

Doctoral Dissertation

THE ROLE OF LANGUAGE IN MANGA:  
FROM THE POINT OF VIEW OF STRUCTURE, VOCABULARY AND  
CHARACTERS

UNSER-SCHUTZ, Giancarla

Graduate School of Social Sciences

Hitotsubashi University

SD091028

マンガにおける言葉の役割

—構造・語彙・登場人物という三つの観点から—

ウンサーシュッツ・ジャンカーラ

博士論文

一橋大学大学院社会学研究科

# Table of Contents

List of Examples .....	v
List of Figures .....	v
List of Graphs .....	vii
List of Tables .....	vii
Abstract.....	1
 1. A Linguistic Analysis of Manga: An Introduction .....	 2
1.1 Chapter themes and overview .....	2
1.1.1 Manga in everyday Japanese life .....	2
1.1.2 Language and manga .....	5
1.1.3 Creating a manga corpus .....	8
1.2 Corpus Design and Construction .....	10
1.2.1 Issues in creating a manga corpus .....	10
1.2.2 Sampling issues .....	11
1.2.3 Structural design .....	14
1.3 Current State of the Corpus .....	20
1.3.1 Data summary .....	20
1.3.2 Potential uses and interests of the corpus .....	21
1.3.3 Unique issues .....	23
1.4 Overview of the following chapters .....	27
 2. The Structural Characteristics of Language in Manga .....	 30
2.1 Chapter themes and overview .....	30
2.1.1 Introduction .....	30
2.1.2 Text in comics and manga .....	32
2.2 Data analysis .....	39
2.2.1 Text types in the corpus: The distribution of categories .....	39
2.2.2 Textual categories and genre .....	43
2.2.3 Defining handwritten lines, thoughts and authorial comments .....	43

2.3	Analyzing <i>Background Lines/Thoughts</i> and <i>Comments</i> .....	49
2.3.1	The relationship of <i>Background Lines/Thoughts</i> and <i>Comments</i> with genre.....	49
2.3.2	Interpreting the roles of <i>Background Lines/Thoughts</i> and <i>Comments</i> .....	53
2.3.3	On the need to use different textual styles .....	55
2.4	Conclusions.....	60
3.	The Orthographic and Lexical Characteristics of Manga .....	62
3.1	Chapter Themes and Overview .....	62
3.2	Orthography in manga: Studies on writing types and furigana.....	64
3.2.1	Orthography in language research .....	64
3.2.2	Orthography and Japanese .....	65
3.2.3	Study 1 – Writing types used: Methodology.....	69
3.2.4	Study 1 – Writing types used: Results .....	70
3.2.5	Study 2 – Use of furigana: Methodology.....	82
3.2.6	Study 2 – Use of furigana: Results.....	84
3.2.7	Discussion: Manga as spoken language, seen through orthography .....	85
3.3	Readability and the skills necessary for reading manga .....	89
3.3.1	Readability and manga.....	89
3.3.2	Methodology .....	91
3.3.3	Results.....	92
3.3.4	Discussion.....	94
3.4	A Morpheme Analysis of Manga.....	97
3.4.1	Japanese morpheme analysis .....	97
3.4.2	Methodology .....	101
3.4.3	Results.....	105
3.4.4	Discussion.....	117
3.5	Conclusions.....	120
4.	Language Patterns and Characterization.....	121
4.1	Chapter themes and overview .....	121
4.1.1	Japanese gendered language .....	122
4.1.2	Gendered language and manga .....	124

4.2	Study 1: Distribution of Lines by sex and number of characters .....	129
4.2.1	Methodology .....	129
4.2.2	Results.....	130
4.2.3	Impact of the distribution of characters on speech patterns .....	134
4.3	Study 2: Distribution of personal pronouns in Lines .....	136
4.3.1	Methodology .....	136
4.3.2	Results.....	139
4.4	Study 3: Distribution of sentence final expressions in Lines .....	143
4.4.1	Methodology .....	143
4.4.2	Results.....	147
4.5	Discussion .....	154
4.5.1	Differences in characters by genres .....	154
4.5.2	Potential explanations: Author interference, character roles and personal distance .....	155
4.5.3	Masculine personal pronouns and girls.....	159
4.6	Conclusions.....	161
5.	Conclusions.....	163
5.1	Overview .....	163
5.1.1	Summary of Chapters 1 to 4 .....	163
5.1.2	Genre, manga and language .....	165
5.1.3	Future issues .....	166
5.2	Education and manga.....	167
5.2.1	Ways and advantages of using manga for learning.....	167
5.2.2	Possible problems made clear by the corpus.....	169
5.2.3	Rethinking manga in the classroom .....	171
5.3	Final words .....	173
6.	Appendix.....	174
6.1	Corpus samples .....	174
6.2	Gloss .....	186
6.3	Previous research .....	190

7.	Bibliography .....	192
7.1	Referenced texts.....	192
7.2	Series used in corpus.....	205

# List of Examples

Example 1: Basic input style of text into corpus.....	19
Example 2: Three examples of [ne:]/[nei] .....	25
Example 3: Text with and without furigana.....	67
Example 4: Text as analyzed by ChaMame .....	102

