading error occurrence patterns than the level-matched groups.

e. The level-matched groups showed greater similarities in writing error occurrence patterns than the L1WS-matched groups.

Findings regarding Research Question B
(Processing unit, retrieval and difficulties in reading and writing)

Reading

a. SJ240, who had high rates of transcription and alternative reading subtypes, have the mora/phoneme or character as the unit of kanji processing.

b. SJ800, with a high rate of component-based analogy errors, have a component-based unit of processing.

c. NJ240 and NJ800, both with high rates of alternative readings, process kanji on a whole character basis.

d. Retrieval is phonologically inclined for SJ groups and circumstantially inclined for NJ groups.

Writing

a. The component subtype of the pseudokanji type was predominant for all four groups, which indicates that they all have the component as a general basic unit of processing for kanji writing.

b. The SJ groups have leanings towards the stroke subtype in comparison with their level-matched NJ counterparts.

c. The LV240 groups were more stroke-based than their respective LV800 counterparts.

d. Both LV240 groups, who had the second highest but rather low rates of the orthographic type writing errors, showed signs of orthographic inclinations (SJ240's inclination was marginal, and NJ240's moderate).

e. SJ800, whose phonological errors were the second most common after the pseudokanji type, are phonologically inclined.

f. NJ800, with equally common phonological and orthographic errors, showed both phonological and orthographic inclinations.

Difficulties

a. The significantly higher rate of the transcription subtype of phonological reading errors indicates that LV240 groups have difficulty in mastering the *kana* orthography.

b. SJ240 has additional difficulty in mastering the *kana* orthography due to L1 phonological transfer.

c. The absolute predominance of the pseudokanji type in all four groups' writing results suggests their underdeveloped configurational awareness and indicates their difficulty in writing configurationally correct characters. The LV240 groups have more difficulty than the LV800 groups and the SJ groups have more difficulty than their level-matched NJ counterparts.

Findings regarding Research Question C
(Characteristics of the kanji learning process of Swedish learners of Japanese)

a. SJ groups have the component as the basic unit of processing and have phonological inclinations of retrieval.

b. Their configurational awareness is less developed than that of level-matched NJ groups and shows a lesser degree of inter-level development than NJ groups.

Findings not covered by the research hypotheses are as follows:

Inter-level differences of the semantic type

The semantic type was significantly more common for the LV240 groups than for their LV800 counterparts in reading, presumably because LV240 target characters include more characters belonging to the same semantic category and/or LV240 learners' have an underdeveloped mental network of such semantically related characters.

L1 phonological transfer

The results suggest that SJ240 learners are under the influence of their L1 phonology, and that the L1 influence on SJ800 learners is smaller than that on SJ240.

Circumstantial breakdowns

The clear L1WS-driven differences observed in the occurrence rates of the circumstantial error type were not present in its breakdown, except in the inter-level comparisons of the inflection subtype. The LV800 groups made significantly higher rates of errors of the inflection subtype, which suggests enhanced grammatical awareness at a higher level of knowledge or maturity.

Reading vs. writing

The substantial number of significant interskill differences suggests activation of different assemblies in the mental kanji lexicon in the tasks of reading and writing (phonological and circumstantial types in reading and orthographic and phonological types in writing).

Summary of characteristics

Finally, the characteristics of the SJ and NJ groups will be summarised according to aspect, skill and level of kanji learning in Table 57. The items in parentheses indicate minor characteristics.

Table 57 Summarised characteristics of SJ and NJ groups according to aspect, skill and level

Aspect	Skill	SJ240	SJ800	NJ240	NJ800
Unit of processing	Reading	Mora-/phoneme-/Character-based	Component-based	Character-/Word-based	Character-/Word-based
	Writing	Component-/Stroke-based	Component-/Stroke-based	Component-/Stroke-based	Component-based
Inclinations of retrieval	Reading	Phonological (circumstantial)	Phonological (circumstantial)	Circumstantial (phonological)	Circumstantial (phonological)
	Writing	(Orthographic)	Phonological	Orthographic	Orthographic-phonological
Difficulties	Reading	Transcription, L1 transfer	(Transcription)	Transcription	No distinctive difficulties
	Writing	Handwriting correct form	Handwriting correct form	Handwriting correct form	(Handwriting correct form)
Other traits	Quintessential error subtype	Transcription	Component-based analogy, Inflection	Alternative reading	Alternative reading, Inflection
	Inter-level comparison	Shifting (difference in preferred strategy between SJ240 and SJ800)		Consistent (NJ240's tendencies escalate into NJ800's preferred strategies)	
	Reading	L1WS-based differences > Inter-level differences			
	Writing	Inter-level differences > L1WS-based differences			

8.2. Conclusion

8.2.1. Swedish learners' developmental characteristics in reading

Although both groups of NJ learners show consistent characteristics (character-based processing, circumstantially inclined retrieval and preference for the alternative reading strategy), the SJ learner groups' characteristics shift from SJ240's mora-/phoneme-/character-based processing, phonological inclination of retrieval and tendency towards transcription errors to SJ800's component-based processing, phonological approach to retrieval and preference for component-based analogy.

Since NJ learners learn kanji as an additional L1WS after *kana*, they learn to process kanji efficiently on a character basis from the start, and they keep to it, only increasing their inclination towards circumstantial and alternative reading errors. On the other hand, SJ leaners' L1WS-influenced phoneme-based processing does not work well with kanji and they have to learn a new processing method. Nevertheless, their L1WS-based phonological inclination does not change, and accordingly, their

interlanguage has developed to show different error occurrence patterns from their NJ counterparts in certain aspects.

In terms of difficulty in reading, the transcription error subtype is characteristic of LV240, especially of SJ240, presumably because they are under stronger influence of their L1 phonology than SJ800. The component-based analogy error subtype is specific to LV800, particularly to SJ800, most probably because the SJ learners had received explicit instructions on this strategy, which was apparently not the case with the NJ learners.

The high rate of the component-based analogy errors implies not only the SJ learners' preference for this strategy, but also the strategy's low success rate. Its failure stems from the learners' inability to properly decompose the character and recall the correct reading to apply; the limiting preconditions for the success of this strategy are another obstructive factor.

Since this strategy is suitable for SJ learners' characteristics of phonological inclination of retrieval and component-based processing, it is advisable to provide them with more instructions to raise the success rate of the strategy rather than encouraging them to try another strategy.

An unexpected finding was that the differences in reading error type occurrences between SJ and NJ were greater for LV800 than LV240. Unlike in previous studies, which merely reported the advanced L2 learners' major similarities with and minor differences from L1 users, this study revealed that L1/L2 differences are present at both levels and that the differences are greater at the advanced level than the novice level.

8.2.2. Swedish learners' developmental characteristics in writing

Both SJ and NJ learners process kanji characters on a component basis, although the SJ groups are more inclined towards stroke-based processing in comparison with NJ groups, and so are the LV240 groups in comparison with the LV800 groups. The SJ240 learners, who have a limited stock of homophonous characters, show marginal orthographic inclinations in retrieval of kanji, whereas the SJ800 learners are clearly phonologically inclined, as a combined consequence of an increased stock of homophonous characters and the L1WS-influenced preference for the phonological route.

With respect to configurational awareness, SJ learners are less developed than their NJ counterparts at both levels. This is possibly due to their difference in age, duration of study, number of class hours and their study

environments, or their L1WS-influenced overconfidence about their kanji production ability, but most probably because of the difference in the amount of handwriting practice.

Handwriting by rote has been and is still a very common way of learning kanji in Japan, and the effect of kinaesthetic facilitation in the retrieval of kanji has been reported in a number of studies. Nevertheless, the frequency and duration of SJ learners' kanji handwriting practice is usually much more limited due to the differences in educational setting and practical needs. However, adequate handwriting practice of the basic characters would help them to pay attention to the structural details of kanji, and kinaesthetic memory would facilitate the character recognition process. Character and component recognition training would also be effective, especially with consideration for L1WS alphabetic learners' difficulties.