# List of Figures

Figure 1 (Left): The manga shelf at a convenience store in Tokyo .....	2
Figure 2 (Middle): Toys and accessories from the popular series <i>One Piece</i> at a convenience store in Tokyo ...	2
Figure 3 (Right): Tie-up magazines for hot spring vacations following the popularity of the manga-cum-movie <i>Thermae Romae</i> .....	2
Figure 4 (Left): Examples of <i>Lines</i> and <i>Thoughts</i> from <i>One Piece</i> (1:178).....	16
Figure 5 (Middle): Examples of <i>Lines</i> and <i>Thoughts</i> from <i>Detective Conan</i> (2:91) .....	16
Figure 6 (Right): Examples of <i>Lines</i> and <i>Thoughts</i> from <i>Death Note</i> (3:62).....	16
Figure 7 (Left): Example of <i>Narration</i> in captions from <i>One Piece</i> (1:9).....	16
Figure 8 (Right): Example of <i>Narration</i> in captions from <i>RabuKon</i> (1:9) .....	16
Figure 9 (Left): Examples of <i>Onomatopoeia</i> , from <i>Naruto</i> (1:19) .....	17
Figure 10 (Right): Examples of <i>Onomatopoeia</i> , from <i>Nodame Cantabile</i> (3:118) .....	17
Figure 11 (Left): Examples of <i>Background Text</i> from <i>Bokura ga Ita</i> (3:25).....	17
Figure 12 (Right): Examples of <i>Background Text</i> from <i>GinTama</i> (2:47).....	17
Figure 13 (Far left): Examples of <i>Background Lines/Thoughts</i> in <i>Naruto</i> (1:18).....	18
Figure 14 (Mid-left): Examples of <i>Background Lines/Thoughts</i> in <i>Kimi ni Todoke</i> (1:84).....	18
Figure 15 (Mid-right): Examples of <i>Comments</i> in <i>Nodame Cantabile</i> (3:128) .....	18
Figure 16 (Far right): Examples of <i>Comments</i> in <i>Kimi ni Todoke</i> (1:134) .....	18
Figure 17 (Left): Examples of <i>Titles</i> from <i>GinTama</i> (2:10) .....	18
Figure 18 (Right): Examples of <i>Titles</i> from <i>Nana</i> (1:106) .....	18
Figure 19: Example of speech bubbles where it is difficult to determine the speaker and natural order in <i>Nodame Cantabile</i> (2:47).....	34
Figure 20: Example of speech bubbles allowing for movement in <i>City of Glass</i> (8) .....	35
Figure 21 (Left): Example of <i>Background Lines/Thoughts</i> surrounded by lines in <i>Naruto</i> (1:121).....	45
Figure 22 (Right): Example of <i>Comments</i> with an arrow from Yazawa Ai's <i>Nana</i> (2:135) .....	45
Figure 23: Example of <i>Narration</i> using past tense from <i>Nana</i> (2:60) .....	46
Figure 24 (Left): Example of <i>Background Lines/Thoughts</i> with shell-like, extra characters from Shiina Karuho's <i>Kimi ni Todoke</i> (1:128) .....	47

Figure 25 (Right): Example of a critical <i>comment</i> in Ninomiya Tomoko's <i>Nodame Cantabile</i> (2:136).....	47
Figure 26: Example of comments and <i>Background Lines/Thoughts</i> in <i>Naruto</i> (1:111) .....	50
Figure 27: Comments-like text seen in <i>Chibi Maruko-chan</i> (1:17) .....	52
Figure 28: <i>Background Lines/Thoughts</i> and comments as seen in Keiko Takemiya's <i>Kaze to Ki no Uta (Song of the Wind and Trees)</i> (1:67) .....	54
Figure 29: Example of comment-like text within bubbles in Takemiya Keiko's <i>Watashi wo Tsuki made Tsuretette! (Take Me to the Moon!)</i> (1:56) .....	58
Figure 30: Non-phonetic furigana from <i>Bokura ga Ita</i> (1:118) .....	68
Figure 31: An example of non-standard furigana in the title of <i>Boku ga Dorobou ni Natta Wake</i> .....	69
Figure 32: An example of a chou-bousen being used sentence-initially <i>RabuKon</i> (2:40).....	78
Figure 33: An example of a strongly ended sentence using the kuten from <i>RabuKon</i> (2:129).....	78
Figure 34 (Left): The touten to describe pauses after a conjunction in <i>Detective Conan</i> (1:12) .....	79
Figure 35 (Right): The space being used to describe pauses after a conjunction in <i>Death Note</i> (3:127) .....	79
Figure 36 (Left): Example of sweat marks in <i>Kimi ni Todoke</i> (1:155) .....	80
Figure 37 (Middle): Example of sweat marks in <i>RabuKon</i> (1:104).....	80
Figure 38 (Right): Example of sweat marks in <i>Detective Conan</i> (1:18).....	80
Figure 39: Example of a bousen being used to express a long vowel in <i>RabuKon</i> (1:119).....	81
Figure 40: Example of a bousen being used for an irregularly elongated vowel in spoken speech in <i>Nodame Cantabile</i> (2:65).....	81
Figure 41: Example of dakuon-ten being used with vowels in <i>GinTama</i> (1:90) .....	82
Figure 42 (Left): Example <i>Emphasis</i> furigana in <i>Bokura ga Ita</i> (1:161).....	83
Figure 43 (Middle): Example of <i>Censorship</i> furigana in <i>RabuKon</i> (1:40) .....	83
Figure 44 (Right): Example of <i>Explanatory</i> furigana (Japanese translation) in <i>Nodame Cantabile</i> (1:34).....	83
Figure 45 (Right): Example of <i>Explanatory</i> furigana (Japanese definition).....	83
Figure 46 (Left): Example <i>Ateji</i> furigana (deictic) in <i>GinTama</i> (3:41).....	83
Figure 47 (Middle): Example of <i>Ateji</i> furigana (gairaigo) in <i>One Piece</i> (1:146).....	83
Figure 48 (Right): Example of <i>Ateji</i> (synonyms) in <i>Bokura ga Ita</i> (1:154).....	83
Figure 49 (Far left): Example of <i>Pronunciation</i> furigana (letters/symbols) in <i>Kimi ni Todoke</i> (3:64).....	84
Figure 50 (Mid-left): Example of <i>Pronunciation</i> furigana (slang) in <i>Detective Conan</i> (1:108).....	84
Figure 51 (Middle): Example of <i>Pronunciation</i> furigana (foreign languages) in <i>Nodame Cantabile</i> (2:100) ...	84
Figure 52 (Mid-right): Example of <i>Pronunciation</i> furigana (katakana) in <i>Nana</i> (3:150).....	84
Figure 53 (Far right): Example of <i>Pronunciation</i> furigana (foreigner talker) in <i>GinTama</i> (1:115).....	84
Figure 54: Example of non-standard writing of <i>desu-ka</i> (ですか→でスカ) in <i>Nodame Cantabile</i> (2:23) .....	99
Figure 55: Lines split between multiple speech bubbles in <i>Detective Conan</i> (2:70) .....	119
Figure 56 (Above): Usage of <i>atashi</i> by a female character in <i>One Piece</i> (1:69) .....	140
Figure 57 (Below): Usage of <i>watashi</i> by a female character in <i>One Piece</i> (1:207).....	140