8.2.3. The significance of this study

This study conducted large-scale rigorous experiments that were unprecedented in the field of L2 kanji research. The experiments had a substantial number of participants (a total of 395), were meticulously designed and performed under strictly controlled conditions, and comparisons were made between two levels of L2 learners and their respective level-matched L1 learner groups for both reading and writing. The reading and writing tests were parallel tasks with identical sets of target characters in a practical task setting, unlike the previous studies that tested relatively small numbers of learners/users on recognition (reading) or production (writing) of single characters or words independent of context under less well-controlled conditions.

Through comprehensive analysis of the error occurrence patterns, this study made a number of new findings and at the same time verified and confirmed various suggestions and observations made in the previous studies.

The findings that are unique to this study are as follows:

(1) Each learner group had a quintessential error subtype of reading errors, namely, the transcription error (L2 novice), the component-based analogy (L2 advanced) and the alternative reading (L1 novice and L1 advanced);

(2) L1 phonological transfer had a significant influence on Swedish learners at the novice level and decreased substantially at the advanced level;

(3) L2 learners' configurational awareness is less developed than that of level-matched L1 learners and shows a lesser degree of inter-level development than L1 learners;

(4) Significantly higher rates of the semantic reading error type at the novice level than the advanced level were found for both L1 and L2 learners;

(5) Significantly higher rates of the inflection subtype of the circumstantial reading error were found at the advanced level than at the novice level for both L1 and L2 learners; and

(6) Shifting inter-level characteristics were observed for L2 learners, as opposed to consistent inter-level characteristics for L1 learners.

As mentioned above, this study has also confirmed, inter alia, the following characteristics of L1WS alphabetic learners of kanji: phonological inclinations of retrieval, component-based units of processing, limited configurational awareness, and a predominance of pseudokanji type writing errors. Also confirmed was the more generic hypothesis of "general similarities and minor differences in error type occurrence tendencies between the level-matched groups".

Another hypothesis verified by this study is that the error occurrence tendencies are L1WS-driven in reading and level-driven in writing (e.g. Komori, 2009; Hatta et al., 2002), as had been observed separately in single-skill tests in previous studies. This study confirmed the said aspect more precisely, i.e. the differences due to L1WS are greater than the inter-level differences in reading and the inter-level differences are greater than the differences due to L1WS in writing.

Based on these findings and observations, this study has gained a precise and comprehensive understanding of the developmental characteristics of Swedish learners' kanji learning process, and, at the same time, has elucidated the developmental characteristics of L2 learners' kanji learning process in comparison with those of L1 learners. The challenges experienced by L2 learners with an alphabetic L1WS stem from the shift from a phoneme-based to component-based processing of the graphemes, taking a less familiar lexical route in the decoding and encoding of grapheme-sound correspondence, and the use of less efficient strategies in reading and writing. It is hoped that the findings and implications of this study will contribute to facilitation of L2 kanji learning in the future.

REFERENCES

Akamatsu, Nobuhiko. "Effects of Second Language Reading Proficiency and First Language Orthography on Second Language Word Recognition." In *Second Language Writing Systems*, edited by Vivian Cook and Benedetta Bassetti, 238-59. Clevedon: Multilingual Matters Ltd., 2005.

Banno, Eri, Yoko Ikeda, Kyoko Shinagawa, Kaori Tajima, and Kyoko Tokashiki. *Kanji Look and Learn: 512 Kanji with Illustrations and Mnemonic Hints*. Tokyo: The Japan Times, 2009.

Banno, Eri, Yutaka Ohno, Yoko Sakane, and Kyoko Shinagawa. *Genki : An Integrated Course in Elementary Japanese I*. 2 vols. Vol. 1. Tokyo: The Japan Times, 1999.

Banno, Eri, Yutaka Ohno, Yoko Sakane, Kyoko Shinagawa, and Kyoko Tokashiki. *Genki: An Integrated Course in Elementary Japanese Ii*. 2 vols. Vol. 2. Tokyo: The Japan Times, 1999.

Bebout, Linda. "An Error Analysis of Misspellings Made by Learners of English as a First and as a Second Language." *Journal of Psycholinguistic Research* 14 (1985): 569-93.

Bernhardt, Elizabeth Buchter, and Michael E. Everson. "Second Language Reading: A Cognitive Perspective." In *The National Reading Conference*. Tucson, AZ, 1988.

Chan, Lily, and Terezinha Nunes. "Children's Understanding of the Formal and Functional Characteristics of Written Chinese." *Applied Psycholinguistics* 19, no. 01 (1998): 115-31.

Chikamatsu, Nobuko. "The Effects of L1 Orthography on L2 Word Recognition: A Study of American and Chinese Learners of Japanese " *Studies in Second Language Acquisition* 18 (1996): 403-32.

Chikamatsu, Nobuko. "L2 Japanese Kanji Memory and Retrieval: An Experiment on the Tip-of-the-Pen (Top) Phenomenon." In *Second Language Writing Systems*, edited by Vivian Cook and Benedetta Bassetti, 71-96. Clevedon: Multilingual Matters, 2005.

Chikamatsu, Nobuko, Shoichi Yokoyama, Hironari Nozaki, Eric Long, and Sachio Fukuda. "A Japanese Logographic Character Frequency List for Cognitive Science Research." *Journal of Behavior Research Methods, Instruments and Computers* 32 (2000): 482-500.

Cook, Vivian J. "L2 Users and English Spelling." *Journal of Multilingual and Multicultural Development* 18, no. 6 (1997): 474-88.

Cook, Vivian J. "The Spelling of the Regular Past Tense in English: Implications for Lexical Spelling and Dual Process Models " In *An International Master of Syntax and Semantics. Papers Presented to Aimo Seppänen on the Occasion of His 75th Birthday.*, edited by Gunnar Bergh, Jennifer Herriman and Mats Mobärg. Göteborg: University of Gothenburg, 2004.

Cook, Vivian J., and Benedetta Bassetti. "An Introduction to Researching Second Language Writing Systems." In *Second Language Writing Systems*, edited by Vivian Cook and Benedetta Bassetti, 1-67. Clevedon: Multilanguage Matters, Ltd., 2005.

Corballis, Michael C. *Human Laterality*. New York: Academic Press, 1983.

Corder, Stephen Pit. "Error Analysis." In *Techniques in Applied Linguistics*, edited by J. Allen and S.P. Corder, 122-54. London: Oxford University Press, 1974.

Cornell, James M. "Spontaneous Mirror-Writing in Children." *Canadian Jounrnal of Psychology* 39, no. 1 (1985): 174-79.

Editors, Alc Japanese Language Publishing. *Kanji Power Handbook for the Japanese Language Proficiency Test*. Tokyo: Alc Co., Ltd., 1994.

Elert, Claes-Christian. *Phonologic Studies of Quantity in Swedish Based on Material from Stockholm Speakers*. Uppsala: Almqvist & Wiksell, 1964.

Ellis, Nick, Miwa Natsume, Katerina Stavropoulou, Lorenc Hoxhallari, Victor H. P. Van Daal, Nicoletta Polyzoe, Maria-Louisa Tsipa, and Michalis Petalas. "The Effects of Orthographic Depth on Learning to Read " *Reading Research Quartery* 39, no. 4 (2004): 438-68.

Flaherty, Mary. "Do Second-Language Learners of Japanese Process Kanji in the Same Way as Japanese Children? ." *Japanese Language Education Around the Globe* 1 (1991): 183-200.

Flaherty, Mary *The Role of Abstract Visual Memory in the Learning of Kanji Reading* Edited by Hsuan-Chih Chen, Cognitive Processing of Chinese and Related Asian Languages. Hong Kong: The Chinese University Press, 1997.

Flores d'Arcais, Giovanni, and Hirofumi Saito. "Lexical Decomposition of Complex Kanji Characters in Japanese Readers. ." *Psichological Research* 55 (1993): 52-63.