Figure 58 (Left): Usage of <i>ore</i> by a female character in <i>GinTama</i> (3:160) .....	141
Figure 59 (Right): Usage of <i>atashi</i> by a male character in <i>RabuKon</i> (3:87).....	141
Figure 60 (Right): Example of a female character using a mildly masculine SFE ( <i>n-da-yo-ne</i> ) in <i>Kimi ni Todoke</i> (1:134).....	150
Figure 61 (Left): Example of female characters using strongly masculine SFEs ( <i>ze</i> ) in <i>Nana</i> (2:135).....	150
Figure 62: Example of a female character using a neutral SFE ( <i>ka-na</i> ) in <i>Naruto</i> (2:177) .....	150
Figure 63 (Left): Example of a female character using a mildly feminine SFE ( <i>mon</i> ) in <i>RabuKon</i> (1:92).....	150
Figure 64 (Right): Example of female characters using a strongly feminine SFE ( <i>wa</i> ) in <i>Detective Conan</i> (3:46) .....	150
Figure 65 (Left): Example of a male character using a moderately masculine SFE ( <i>yo</i> ) in <i>Death Note</i> (3:62)	153
Figure 66 (Right): Example of male characters using strongly masculine SFEs ( <i>zo</i> ; <i>nai</i> → <i>nee</i> ; the command form) in <i>One Piece</i> (1:10) .....	153
Figure 67: Example of a male character using a neutral SFE ( <i>jan</i> ) in <i>Kimi ni Todoke</i> (2:14) .....	153
Figure 68: Example of a male character using a mildly feminine SFE ( <i>desho</i> ) in <i>Detective Conan</i> (2:91) .....	153
Figure 69: Example of a male character using a strongly feminine SFE ( <i>kashira</i> ) in <i>Nodame Cantabile</i> (2:27) .....	153

## List of Graphs

Graph 1: Distribution of text across series by text type, with details for <i>Background Lines/Thoughts</i> and <i>Comments</i> .....	50
Graph 2: Visualization of the distribution of text types .....	71
Graph 3: Distribution of text within onomatopoeia .....	74
Graph 4: Distribution of parts of speech - unique words vs. number of occurrences .....	107
Graph 5 (Left): Scatter plot of the correlation between female and male characters in shoujo-manga .....	131
Graph 6 (Right): Scatter plot of the correlation between female and male characters in shounen-manga .....	131
Graph 7 (Left): Number of identifiable male characters by series .....	133
Graph 8 (Right): Number of identifiable female characters by series .....	133
Graph 9: Distribution of <i>Lines</i> by character types .....	133
Graph 10: Ordered distribution of most common SFPs by gender and genre .....	154