Flores d'Arcais, Giovanni, Hirofumi Saito, and Masahiro Kawakami. "Phonological and Semantic Activation in Reading Kanji Characters." *Journal of Experimental Psychology: Learning, Memory and Cognition* 21 (1995): 3-42.

Gamage, Haththotuwa. "Perceptions of Kanji Learning Strategies: Do They Differ among Chinese Character and Alphabetic Background Learners? ." *Australian Review of Applied Linguistics* 26 (2003): 17-30.

Goswami, Usha. "The Relationship between Phonological Awareness and Orthographic Representation in Different Orthographies." In *Learning to Read and Write: A Cross Linguistic Perspective*, edited by M. Harris and G. Hatano, 134-56. Cambridge: Cambridge University Press, 1999.

Han, Mieko. "The Feature of Duration in Japanese." *Journal of Phonetic Society of Japan* 10 (1965): 65-80.

Hanley, Richard, Ovid Tzeng, and Tzung-Hsien Huang. "Learning to Read Chinese." In *Learning to Read and Write: A Cross-Linguistic Perspective*, edited by M. Harris and Hatano. G., 173-95. Cambridge: Cambridge University Press, 1999.

Hatta, Takeshi, Ayako Kawakami, and Katsuo Tamaoka. "Errors in Writing Japanese Kanji: A Comparison of Japanese Schoolchildren, College Students and Second-Language Learners of Japanese." *Asia Pacific Journal of Speech, Language and Hearing (Australia)* 7, no. 3 (2002): 157-66.

Hatta, Takeshi, Ayako Kawakami, and Katsuo Tamaoka. "Writing Errors in Japanese Kanji: A Study with Japanese Students and Foreign Learners of Japanese." *Reading and Writing* 10 (1998): 457-70.

Huang, Tzung-Hsien, and Richard Hanley. "Phonological Awareness and Visual Skills in Learning to Read Chinese and English." *Cognition* 54, no. 1 (1995): 73-98.

Hue, Chih-Wei. "Recognition Process in Chinese Naming." In *Language Processing in Chinese*, edited by Hsuan-Chih Chen and O.J.L. Tzeng, 93-107. Amsterdam: North-Holland, 1992.

Inoue, Miyoko. "Perception of Japanese Quantity by Swedish Speaking Learners: A Preliminary Analysis." Paper presented at the Fonetik 2009, Stockholm, 2009.

Inoue, Miyoko. "Production of Japanese Quantity by Swedish Native Speakers: With a Focus on Cvcv Type Words." In *International Symposium on Japan as Seen from Different Cultural Perspective*, 37. Nagoya: The Graduate School of Languages and Cultures 2008.

Ivarsson, Fusae. "Analyses of Kanji Reading and Writing Errors of Swedish Learners of Japanese in Comparison with Proficiency-Matched Japanese Schoolchildren." In *Translation, Education, and Innovation in Japanese and Korean Societies*, edited by David G. Hebert. London: Springer, 2016 forthcoming.

Ivarsson, Fusae. "Kenkyuu-Nooto: Nihonjin-Gakudoo Tono Gotootaishoo-Bunseki O Chuushin Ni Shita Sweedenjin-Nihongogakushuusha No Kanjishuutoku-Katee No Ninchiteki Kenkyuu (Research Notes: A Cognitive Study of the Kanji Learning Process of Swedish Students of Japanese Focusing on Comparative Error Analyses with Japanese School Children)." *Ryugakusei Kyoiku, Bulletin of International Student Center, The University of the Ryukyus* 8 (2011): 93-101.

Iwasaki, Shoichi. *Japanese*. Revised edition ed. Vol. 17, London Oriental and African Language Library. Amsterdam: John Benjamins, 2013.

Japan, Basic Academic Ability Survey Committee of Teaching Skill Sharing Society of. *Dēta Kara Miru Kanji Shūtoku Itsutsu No Pointo: Shōgakkō Gakushū-Kanji Shūtoku-Jōkyō No Chōsa-Hōkoku (the Five Points of Kanji Acquisition Based on Data: Survey Report on Primary School Kanji Acquisition Situation)*. Tokyo: Mitsumura Educational, 2007.

Kadota, Shuhei. "Shikaku-Tēji Sareta Ētango Pea No Kankē-Handan (Relationship Judgment of Visually Presented English Word Pairs) " *Gaikokugo gaikokubunka kenkyū* 11 (1998): 205-20.

Kadota, Shuhei. *Sla Kenkyū Nyūmon: Dai Ni Gengo No Shori & Shūtoku Kenkyū No Susumekata (Introduction to Sla Research: How to Conduct Research on Second Language Processing and Acquisition)*. Tokyo: Kuroshio Shuppan, 2010.

Kano, Chieko, Yuri Shimizu, Hiroko Takenaka, Eriko Ishii, and Satoru Akutsu. *Intermediate Kanji Book*. Vol. 1. Tokyo: Bonjinsha, 1993.

Kano, Chieko, Yuri Shimizu, Hiroko Takenaka, Satoru Akutsu, Eriko Ishii, Hiroyuki Kaiho, and Takeshi Deguchi. "Jiyū-Hōshutsu-Hō Ni Yoru Gaikokujin No Kanji-Chishiki No Bunseki (Analysis of L2 Learners' Kanji Knowledge by Free Recall Method)." *Tsukuba Daigaku Ryūgakusē Kyōiku Sentā Nihongo Kyōiku-ronshū* 4 (1989): 65-91.

Keightley, David Noel. *Sources of Shang History: The Oracle-Bone Inscriptions of Bronze Age China*. Berkeley, Los Angeles: University of California Press, 1978.

Kern, Martin. "Early Chinese Literature, Beginnings through Western Han." In *The Cambridge History of Chinese Literature, Volume 1: To 1375*, edited by Stephen Owen, pp. 1–115. Cambridge: Cambridge University Press, 2010.

Kess, Joseph F., and Tadao Miyamoto. *The Japanese Mental Lexicon*. Amsterdam: John Benjamins, 1999.

Koike, Toshihide, Miyoshi Kumoi, and Tsutomu Kuboshima. *Ld-Ji No Tame No Hiragana Kanji Shien: Kobetsu Shien Ni Ikasu Shoji Kyōzai (Supporting Children with Learning Disorder in Their Hiragana and Kanji Learning: Teaching Materials in Writing Useful for Individual Support)*. Kyoto: Airi Shuppan, 2003.

Komori, Saeko. *A Study of Kanji Word Recognition Process for Japanese as a Second Language*. Tokyo: Kazama Shobo, 2009.

Kono, Toshihiro. *Kodomo No Shoji to Hattatsu: Kensa to Shien No Tame No Kisobunseki (Childrens' Writing and Devrlopment: Basic Analysis for Screening and Support)*. Tokyo: Fukumura Shuppan, 2008.

Leong, Che Kan, and Katsuo Tamaoka. "Use of Phonological Information in Processing Kanji and Katakana by Skilled and Less Skilled Japanese Readers." *Reading and Writing: An Interdisciplinary Journal* 7 (1995): 377-93.

Matsumoto, Kazumi "Kanji Recognition by Second Language Learners: Exploring Effects of First Language Writing Systems and Second Language Exposure." *Modern Language Journal* 97, no. 1 (2013): 161-77.

Matsumoto-Sturt, Yoko. "Measuring Accuracy and Speed of L2 Kanji Word Identification." Paper presented at the Vocabulary knowledge and use: measurements and applications, Bristol, 2004.

Matsuo, Kayako. *Kūsho No Nōnai Mekanizumu (the Celebral Mechanisms of Finger Writing Gesture)*. Tokyo: Kazama Shobō, 2004.

Ministry of Education, Culture, Sports, Science and Technology. "The Enforcement Regulations of the School Education Law (Gakkō Kyōiku Hō Shikō Kisoku)." edited by Culture Ministry of Education, Sports, Science and Technology, 1998.

Ministry of Education, Culture, Sports, Science and Technology. "Gakunenbetsu Kanji Haitōhyō (List of Kanji by School Year)." edited by Culture Ministry of Education, Sports, Science and Technology, 1989.

Mori, Yoshiko. "Effects of First Language and Phonological Accessibility on Kanji Recognition." *The Modern Language Journal* 82, no. 1 (1998): 69-82.

Mori, Yoshiko. "Epistemological Beliefs and Language Learning Beliefs: What Do Language Learners Believe About Learning?" *Language Learning* 49 (1999): 377–415.

Mori, Yoshiko, and Junko Mori. *Language Teachin* 44, no. 04 (2011): 447-84.

Mori, Yoshiko, Kumi Sato, and Hideko Shimizu. "Japanese Language Students' Perceptions on Kanji Learning and Their Relationship to Novel Kanji Word Learning Ability." *Language Learning* (2007).

Mori, Yoshiko, and Hideko Shimizu. "Japanese Language Students' Attitudes toward Kanji and Their Perceptions on Kanji Learning Strategies." *Foreign Language Annals* 40 (2007): 472–90.

Murakami, Shiho. "Kūsho Ni Miru Hyōshō No Keisei (Formation of Representations Observed in the Finger Writing Gesture)." *Hattatsu shinrigaku kenkyū* 2 (1991): 25-31.

Muter, Valerie, and Margaret Snowling. "Grammar and Phonology Predict Spelling in Middle Childhood." *Reading and Writing: An Interdisciplinary Journal* 9 (1997): 407-25.

Nesbitt, Dallas. "Achieving Unconscious Recall of Kanji: Can Rote Learning Help?" *New Zealand Studies in Applied Linguistics* 15, no. 2 (2009): 61-73.

Nomura, Masaaki. "Kanji No Tokusē O Hakaru: Kanji No Kēryō-Kokugogaku (Measuring the Characteristics of Kanji: The Mathematical Linguistics of Kanji)." In *Kanji O Kagak Suru (the*

Scientific Exploration of Kanji), edited by Hiroyuki Kaiho, 1-34. Tokyo: Yuhikaku, 1984.

Okada, Takeshi. "Remarks on Japanese Poor Spellers of English." *Yamagata English Studies* 7 (2002): 21-41.

Okita, Yoko. "Kanji No Kakiayamari Ga Kanjikyōiku Ni Shisasuru Koto (What Do Handwriting Kanji Errors Tell Us?)" (2001): : 19-28." In *Nihongo-Kyōiku No Tame No Ajia-Shogengo No Taiyak-Sakubun-Dēta No Shūshū to Kōpasu No Kōchiku*, edited by Kokuritsu Kokugo Kenkyūjo, 19-28. Tokyo: Kokuritsu Kokugo Kenkyūjo, 2001.

Okita, Yoko. "Shokyū Kyōkasho-No Kanji Gakushū Sutorategī Shiyō Oyobi Kanji Gakushū Shinnen Ni Ataeru Eikyō [the Influence of Introductory Textbooks on Kanji Learning Strategy Use and Beliefs About Kanji Learning]. ." *Sekaii no Nihongo Kyoiku* 8 (1998): 31–45.

Okita, Yoko. "Students' Beliefs About Learning Japanese Orthography: Beyond the Textbooks." In *New Trends and Issued in Teaching Japanese Language and Culture*, edited by H. M. Cook, K. Hijirida and M. Tahara, 61–75. Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center, 1997.

Oller Jr., John W., and Seid M. Ziahosseiny. "The Contrastive Analysis Hypothesis and Spelling Errors." *Language Learning* 20 (1970): 183-89.

Ozeki, Hiromi. *Nihongo O Oshieru Tame No Dai-Ni Gengo Shūtoku Nyūmon (Introduction to Second Language Acquisition for Teaching Japanese)*. Tokyo: Kuroshio Shuppan, 2010.

Perfetti, Charles A., and Susan Dunlap. "Learning to Read: General Principles and Writing System Variations." In *Learning to Read across Languages*, edited by Keiko Koda and Annette M. Zehler, 13-38. New York: Routledge, 2008.

Saito, Hirofumi. "Kanji No Ninshiki to Hattatsu (Recognition and Development of Kanji)." In *Kanji No Hataraki (the Functions of Kanji)*, edited by Tomiyoshi Maeda and Masaaki Nomura. Tokyo: Asakura shoten, 2006.

Saito, Hirofumi, Michio Inoue, and Yukimasa Nomura. "Information Processing of Kanji (Chinese Characters) and Kana (Japanese Characters): The Close Relationship among Graphemic, Phonemic, and Semantic Aspects. ." *Psychologia* 22 (1979): 195-206.

Saito, Hirofumi, Masahiro Kawakami, Yoshinobu Yanase, and Hisashi Masuda. "Evaluating the Wordlikeness of Kanji Pseudo-Compounds: Holistic and Analytic Processes in Reading Kanji Compounds." *Acta Psychologica Sinica* 32 (2000): 47-55.

Saito, Hirofumi, Hisashi Masuda, and Masahiro Kawakami. "Form and Sound Similarity Effects in Kanji Recognition." *Reading and Writing: An Interdisciplinary Journal* 10 (1998): 323-57.

Saito, Hirofumi, Hisashi Masuda, and Masahiro Kawakami. "Subword Activation in Reading Japanese Kanji Character Words." *Brain and Language* 68 (1999): 75-81.

Sakuma, Naoko, Sumiko Sasanuma, Itaru F. Tatsumi, and Shinobu Masaki. "Orthography and Phonology in Readng Japanese Kanji Words: Evidence from the Semantic Decision Task with Homophones." *Memory & Cognition* 26, no. 1 (1998): 75-87.

Sasaki, Masato, and Akira Watanabe. "Kūsho Kōdō No Bunkateki Kigen: Kanjiken, Hikanjiken Tono Hikaku (the Cultural Origin of the Finger Writing Gesture: A Comparison between the Kanji- and Non-Kanji-Spheres)." *Kyōiku shinrigaku kenkyū* 32 (1984): 182-90.

Sasaki, Masato, and Akira Watanabe. "Kūsho Kōdō No Shutsugen to Kinō - Hyōshō No Undōkankakuteki Na Seibun Ni Tsuite (the Appearances and Functions of the Finger Writing Gesture: On the Motorsensory Aspects of Representations)." *Kyōiku shinrigaku kenkyū* 31 (1983): 1-10.

Schott, Geoffrey. "Mirror Writing: Neurological Reflections on an Unusual Phenomenon." *Journal of Neurology, Neurosurgery and Psychiatry* 78, no. 1 (2007): 5-13.

Sears, Chris R., Yasushi Hino, and Stephen J. Lupker. "Neighborhood Size and Neighborhood Frequency Effects in Word Recognition." *Journal of Experimental Psychology: Human Perception and Performance* 21 (1995): 876-900.

Seidenberg, Mark S. "The Time Course of Phonological Code Activation in Two Writing Systems." *Cognition* 19 (1985): 1-30.

Shirahata, Tomohiko, Shigenori Wakabayashi, and Muranoi Hitoshi. *Shōsetsu Dai-Ni Gengo Shūtoku Kenkyū: Riron Kara Kenkyūhō Made (Detailed Explanations on Second Language Acquisition Studies: From Theory to Research)*. Tokyo: Kenkyusha, 2010.

Shu, Hua, and Richard C. Anderson. "Learning to Read Chinese: The Development of Metalinguistic Awareness." In *Reading Chinese Scripts: A Cognitive Analysis*, edited by Jian Wang, Hsuan-Chih Chen, Ralph Radach and Albrecht Inhuff, 1-18. Mahwah, NJ: Lawrence Erlbaum, 1999.