## List of Tables

Table 1: Series in corpus .....	14
Table 2: Overall data totals .....	20



Table 3: Summary of data by categories, characters and entries .....	21
Table 4: Summary of the roles and environments of the major text categories in manga .....	40
Table 5: Distribution of text categories by genre and series .....	42
Table 6: Distribution of text across genres .....	71
Table 7: Distribution of script types by category .....	73
Table 8: Text types by manga-genre .....	73
Table 9: Use of kanji by type .....	75
Table 10: Kanji usage by genre .....	76
Table 11: Types of symbols seen .....	77
Table 12: Types of furigana .....	83
Table 13: Distribution of furigana .....	84
Table 14: Forms and distribution of [ne:] .....	88
Table 15: Comparison of the readability test results .....	93
Table 16: Comparative genre statistics using the Japanese Readability Test .....	93
Table 17: Comparative genre statistics using Obi2 .....	93
Table 19: Parts of speech for various mediums .....	106
Table 18: Number of words by type .....	106
Table 20: Types of word seen by origin for various mediums .....	109
Table 21: Number of sentence-forming predicates seen by genre and series .....	110
Table 22: Most frequently seen nouns .....	115
Table 23: Most frequently seen verbs .....	115
Table 24: Most frequently seen adjectives .....	116
Table 25: Most frequently seen adverbs .....	116
Table 26: Number of <i>Lines</i> by the sex of speakers and genre .....	129
Table 27: Number of characters seen in each series .....	130
Table 28: Percentages of text (characters) spoken by the main male and female characters for each series ....	134
Table 29: Types of personal pronouns .....	137
Table 30: Number and types of personal pronouns seen by genre, series and gender (details) .....	138
Table 31: Types of personal pronouns used by female characters .....	139
Table 32: Types of personal pronouns used by male characters .....	141
Table 33: Types of sentence final expressions by gender .....	144
Table 34: Types of sentence final expressions seen .....	146
Table 35: Sentence final expressions used by female characters .....	148
Table 36: Sentence final expressions used by male characters .....	151
Table 37: SFP distributions in other studies .....	155

Table 38: Age of authors.....	157
-------------------------------	-----

# Abstract

This thesis details a corpus-based analysis of the role and nature of language in manga, thereby adding to the understanding of how manga functions as a medium and how language affects how manga are read. Research on manga has generally focused on their visual properties, with most of their linguistic properties still not clear. In Chapter 1, after reviewing the previous literature, I introduced a corpus of 10 popular series, which I compiled in order to rectify these gaps in knowledge, and which is characterized by its inclusion and categorization of all language in manga. The subsequent chapters then dealt with language in manga at three, interrelated levels. First, I looked at the *structural* characteristics of manga through an examination of the distribution of the types of text seen (Chapter 2). I particularly examined how two text types—*Background Lines/Thoughts* and *Comments*—seem to differ in usage by genre, arguing that they are the result of different reading experiences being offered. Then, I moved to the *orthographic* and *lexical* characteristics of manga, particularly those found in *Lines* (Chapter 3). To do so, I conducted three studies on the use of scripts and orthographic symbols; the readability of manga; and a morpheme analysis. The results of these studies helped to understand how manga is located within a spectrum of different texts, showing that they have many similarities with other new media as well as to the spoken speech they emulate. Finally, I discussed how language is manipulated to create characters, focusing on the use of gendered language (Chapter 4). By looking at how characters use personal pronouns and sentence final expressions, I demonstrated that manga for boys use more stereotypical speech patterns, particularly in regards to female characters. I argued that these differences are caused not only by author interference, but also by the roles of female characters in the text and the interpersonal distance between them and other characters. In this way, the discussion moved from more general linguistic characteristics to those more in-depth, creating a fluid transition towards more content-oriented analyses.

# 1. A Linguistic Analysis of Manga: An Introduction

## 1.1 Chapter themes and overview

### 1.1.1 Manga in everyday Japanese life

Whether one is a reader or not, *manga*, or Japanese comics, are as a part of Japanese life today—if not more—than the stereotypical images of Japan such as kimono or sushi. They are so central a part of everyday Japanese life that it is literally almost impossible to ignore them. Regardless of one's interest, or lack thereof, manga will be one of the first things that enters one's vision at any of the many common *konbini* (convenient stores), where they often take up one of the best spots right by the door, and it is the local *raamen-ya* (ramen shop), *chuuka-ryouriya* (Chinese restaurants) or *sentou* (public baths) that does *not* have any manga lying about that is the unusual one (Figure 1).

**Figure 1 (Left):** The manga shelf at a convenience store in Tokyo

**Figure 2 (Middle):** Toys and accessories from the popular series *One Piece* at a convenience store in Tokyo

**Figure 3 (Right):** Tie-up magazines for hot spring vacations following the popularity of the manga-cum-movie *Thermae Romae*



(Photos by author)

Manga by some accounts make up 40% of all items published in Japan (Schodt, 1996), and while some (e.g., Nakano (2009)) have pointed to recent decreases in readership, they continue to be of great importance

because of their overarching reach within Japanese media, with huge licensed industries in animation and live-action television and movies, video games, novels and various non-comic forms of literature, and character goods, etc. (Figure 2, Figure 3). While sales numbers going down for read-and-toss manga magazines in face of increasing use of other forms of entertainment such as cellular phones may mean that it is not as common as it once was to find people reading manga in the subways—popularly noted as a surprising point of difference between Japan and other countries (Pink, 2007)—ads for popular manga magazines and titles still are a common part of train advertising, and they themselves are now diversifying onto online markets (Takekuma, (2006), Mihara, Nagamori & Sugimoto (2011)).