Su, Xiaoxiang, Young-Suk Kim, Christopher Reenock, Alysia Roehrig, Gretchen Sunderman, and Shelbie Witte. "Radical Awareness among Chinese-as-a-Foreign-Language Learners." Florida State University, 2010.

Sumiyoshi, Chika. "Eitango Sōki Ni Mirareru Kūsho Kōdō - Nihongo-Washa to Chūgokugo-Washa Ni Okeru Kentō (Finger Writing Gesture Observed in Recalling English Words: A Study with Japanese and Chinese Speakers)." *Kyōiku shinrigaku kenkyū* 44 (1996): 75-84.

Tainaka, Hidetsugu. "Hissha-Sagyō-Jikan Ni Tsuite: Kanji-Kana-Majiri to Hiragana No Hikaku (Regarding Time for Handwriting Tasks: Comparison between Kanji-Kana-Mixture and Hiragana)." *Osaka Prefectural Research Institute of Public Health Research Report: working health* 17 (1979): 39-44.

Tamaoka, Katsuo. "Ēgo O Bogo to Suru Nihongo Gakushūsha No Tango Shori No Kōritsu (Word Processing Effectiveness of Japanese Learners with English L1 Backgrounds)." *Intercultural/Transcultural Education* 6 (1992): 99-113.

Tamaoka, Katsuo. "The Processing Strategy of Words Presented in Kanji and Kana by Chinese and English Speakers Learning Japanese." *Studies in Language and Literature* 17 (1997): 65-77.

Tamaoka, Katsuo, Kim Kirsner, Yushi Yanase, Yayoi Miyaoka, and Masahiro Kawakami. "A Web-Accessible Database of Characteristics of the 1,945 Basic Japanese Kanji." *Behavior Research Methods, Instruments and Computers* 34 (2002): 260-75.

Tamaoka, Katsuo, and Hiroyuki Yamada. "The Effects of Stroke Order and Radicals on the Knowledge of Japanese Kanji Orthography, Phonology and Semantics." *Psychologia* 43 (2000): 199-210.

Tanaka, Toshitaka. "Hōkō-Ninchi Ni Kansuru Hattatsu: Zukei to Moji Ni Tsuite (Development of Directional Cognition: On Figures and Letters)." *Kyōiku shinrigaku kenkyū* (1978).

Toda, Takako. "Gaikokujin-Gakushūsha No Nihongo Tokushuhaku No Shūtoku (Acquisition of Special Morae in Japanese as a Second Language) " *Journal of the Phonetic Society of Japan* 7, no. 2 (2003): 70-83.

Tollini, Aldo. "Hikanjikei Gakushūsha No Tame No Nyūmonki Ni Okeru Kanji-Gakushū-Shidō No Ichi-Kōsatsu (a Consideration on the Teaching of Kanji at Introductory Level to Non-Kanji Area Students)." *Sekai no nihongo-kyōiku* 2 (1992): 65-76.

Tollini, Aldo. "The Importance of Form in the Teaching of Kanjis." *Sekai no nihongo-kyōiku* 4 (1994): 107-16.

Toyoda, Etsuko. "Kanji Gakushū Ni Taisuru Gakushū-Sha No Ishiki [Student Perception About Kanji Learning]." *Nihongo Kyoiku* 85 (1995): 101–13.

Toyoda, Etsuko. "Teaching Kanji by Focusing on Learners' Development of Graphemic Awareness." *Australian Review of Applied Linguistics* 15 (1998): 155–68.

Tsujimura, Natsuko. *An Introduction to Japanese Linguistics*. Oxford: Blackwell Publishers, 1996.

Vance, Timothy J. *An Introduction to Japanese Phonology*. New York: SUNY Press, 1987.

Wang, Min, Keiko Koda, and Charles A. Perfetti. "Alphabetic and Nonalphabetic L1 Effects in English Word Identification: Lexical and Visual-Orthographic Processes." *Cognition* 87, no. 2 (2003): 129-49.

Wydell, Taeko N. "Nihongo Ni Okeru Yomikaki Shōgai to Moji-Retsu Kara On'in-Retsu Henkan-Ji No Tsubu No Ōkisa to Tōmēsē No Kasetsu (Dyslexia in Japanese and Hypothesis of Granularity and Transparency)." In *Disurekusia: Yomikaki No Ld - Oya to Senmonka No Tame No Gaido (Dislexia: Reading and Writing Ld- a Guide for Parents and Experts)*, edited by Margaret J. Snowling, 271-99. Tokyo: Tokyo Shoseki, 2008.

Yamada, Jun, and Hiroomi Takashima. "The Semantic Effect on Retrieval of Radicals in Logographic Characters." *Reading and Writing* 14, no. 1-2 (2001): 179-94.

Yamashita, Hiroko, and Yukiko Maru. "Compositional Features of Kanji for Effective Instruction." *Journal of the Association of Teachers of Japanese* 34 (2000): 159–78.

Yamashita, Yasufumi. *Suēdengo-Bumpō (Swedish Grammar)*. Tokyo: Daigaku Shorin, 1990.

Japan Foundation. "Japanese-Language Education Overseas 2011: Sweden." http://www.jpf.go.jp/j/japanese/survey/country/2011/sweden.html: 2015-07-30

SAMMANFATTNING PÅ SVENSKA

Denna avhandling utforskar egenskaperna av kanjiinlärningsprocessen hos studenter (med alfabetisk bakgrund) i japanska som andraspråk i jämförelse med nivåanpassade modersmålstalare. Ytterst noggranna och omfattande experiment genomfördes under mycket välkontrollerade villkor med ett stort antal deltagare. Svenska universitetsstuderande i japanska på nybörjar- och avancerad nivå jämfördes med nivåanpassade grupper av japanska modersmåltalare (japanska skolelever i årskurs 2 respektive 5). Experiementen bestod av läs- och skrivtest av kanji med praktiska parallela uppgifter och identiska uppsättningar av måltecken för de respektive nivåanpassade grupperna.

Felklassifikationen baseras på kanjins kognitiva aspekter. Läsfel klassificeras i fonologisk, omständlig, ortografisk och semantisk feltyp, och skrivfel i samma fyra typer samt även pseudokanjityp. Feltyps-förekomstmönster analyseras enligt kunskapstyp (läsning/skrivning), nivå (nybörjare/avancerad) och modersmål (svenska/japanska), med fokus på kanjiprocessningsenhet, föredragna minnesinhämtningsmetoder av tecken/uttal samt läs- och skrivsvårigheter.

Denna studie har kommit fram till ett antal nya slutsatser och bekräftade många olika observationer av tidigare studier. Några av slutsatserna som är unika för denna studie är: (i) L1 fonologisk transfer för svensk nybörjare och dess förminskning på den avancerade nivån; (ii) L2 gruppernas mindre utvecklade ortografiska medvetenhet och dess mindre förbättringsgrad mellan nivårerna; och (iii) olika nivågruppernas egenskaper som visade övergångar mellan nivåerna för L2 grupperna gentemot L1 gruppernas egenskaper som var oförändrad mellan nivåerna. I de bekräftade hypoteserna ingår följande egenskaper av L1 alfabetiska studerande: fonologiska tillvägagångssätt av minnesinhämtning, komponentbaserad processningsenhet, dominans av skrivfel av pseudokanjityp, samt större modersmålsbaserade skillnader än nivåskillnader i läsning men tvärtom i skrivning.

Denna avhandling har visat att svårigheter som studerande med alfabetisk bakgrund upplevt härstammar från övergången från fonembaserad till komponentbaserad grafemprocessning, att använda den mindre bekanta lexikaliska rutten i avkodning och inkodning av grafem-ljud korrespondens, samt användning av mindre effektiva strategier i läsning och skrivning.