With the expansion of the international market, manga have also become a crucial part of what represents Japan abroad. Translated and published internationally, they are also increasingly the face of Japan abroad, and the number of people studying Japanese as a second language has also sharply increased in recent years, for which manga and Japanese popular culture have often been given as one of the important motivating reasons (Kumano & Hirokaga, 2008). This is all the more crucial given the decline of other reasons, such as business, for studying Japanese (Wasabi Brothers Trading Company, 1998). In 2006, the manga market peaked at over 200 million dollars in the United States (Japan External Trade Organization, 2012), and series such as Masashi Kishimoto's *Naruto* regularly break out onto the best-selling book lists, something uncommon even for popular American comics (Goldberg, 2010). While sales numbers have since gone down dramatically, popular titles still feature regularly on the top-50 graphic novels sales list, and the reasons for the declines seem to be more related to changes in the markets, such as the disappearance of traditional bookstores (ICV2, 2012) and the time it has taken for manga publishers in the U.S. to go digital (MacDonald, 2012). Recognition of manga as an important cultural resource has led many to call for more serious work into establishing Japan's role as a source of cultural brands, such as Nakano (2009) and Sakurai (Anime Bunka Gaikou, 2009), (Sekai Kawaii Kakumei, 2009). While not all attempts to cash into manga's popularity have been successful—the then-prime minister Taro Aso's attempt to establish a National Comprehensive Media Arts Center (*Kokuritsu Media Geijutsu Sougou Sentaa*) were widely criticized as being just *manga-kissa* (manga cafes), and a waste of public resources in poor economic times (Hayakawa & X(Ekkusu), 2009)—many real achievements have been made, such as the establishment of the Kyoto International Manga Museum, which in its first year alone drew in 220,700 people (Corkill, 2008), the equivalent of two-thirds of the people who went to the Kyoto National Museum (331,605 in 2007 (Kyoto National Museum, 2009)), a long-established, well-respected, definite must-see on the Kyoto museum lists; interesting, it even managed to draw in 30,000 foreigners that year, somewhat less than 15% of all of their visitors, a feat for any Japanese museum (Corkill, 2008). The founding of the Japan Society for Studies in Cartoon and

Comics in 2001 also represented a new beginning of serious academic study of manga, thus reflecting the new times.

This thesis, too, could be said to be a reflection of these changes in attitudes towards manga, as one of the many works to have started approaching manga in an academic way. The focus specifically will be on the language seen in manga, keeping in mind the important role that language plays, and how manga are now an important source of Japanese data, both domestically and internationally. To do this I constructed an original, linguistic corpus of manga, collecting all of the textual data found within the first three volumes of ten series. While the following chapters will look into data analysis—specifically focusing on the structural and lexical characteristics of manga, followed by an examination of characters’ speech patterns—as I will demonstrate in the following sections, language has hitherto been a relatively un-approached issue within manga studies, often quickly dismissed as secondary to manga’s visual aspects, and plagued by speculation in the studies that do exist. By conducting this study, I hope to offer some unique insights into how manga function as a medium by comprehensively looking at how language works in manga, and how that might differ from other media. Additionally, however, I also intend to outline how manga could be viewed as resources for other kinds of linguistic research, particularly through demonstrating the unique qualities of the corpus I constructed. I would also like to suggest that this might form a useful resource for others dealing with similar texts.

With these points in mind, in the next sections of this introductory chapter I will first describe how manga are produced and read, paying particular attention to the role of language in those processes and their historical contexts. This will set the stage for talking about how manga are currently researched, wherein I will argue that sufficient attention has not been given to language. With this background in place, I will then describe how and why the corpus was constructed, particularly in light of sampling and structural issues. After describing how the corpus was sampled and designed within a larger corpus linguistic framework, and describing the current amount of data in the corpus, I then go over possible ways to use the corpus, discussing some problems that have been encountered thus far. As a second line of analysis, this introduction will also serve to outline some of the reasons for being interested in a source of text that has been traditionally undervalued. As I will argue, manga form a potentially interesting resource pool not only to those interested in researching Japanese popular culture and how Japanese culture is disseminated and absorbed abroad, but also to those interested in researching topics in Japanese, without any necessary interest in popular culture itself. Finally, I will conclude the chapter with a brief overview of what the following chapters have in store.

### 1.1.2 Language and manga

In assessing the role that language plays in manga, it is important to take a step back and consider the background of manga. Although manga may only just recently have started to gain attention as a legitimate subject of academic research, the most common forms of manga today have not changed very much since the current market developed in the postwar period. Even with the increasing market for online-manga (Takekuma, (2006), Mihara, Nagamori & Sugimoto (2011)), manga today generally appear first in serial chapters of approximately 20 to 40 pages in large, phone-book size magazines, which generally include several series for anywhere between 250 to 600 pages. Magazines themselves are generally viewed as belonging to different genres based upon their imagined readership, with the traditional division being between titles for young boys (*shounen-manga*) and titles for young girls (*shoujo-manga*). Most such titles are monthly, but some, such as *Shounen Jump*, the best-selling title, are biweekly. After several chapters have been printed, series are republished as *tankoubon*, pocket novel sized books containing approximately 180 pages. Popular series may continue indefinitely, with some continuing to over 100 volumes.

Manga themselves have both a short and a very long history. Depending on one's definition of manga, one might point to 12th century scrolls depicting people and animals humorously (Ito, 2005, p. 458). The word manga itself was first used to describe humorous, one-frame woodblock prints by Hokusai in the 18th century (Shimizu, 1991, pp. 15-28). Along with Westernization after the Meiji Restoration (1868) came new political comics in magazines like Japan Punch. Longer, story-based series began to appear in the early 20th century, but manga could be said to have come to their current form in the post-war period. Natsume (1997, pp. 14-17) ascribes this post-war boom to the ease of printing manga: They were a cheap, easy-to-produce form of entertainment, with a large audience of entertainment-deprived Japanese readers. The form of manga that developed during that period is commonly called *sutoorii-manga* (story-manga) and continues to form the majority of the current market. One of the primary characteristics of story-manga are their strong narratives (Takeuchi I. , 2006, pp. 225-227); as a result, they usually feature continuing story arcs which, while extended in length, have eventual ends and whose characters develop over time.