APPENDIX 1 Target kanji for LV240

90 characters (113 readings)

遠何夏家歌　画会海外楽　間帰牛魚強　教近元言古　午後語広行
高国黒今作　市紙寺自時　室社首秋週　春書少色食　心新親西切
前走多太体　茶昼長鳥朝　店電冬東道　読南肉馬売　買半父風分
聞米母方北　毎万明門夜　友曜来理話

No.	1	2	3	4	5
Kanji	遠	何	夏	家	歌
Reading	とお(い)	な　に なん	なつ	いえ	うた
No.	6	7	8	9	10
Kanji	画	会	海	外	楽
Reading	ガ	カイ あ(う)	うみ	ガイ そと	ガク たの(しい)
No.	11	12	13	14	15
Kanji	間	帰	牛	魚	強
Reading	カン あいだ	かえ(る)	ギュウ うし	さかな	キョウ つよ(い)
No.	16	17	18	19	20
Kanji	教	近	元	言	古
Reading	キョウ おし(える)	ちか(い)	ゲン	い(う)	ふる(い)
No.	21	22	23	24	25
Kanji	午	後	語	広	行
Reading	ゴ	ゴ あと うし(ろ)	ゴ	ひろ(い)	い(く)
No.	26	27	28	29	30
Kanji	高	国	黒	今	作
Reading	コウ たか(い)	コク くに	くろ(い)	コン いま	つく(る)
No.	31	32	33	34	35
Kanji	市	紙	寺	自	時
Reading	シ	かみ	てら	ジ	ジ
No.	36	37	38	39	40
Kanji	室	社	首	秋	週
Reading	シツ	シャ	くび	あき	シュウ

(continued)

APPENDIX 1:Table *Continued*

No.	41	42	43	44	45
Kanji	春	書	少	色	食
Reading	はる	か(く)	すこ(し)	いろ	た(べる)
No.	46	47	48	49	50
Kanji	心	新	親	西	切
Reading	シン こころ	あたら(しい)	シン おや	にし	セツ き(る)
No.	51	52	53	54	55
Kanji	前	走	多	太	体
Reading	ゼン まえ	はし(る)	おお(い)	ふと(い)	からだ
No.	56	57	58	59	60
Kanji	茶	昼	長	鳥	朝
Reading	チャ	ひる	チョウ なが(い)	とり	あさ
No.	61	62	63	64	65
Kanji	店	電	冬	東	道
Reading	みせ	デン	ふゆ	トウ ひがし	みち
No.	66	67	68	69	70
Kanji	読	南	肉	馬	売
Reading	よ(む)	みなみ	ニク	うま	う(る)
No.	71	72	73	74	75
Kanji	買	半	父	風	分
Reading	か(う)	ハン	ちち	かぜ	ブン わ(かる)
No.	76	77	78	79	80
Kanji	聞	米	母	方	北
Reading	き(く)	こめ	はは	がた	きた
No.	81	82	83	84	85
Kanji	毎	万	明	門	夜
Reading	マイ	マン	あか(るい)	モン	ヤ よる

(continued)

APPENDIX 1:Table *Continued*

No.	86	87	88	89	90
Kanji	友	曜	来	理	話
Reading	とも	ヨウ	ライ く(る)	リ	ワ はなし

APPENDIX 2 Target kanji for LV800

118 characters (143 readings)

圧移因営衛　易益演応往　可仮価河過　快解格確額　刊慣基寄規
技久旧居許　境均禁群経　件券険検限　現減効厚耕　鉱構講混再
妻採際財罪　雑賛志師資　飼識修招承　条情制勢精　製税績設絶
総造増則測　率損退貸断　築程適導独　燃能破犯判　版評布婦富
武復複編報　豊防貿暴務　夢綿輸有余　預留領

No.	1	2	3	4	5
Kanji	圧	移	因	営	衛
Reading	アツ	うつ(る)	イン	いとな(む)	エイ
No.	6	7	8	9	10
Kanji	易	益	演	応	往
Reading	エキ	エキ	エン	オウ	オウ
No.	11	12	13	14	15
Kanji	可	仮	価	河	過
Reading	カ	カ かり	カ	ガ	カ
No.	16	17	18	19	20
Kanji	快	解	格	確	額
Reading	カイ こころよ(い)	カイ と(く)	カク	カク	ガク
No.	21	22	23	24	25
Kanji	刊	慣	基	寄	規
Reading	カン	な(れる)	キ	よ(る)	キ
No.	26	27	28	29	30
Kanji	技	久	旧	居	許
Reading	ギ	キュウ	キュウ	キョ い(る)	ゆる(す)
No.	31	32	33	34	35
Kanji	境	均	禁	群	経
Reading	キョウ さかい	キン	キン	グン む(れ) む(らがる)	ケイ
No.	36	37	38	39	40
Kanji	件	券	険	検	限
Reading	ケン	ケン	ケン けわ(しい)	ケン	ゲン かぎ(り)

(continued)

APPENDIX 2:Table *Continued*

No.	41	42	43	44	45
Kanji	現	減	効	厚	耕
Reading	あらわ(す)	へ(る)	コウ き(く)	あつ(い)	コウ たがや(す)
No.	46	47	48	49	50
Kanji	鉱	構	講	混	再
Reading	コウ	コウ	コウ	コン ま(ぜる)	サ ふたた(び)
No.	51	52	53	54	55
Kanji	妻	採	際	財	罪
Reading	つま	と(る)	サイ	ザイ	ザイ
No.	56	57	58	59	60
Kanji	雑	賛	志	師	資
Reading	ザツ ゾウ	サン	シ こころざし こころざ(す)	シ	シ
No.	61	62	63	64	65
Kanji	飼	識	修	招	承
Reading	か(う)	シキ	シュウ	ショウ	ショウ
No.	66	67	68	69	70
Kanji	条	情	制	勢	精
Reading	ジョウ	ジョウ なさ(け)	セイ	セイ	セイ
No.	71	72	73	74	75
Kanji	製	税	績	設	絶
Reading	セイ	ゼイ	セキ	セッ もう(ける)	た(える)
No.	76	77	78	79	80
Kanji	総	造	増	則	測
Reading	ソウ	ゾウ	ふ(える)	ソク	ソク はか(る)
No.	81	82	83	84	85
Kanji	率	損	退	貸	断
Reading	リツ ひき(いる)	ソン	しりぞ(ける)	か(す)	ダン ことわ(る)
No.	86	87	88	89	90
Kanji	築	程	適	導	独
Reading	きず(く)	テイ	テキ	ドウ みちび(く)	ドク

(continued)

APPENDIX 2:Table *Continued*

No.	91	92	93	94	95
Kanji	燃	能	破	犯	判
Reading	ネン	ノウ	ハ やぶ(る)	ハン	バン
No.	96	97	98	99	100
Kanji	版	評	布	婦	富
Reading	ハン	ヒョウ	プ	フ	フ
No.	101	102	103	104	105
Kanji	武	復	複	編	報
Reading	ム	フク	フク	ヘン あ(む)	ホウ
No.	106	107	108	109	110
Kanji	豊	防	貿	暴	務
Reading	ホウ	ボウ	ボウ	ボウ	つと(める)
No.	111	112	113	114	115
Kanji	夢	綿	輸	有	余
Reading	ム	わた	ユ	あ(る)	ヨ
No.	116	117	118		
Kanji	預	留	領		
Reading	ヨ	ル と(める)	リョウ		

APPENDIX 3 Reading Test for LV240

Instruction in Swedish

Skriv den rätta läsningen av de kanji i fetstil med hiragana i ( ). Försök att fylla i ( ) med din bästa gissning även när du är osäker på den rätta läsningen.
(**Translation:** Write the correct reading of the kanji in boldface with hiragana in ( ). Try to fill in ( ) with your best guess even when you are not sure of the correct reading.)*

Exempel (Example)：学(がっ こう) **校** の **先生**(せんせい) を **見**(み) た。

Instruction in Japanese

ふといかんじのよみを かんじの上の( )の中に かきなさい。
(**Translation:** Write the reading of the kanji in boldface in the ( ) on top of kanji.)