The emphasis on narrative in the modern story-manga might be considered to be directly associated with an emphasis on language: The length and complexity of plots necessitate a way to transmit those stories, and language is clearly one important way of expressing detailed narratives. As Osamu Takeuchi (2005) describes, the use of language found in story-manga was both unique in comparison to previous manga types, and essential to creating a sense of time and space in stories. Likewise, while wordless comics do exist, they require more work on the part of both the author and the reader, and "(t)he narrative in wordless comics, therefore,

demands more visual competency from readers than in traditional comics” (Beronă, 2001, p. 20). In the case of mass-produced, mass-consumed story-manga, it seems safe to assume that language is a core way of explicating plot to readers.

Manga thus constitute an enormous base of linguistic information, and popular reaction to them supports that. A 2008 survey by the Japanese Agency for Cultural Affairs (2010) showed that approximately 45% of respondents gave manga as being influential in the speech of young people, coming in only after television, friends and parents. Manga are indeed a wide-read linguistic form amongst young people. In fact, in a survey I conducted in 2007 and 2008 of manga reading habits amongst high school students in Tokyo and Okinawa, not only did a majority of students (385 out of 501) say that they had talked to friends about manga before, but a majority (288) also said they had a favorite manga series. Perhaps because manga are widely read amongst young people, they are not uncommonly given as a source of ‘unusual’ speech patterns. Manga are, for example, often given as the reason some young girls have taken to use the masculine first person pronouns *ore* and *boku*. In that vein, Endo (2001) found that manga were one of the reasons college students’ gave for why girls were thought to use *ore* and *boku*. As I will discuss in Chapter 4, this is an explanation common within popular discourse, expressed particularly clearly in posts about girls’ usage of *ore* and *boku* on Yahoo Chiebukuro, the Japanese equivalent of Yahoo Q&A.

The question of whether or not manga are really influential on young people’s speech, and if so, how, is a difficult one and may be better thought of as a complex process. Ueno (2006, pp. 16-17) comments in her analysis of gendered sentence-final particles in manga for young girls and older women that manga are on the one hand “likely to reflect the characteristics of the desired readership” in order to create a sense of rapport, yet at the same time they may “...influence the readers in framing their gender identities as well as their linguistic behavior.” Language in manga is subject to the same burdens that other works of fiction are, and there is a clear tension between representing realistic speech that readers can relate to while using speech to create characters and dealing with the interference of the author’s own speech patterns. As I found in my own examination of gendered sentence-final particles—described in Chapter 4—in at least some types of manga, sentence-final particles seem distributed somewhat differently from real-life usage. For example, female characters use more realistic forms in series for girls and more conservative ones in series for boys, and factors such as the roles of characters and the sex of authors were all raised as possible reasons for this. Language in manga is thus subject to many different factors, offering a rich layer of complexity, and this gap between how manga are perceived and their linguistic reality is one worthy of interest from both social and linguistic standpoints.

Whatever the status of their role in society, however, with the clear importance of manga in Japan and their role as a source of linguistic information, they would seem an obvious choice of research. Indeed, as I



previously noted, with the attention given to manga as a cultural symbol there has been a rise in interest in their potential as a subject of serious academic pursuit. Yet for all that manga research is coming to the fore, one notices that there are distinct limiting trends. In particular, there is a tendency to examine them on a visual basis. Such visual-based research tends to examine the role of frames, panels, gutters, etc., in creating a “comics-grammar”. An unfortunate result of this focus is that much of the research on manga, and comics in general, assumes language to be secondary to their visual aspects. McCloud (1994, p. 9) is particularly definitive in his dismissal of language, defining comics as “(j)uxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer,” and asserting that language text is an ‘iconic’ element that can be treated as part of comics’ visual repertoire. While not as radical, Groensteen (2007) argues that the development of new visual-narrative forms has eclipsed that of the linguistic-narrative in comics. This is something that Ichirou Takeuchi might agree with, arguing that while story-manga are a mixed visual-linguistic media, visual elements are necessary precisely because of language’s “incompleteness” and its inability to describe the “essence” of things (2006, pp. 257-258). Others have in fact suggested that linguistic text is essentially ignored by readers in story-manga. Rommens (2000) explicitly argues that one of the supposed key methods of reading manga, speed-reading, “...is made possible through the elimination of textual information.”

This of course does not mean that there has been no work on the language of manga. Some recent examples of research on language in manga include Kinsui’s work on *yakuwari-go* (literally “role-language”, or “speech patterns which are aligned to the images of their speakers” (2003, p. 205)) in comics, where he stresses the function of stereotypical speech patterns not found in natural speech in character development, in respect to the issue raised above on tension in manga as works of fiction. Takahashi’s (2009) work touches upon Kinsui’s as she examines the role stereotypical female-language has in developing characters in the shoujo-manga *Life*, demonstrating that characters manipulate gendered speech patterns in accordance with the images they are trying to project. Ueno’s (2006) research on the usage of gendered sentence-final particles in female-language in comics for girls and women mentioned above also demonstrates that there are distinctly different patterns between differently aged characters as well as between individual series depending on the age of their readership. Inoue (2009), on the other hand, takes on a more specific aspect of language in manga, that of onomatopoeia, and how it is used and structured differently in manga as opposed to other linguistic environments.