れい (Example): 学(がっ こう) **校** の **先生**(せんせい) を **見**(み) た。

** The latter half of the Swedish instruction was given to the Japanese participants orally (in Japanese).*

Tasks (for both Japanese and Swedish participants)

1. **今**( )、 **風**( ) が **強**( )い。
2. **心**( )が **広**( )い。
3. **友**( )だちの **家**( )。
4. **市長**( )に **会**( )う。
5. **午後**( )のお **茶**( )。
6. **自分**( )で **作**( )る。
7. **来週**( )の **金曜**(きん)日(び)。
8. **秋**( )と **冬**( )の **間**( )。
9. **牛肉**( )を **買**( )う。
10. **母**( )の **話**( )を **聞**( )く。
11. **体**( )が **小**(ちい)さい。
12. **外**( )で **元**( )気に **遊**(あそ)ぶ。
13. **何**( )を **食**( )べますか。
14. **親切**( )な **高校生**( )。
15. **楽**( )しい **夏休**(やす)み。
16. **教室**( )が **多**( )い。
17. **朝**( )と **昼**( )と **夜**( )。
18. **駅**(えき)の **西口**(ぐち)に **行**( )く。
19. **遠**( )い **北**( )の **国**( )。
20. **東**( )の **空**(そら)が **明**( )るい。

21. 首の 長い 鳥。
22. 南の 海の 魚。
23. 紙に 書く。
24. 親が 安心する。
25. 古い 映画。
26. 会社に 近い。
27. 黒い 馬。
28. 少し 高い。
29. 後で 言う。
30. 春の 夕方。
31. 東京の お寺。
32. 歌と 音楽。
33. 門の 前の 道 。
34. 新しい本を読む。
35. 牛が 走って 来る。
36. 父に 電話する。
37. 毎月 一万円 払う。
38. 午前中の勉強。
39. 木を 何本も 切る。
40. 日本語が 分かる。
41. 今夜の月の色。
42. 外国から 帰る。
43. 二 時間半。
44. 太い木の後ろ。
45. お米を売る店。
46. 料理を 教える。

APPENDIX 4 Reading Test for LV800

Instruction in Swedish

Komplettera meningarna/fraserna genom att fylla i rutorna med det rätta kanji. Försök fylla i rutorna med din bästa gissning även när du är osäker på det rätta kanji.
(**Translation:** Write the correct reading of the kanji in boldface with hiragana in ( ). Try to fill in ( ) with your best guess even when you are not sure of the correct reading.)*

Instruction in Japanese

太字の漢字の読みを( )の中に書きなさい。
(**Translation:** Write the reading of the kanji in boldface in the ( ) on top of kanji.)

れい (Example)： 学(がっ こう)**校** の **先生**(せんせい) を **見**(み)た。

** The latter half of the Swedish instruction was given to the Japanese participants orally (in Japanese).*

Exempel (Example)：学(がっ こう)**校** の **先生**(せんせい) を **見**(み)た。

Tasks (for both Japanese and Swedish participants)

1. **圧**力を **測**(　)る。
2. 原**因**を 調(しら)べる。
3. 犬の **群**(いぬ　)れ。
4. 理**解**(り　) 不**可能**(ふ　) 。
5. 事業を **営**(じぎょう　) む。
6. **衛**(　)星 **放**(ほうそう) **送**。
7. **貿易**の 利**益**(　)(り　)。
8. 天気は **快**(てんき)晴(　)。
9. 上手な **演**(じょうず) **技**(　)。
10. 研究を **志**(けんきゅう　) す。
11. **往**(　) **復**(いちじかん) 一時間。
12. 永**久**(えい　) に 死(し)なない。
13. 中 国 **製** 品の **輸**(ちゅうごく　)入(ひん)(　)。
14. **婦**(　)人服(じんふく)の **価格**(　)。
15. **仮**(　)面(めん)で 隠(かく)す。
16. 銀**河**(ぎん　)系(けい)の 星(ほし)。
17. **綿**(　)を つめる。
18. 駅(えき)を 通**過**(つう　)する
19. 問題(もんだい)を **解**(　)く。
20. 正**確**(せい　)な 平**均**(へい　) 気温(きおん)。