However, while much of the recent research on language in manga offers interesting suggestions, the majority of such studies rely upon a smaller number of examples, usually focusing on a few select elements with limited attention to the larger context of such patterns. This makes it difficult to evaluate their arguments. At present, even basic questions such as what kinds of linguistic forms appear, how closely related language in

manga is to other literary forms or spoken language, and even more concrete questions such as the actual ratio of words to pages or the ratio of words to pictures have yet to be answered. The present situation, then, is not much different from over 20 years ago, when Kai (1989) posed several key points as being necessary to look at in future manga language research. These questions are not only pertinent to assessing how comics themselves function, but also to assessing the linguistic relationship they have with their readers. Without any real concept of what the characteristics of language in manga are to begin with, such questions may only really be answered with speculation.

### 1.1.3 Creating a manga corpus

This project, including the subsequent data analyses described in Chapter 2 to Chapter 4 was started in response to this lack of quantitative, empirical research on language in manga. While many different approaches were possible, I chose in particular to create a corpus—the term used here in the sense that is used in linguistics, that of being a body of text collected primarily for language research—of manga because corpora offer unique advantages when conducting quantitative linguistic research. Corpus linguistics is primarily a *method*, and is data-driven, meaning that it attempts to influence linguistic theory-making through the examination of patterns on a large-scale basis, rather than picking and choosing examples as per the conditions of the theories being considered. As Leech (2004) argues, this has the advantage of offering data to support—or contradict—theories, and so can be used in combination with theory-oriented linguistics. Others, as McEnery & Hardie (2012) describe in detail, stress the corpus as a complete source of all theory. Whatever position one takes, corpus linguistics is characterized by its stress of real, quantitative data. As a field, corpus linguistics is relatively new, started with major English language corpora such as the Brown corpus, which was first released in 1964 based upon data compiled in 1961 (Francis & Kucera, 1979). However, it has increasingly gained in popularity with the development of computers and other large-scale corpora databases, and today, it is a popular and widely adopted form of research, and is one of the sub-areas of linguistics that has been given the most attention in recent years. (For a good overview of the history and issues concerning corpus linguistics, please see Marzá (2009).)

While manga may seem not appear to be an intuitive choice for corpus-based research, such an approach might be considered useful for a number of reasons. Not only would it give clear answers to some basic questions such as the ratio of words to the page, it would also shed light on issues such as Kinsui's role-language or, like Ueno's research, the distribution of gendered speech patterns amongst different characters and series. Indeed,

Ingulsrud & Allen (2009, p. 81), in their important work on how manga are read, note that a corpus approach would be ideal for answering basic questions about manga such as the use of *kanji* (Chinese characters) and different types of vocabulary. Furthermore, such a corpus would also be usable by those interested in some of the specific kinds of texts found in manga—such as thoughts or onomatopoeia—for other, non-comics research.

In many ways, the development of a manga corpus could be said to be in-line with trends within linguistics. With the development of corpus-based linguistic research, the demand for corpora of non-traditional text sources has seen a sharp increase. Many have called for the expansion into visual mixed media as television scripts and advertisements, pointing to their importance in society and their rich contextual data (overviewed in Allwood (2008)). Yet there are many important practical and theoretical problems in designing such corpora, not least how to input the textual elements as meaningful linguistic data. As Bateman, Delin & Henschel (2002, p. 3) describe, while there have been a wide variety of tools developed for linguistics (=monomodal documents), “(f)or multimodal documents, however, such models and techniques are still largely lacking. A substantial set of problems is raised by the fact that the object of study is not linear, either temporally or in terms of the principles for its consumption; moreover, its multichannel nature makes it difficult to reconcile and peg together the methods of recording, transcription, analysis and annotation that have been developed separately for each mode.” The analysis of multimodal texts, then, requires the development of systematic methods of thinking about the text so as to give it the “order” that it currently lacks.

Indeed, with its linguistic data embedded directly into the drawings, manga are a true marriage of language with the visual. With text of inherently different natures being presented in a non-linear manner in the same page area, a crucial task in corpus design was to reanalyze the text types found within manga, such as lines (text representing conversational dialogue in speech bubbles; see Section 1.2 below) and onomatopoeia. It was also necessary to re-evaluate the relationship between such different types of text in order to give meaning to the abstracted texts, i.e., once they were taken out of their original embedded environments. Ultimately this is a text-oriented corpus, insofar as it is an attempt to deal with how to think about and link together non-linear text which is by nature a part of a larger visual structure. Yet the corpus described here could be said to be a part of the trend to try to link the visual and the linguistic. Furthermore, many of the issues encountered in its design are found with other media—newspaper stories, advertisements, political cartoons—which mix the visual with the linguistic, and the approaches taken here may help others with similar projects.