せい(　　) よ
21. 成**績**が 良い。

(　　)りょく はんたい
22. **暴** 力 反 対。

がっ こう　(　　)
23. 学校に **慣**れる。

(　　)ほんてき　もん だい
24. **基**本的な 問題。

(　　) みち　(　　)
25. **寄**り道を **許**す。

(　　)しき
26. **旧** 式 の テレビ。

たちいり(　　)し
27. 立入 **禁**止。

えだ　と　(　　)
28. 枝に 飛び **移**る。

いえ　(　　)
29. 家に **居**る日。

ひつじ　たい(　　)
30. 羊の 大**群**。

(　　)ちゅう　あそ
31. **夢**中で 遊ぶ。

(　　)め
32. **境**目が ない。

(　　) じゅうしょ
33. **仮**の 住所。

(　　) (　　)けん
34. **豊富**な **経** 験。

ちょう(　　) はいたつ
35. 朝 **刊** を 配達する。

わりびき(　　) (　　)
36. 割引**券** を **混**ぜる。

(　　)ど ない　(　　) しろ
37. **領**土内に **有**る 城。

ほ(　　)
38. 保 **険** を かける。

(　　)かい
39. ビルを **破**壊する。

(　　) (　　)
40. **制** **限** を **設**ける。

(　　)
41. **評** **判** が いい。

かい ちょう(　　)
42. 会長を **務**める。

(　　)で あ
43. **再** び 出会う。

(　　)きん　まち　(　　)
44. **税** 金で 町を **築**く。

(　　)が　(　　)
45. **版** 画を **破**る。

ひつよう　きん(　　)
46. 必要な 金**額**。

こっ(　　)まち
47. 国**境**の 町。

い　(　　)　よわ
48. 意**志**が 弱い。

ろう か　(　　)し
49. 老化の **防**止。

(　　) しん(　　)
50. **妻** の 心**情**。

のう ち　(　　)さく
51. 農地の **耕** 作。

くすり　(　　)
52. 薬が **効**かない。

じゅう(　) せま
53. 住**居**が 狭い。

てっ(　)せき　ほ
54. 鉄**鉱**石を 掘る。

こん げつ(　　)
55. 今月 **限** り。

かん じ　(　　)せい よう そ
56. 漢字の **構**成要素。

ほん　(　　)
57. 本の **厚** さ。

き かい　(　　)り
58. 機械の **修** 理。

さくら　ぶん(　)ず
59. 桜の 分**布**図。

(　　)
60. セーターを **編**む。

そうきん　(　　)
61. 送金が **絶**える。

ぎんこう(　　)きん　ぞうか(　　)
62. 銀行**預**金 増加**率**。

(　)しゅうかい　　し(　　)
63. **講**習会での 指 **導**。

(　)すう　(　　)
64. **複**数の **犯** **罪** 。

(　　)　かね　(　　)
65. **快**く お金を **貸**す。

(　)ごうびょういん　(　　)けい
66. **総**合病院の **設** 計。

(　)せつ　くるま　かい(　　)
67. **適**切な 車の 改 **造**。

(　)らいしゅう(　)たい
68. **再**来週 **招**待する。

あ ち（ ）
69. 荒れ地を **耕**す。
とり（ ）
70. 鳥が **群**らがる。
しょう たい（ ）
71. 正体を **現** す。
くるま（ ）りょう
72. 車の **燃** 料。
きょう（ ）（ ）
73. 教 **師**が **導**く。
ち（ ）（ ）さん
74. 知**識**は **財** 産だ。
かい ぎ（ ）りょう
75. 会議の **資**料。
（ ）けい（ ）
76. **余**計な **条件** 。
（ ）す ばん でん わ
77. **留**守番電話。
なか ま（ ）
78. 仲間を **率** いる。
くすり（ ）か
79. 薬の **効** 果。
（ ）がい（ ）ど
80. **損**害の **程**度。
（ ）
81. **険**しい山。
（ ）てき せい かつ
82. **規則**的な 生活。
か（ ）
83. 書き**留**める。
（ ）おな ひと
84. **志**の同じ人。
（ ）ぬし はん（ ）
85. **飼**い主の判 **断**。
（ ）ばやし（ ）
86. **雑木**林が **増**えた。
もり（ ）
87. 森できのこを**採**る。
こく（ ）たいかい（ ）えん
88. 国**際**大会の **応**援
（ ）りつ（ ）せい
89. **独**立に **賛**成する。
でん（ ） あそ
90. 伝**承**される遊び。
（ ） みせ
91. **混雑**した 店。
ひと（ ）（ ）
92. 人の**情**けを **断**る。
（ ）しゅうしゃ（ ）こくしょ
93. **編**集者の **報**告書。
む り たい（ ）け
94. 無理な体**勢**で 蹴る。
なん きょく かん（ ）たん（ ）
95. 南極の観 **測** 探 **検**。
（ ）しゃ しゅ ぎょう（ ）しん
96. **武**者修行の **精** 神。
てき（ ）
97. 敵を **退**ける。
せいさん（ ）
98. 生産が **減**る。

APPENDIX 5 Writing Test for LV240

Instruction in Swedish

Komplettera meningarna/fraserna genom att fylla i rutorna med det rätta kanji. Försök fylla i rutorna med din bästa gissning även när du är osäker på det rätta kanji.
(**Translation:** Complete the sentences/phrases by filling in the boxes with the correct kanji. Try to fill in the box with your best guess even when you are not sure of the correct kanji.)*

Exempel (Example)：学(がっこう)校の先(せん)生(せい)を見(み)た。

Instruction in Japanese

文のいみに あうように、□の中に かんじを かきなさい。
(Translation: Fill in the box with kanji to complete the sentence.)

れい (Example)：学(がっこう)校の先(せん)生(せい)を見(み)た。

** The latter half of the Swedish instruction was given to the Japanese participants orally (in Japanese).*

Tasks (for both Japanese and Swedish participants)

1. □(いま)、□(かぜ)が□(つよ)い。
2. □(こころ)が□(ひろ)い。
3. □(とも)だちの□(いえ)。
4. □□(しちょう)に□(あ)う。
5. □□(ごご)のお□(ちゃ)。
6. □□(じぶん)で□(つく)る。
7. □□(らいしゅう)の金□(きんようび)日。
8. □(あき)と□(ふゆ)の□(あいだ)。
9. □□(ぎゅうにく)を□(か)う。
10. □(はは)の□(はなし)を□(き)く。
11. □(からだ)が□(ちい)さい。
12. □(そと)で□(げんき)気に□(あそ)ぶ。
13. □(なに)を□(た)べますか。
14. □□(しんせつ)な□(こうこうせい)校生。
15. □しい□(なつやす)休み。
16. □□(きょうしつ)が□(おお)い。
17. □(あさ)と□(ひる)と□(よる)。
18. 駅(えき)の□(にしぐち)口に□(い)く。
19. □(とお)い□(きた)の□(くに)。
20. □(ひがし)の空(そら)が□(あか)るい。

21. □の□い□。
22. □の□の□。
23. □に□く。
24. □が安□する。
25. □い映□。
26. □ □に□い。
27. □い□。
28. □し□い。
29. □で□う。
30. □の夕□。
31. □ 京のお□。
32. □と音□。
33. □の□の□ 。
34. □しい本を□む。
35. □が□って□る。
36. □に□ □する。
37. □月一□円払う。
38. □ □中の勉□。
39. 木を□本も□る。
40. 日本□が□かる。
41. □ □の月の□。
42. □ □から□る。
43. 二□ □ □。
44. □い木の□ろ。
45. お□を□る□。
46. 料□を□える。

APPENDIX 6 Writing Test for LV800

Instruction in Swedish

Komplettera meningarna/fraserna genom att fylla i rutorna med det rätta kanji. Försök fylla i rutorna med din bästa gissning även när du är osäker på det rätta kanji.

Exempel: 先(せん)｜生(せい)が 空(そら)を 見(み)た。

Instruction in Japanese

文のいみに 合うように、□の中に 漢字を 書きなさい。

1. □(あつ)力を□(りょく)(はか)る。
2. 原□(げん)(いん)を調(しら)べる。
3. 犬の□(いぬ)れ(む)。
4. 理□(り)不□(かい)(ふ)□(か)(のう)。
5. 事業を□(じぎょう)む(いとな)。
6. □(えい)星放送(せい)(ほうそう)。
7. □(ぼう)□(えき)の利□(えき)。
8. 天気は□(てんき)晴(かいせい)。
9. 上手な□(じょうず)□(えん)(ぎ)。
10. 研究を□(けんきゅう)す(こころざ)。
11. □(おう)□(ふく)一時間(いちじかん)。
12. 永□(えい)に死(し)なない。
13. 中国□(ちゅうごく) 品の□(ひん)入(にゅう)。
14. □(ふ)人服の□(じんぷく)□(か)(かく)。
15. □(か)面(めん)で隠(かく)す。
16. 銀□(ぎん)系の星(けいぼし)。
17. □(わた)をつめる。
18. 駅を通□(えき)(つう)する。
19. 問題を□(もんだい)く(と)。
20. 正□(せい)(かく)な平□(へい)気温(きおん)。
21. 成□(せいせき)が良(よ)い。
22. □(ぼう)力反対(りょくはんたい)。
23. 学校に□(がっこう)れる(な)。
24. □(き)本的(ほんてき)な問題(もんだい)。
25. □(よ)り道(みち)を□(ゆる)す。
26. □(きゅうしき)式のテレビ。
27. 立入(たちいり)□(きん)止(し)。
28. 枝に飛(えだ)び□(と)る(うつ)。
29. 家に□(いえ)る(い)日。
30. 羊の大□(ひつじ)(たいぐん)。
31. □(む)中(ちゅう)で遊(あそ)ぶ。
32. □(さかい)目(め)がない。
33. □(かり)の住所(じゅうしょ)。
34. □(ほう)□(ふ)な□(けい)験。

35. 朝□を配達する。
36. 割引□を□ぜる。
37. □土内に□る城。
38. 保□をかける。
39. ビルを□壊する。
40. □ □を□ける。
41. □ □がいい。
42. 会長を□める。
43. □び出会う。
44. □金で町を□く。
45. □ 画を□る。
46. 必要な金□。
47. 国□の町。
48. 意□が弱い。
49. 老化の□止。
50. □の心 □。
51. 農地の□作。
52. 薬が□かない。
53. 住□ が狭い。
54. 鉄□石を掘る。
55. 今月□り。
56. 漢字の□成要素。
57. 本の□さ。
58. 機械の□理。
59. 桜の分□図。
60. セーターを□む。
61. 送金が□える。
62. 銀行□金増加□。
63. □習会での指□。
64. □数の□ □。
65. □くお金を□す。
66. □合病院の□計。
67. □切な車の改□。
68. □来週□ 待する。
69. 荒れ地を□す。
70. 鳥が□らがる。
71. 正体を□す。
72. 車の□ 料。
73. 教□が □く。
74. 知□は□産だ。
75. 会議の□料。
76. □計な□ □ 。
77. □守番電話。
78. 仲間を□いる。
79. 薬の□果。
80. □害の□度。
81. □しい山。
82. □ □的な生活。

83. 書き□める。
84. □の同じ人。
85. □い主の判□。
86. □木林が□えた。
87. 森できのこを□る。
88. 国□大会の□援
89. □立に□成する。
90. 伝□される遊び。
91. □□した店。
92. 人の□けを□る。
93. □集者の□告書。
94. 無理な体□で蹴る。
95. 南極の観□探□。
96. □者修行の□神。
97. 敵を□ける。
98. 生産が□る。