## 1.2 Corpus Design and Construction

### 1.2.1 Issues in creating a manga corpus

While the discussion above may have made clear how useful a manga corpus could be, it should be noted that there are many issues unique to a medium like manga that had to be dealt with in its construction. In particular, two issues—sampling and the proper representation of text—had to be dealt with before data could be collected, and they are both questions especially worthy of attention. First, in regard to the issue of how to choose an appropriate sample, with thousands of titles published annually, there is an enormous collection of texts to select from. While attending to genre might appear to be one possible option, genre in manga has been traditionally defined by market segments based primarily upon readers' sex and, to a lesser degree, their age, rather than the themes or styles of the texts themselves. These segments are rigorously defined by the magazines that manga series are serialized in, with individual magazines marketed towards different sexes. The legitimacy of such a genre scheme has come into question in recent years, however, with the breakdown of those markets. Young female readers in particular are increasingly reading titles from traditionally male-oriented magazines (Ogino (2001), Allen & Ingulsrud (2005)). Such changes in manga reading patterns have led some researchers, such as Itou (2005), to press for an expression-oriented definition of manga genre which would extend beyond market segments. As an expressive media wherein the author intends to convey some story, or more generally, information in the form of emotions, etc., to its readers, manga can be thought of as utilizing many different elements as tools to express that information. Those elements in manga include but are not limited to the type of story they tell, the style of drawings utilized and structural elements such as their use of frames and panels. Under an expression-oriented genre scheme, genres would be defined by their use of such elements, focusing on the content one actually sees rather than readership, which may reflect little in terms of common structural or narrative elements and language itself could of course be one such element. Until such attempts are made, however, the extent of the market and the ambiguity of genre lines are such that there is a lack of clear standards for what would be an appropriate general sample.

Second, in regard to representing the text, as Ingulsrud & Allen (2009) point out, language in manga is embedded directly in the images, making them inappropriate for text-scanning and thus requiring intensive data entry. A more important corollary, however, is that the non-linear relationship between textual elements demands that any linguistic corpus represent the relationship between different elements dynamically. Maekawa (2008) comments that manga should be included as a part of the large-scale written corpus that the National Japanese

Language Institute is producing, due to their importance as a symbol of modern Japanese culture and as a genre which often motivates non-Japanese speakers to begin studying the language. Yet how to deal with the linguistic data remains a problem. Some previous attempts at a manga corpus, such as Tokuhisa, Murakami & Ikehara (2006), who created a manga corpus tagged for facial expressions, have eliminated this concern by selecting only particular elements like lines. But this fails to capture the interdependent relationship that different elements, such as onomatopoeic expressions, notes and narration, have in creating manga, and which would theoretically be a major point of interest. Some categorical system for 1) distinguishing between the different types of linguistic text and 2) tagging where they appeared in the original manga is essential for properly representing manga's unique linguistic data in any kind of meaningful way.

Before looking in more detail at such questions, to summarize the responses described below, in regard to the first problem, I limited the corpus to series popular with young people, the group said to be influenced by manga. This was done in a two-part process, first by creating a list of top-selling series, and then comparing it with the results from surveys of high school students' comic reading habits. Ten series were selected, with an even number of young girl-oriented and young boy-oriented titles. On the other hand, in order to effectively represent the text, two sets of tags—one including contextual information such as the series and page number, and one including categorical information, noting information such as the type of text and the sex and name of the speaker—were designed and assigned to all elements. For categorical tags, all text was further determined to be in one of eight categories distinguished primarily by their presentation in the text: *Lines*, *Thoughts*, *Narration*, *Onomatopoeia*, *Background Text*, *Background Lines/Thoughts*, *Comments* and *Titles*.

### 1.2.2 Sampling issues

As summarized above, in response to the problem of selecting titles from the mass number of possible series, I first restricted the sample to those current or recently ended series popular with young people at the time of the original sampling (summer 2007 to summer 2008). As the previous survey results by the Agency for Cultural Affairs (2010) suggest, young people are the group that manga are assumed to have some influence over, and by restricting the sample to titles popular with said group, it may be possible to examine how such readers and manga interact. It must be noted that while genre is not a fully reliable marker in manga, it is also possible that the language patterns that may be discovered through a corpus could lead to more dynamic ways of thinking about genre, as has been found for literature. With the massive quantity of manga published each year, an

important starting point for considering genre in manga could be to examine what, if any, actual common elements may be found in the currently-accepted market-segment genres. Thus, while it may seem somewhat counterintuitive given the discussion above, and while holding some reservations and concerns as to their applicability, I selected equal numbers of titles from the most commonly accepted genre groups, shoujo-manga and shounen-manga, as a starting point to examine whether or not there are any salient characteristics to be found within the two. It should be noted that genre in this case was determined by the market-segment of the magazines running the titles, generally following the Japanese Magazine Publishers Association's (2012) groupings. The only exception was *Nodame Cantabile*, whose publishing magazine, *Kiss*, is listed as a being for women by the Japanese Magazine Publisher Association, but which has generally been received as a shoujo-manga series, winning awards such as the Kodansha shoujo-manga award in 2004. The corpus, in this sense, is not a full representation of the market, and does not have that as its aim; rather, it is a sample of male- and female-oriented series popular amongst high school students, and should be interpreted as such.

In order to establish what series were popular amongst young people, I conducted two surveys on high school students' manga reading habits. The surveys in question were conducted in July 2007 at a private all-girls school near downtown Tokyo, with 125 female respondents and at two co-ed public schools in western Tokyo in July 2008 with 148 respondents (73 male, 75 female). The survey itself consisted of seven questions, and was focused on their manga reading habits. Specifically, they asked what comic magazines they had ever read; how often do they read comic magazines; do they buy comic books (non-magazines); had they ever talked to their friends about comics; had they bought any non-comic/non-academic books recently; their favorite series; and spending money. The survey had multiple-choice selections available, with space to write additional comments. However, the structure of the survey varied slightly between the earlier and the later surveys. In the first survey, there were three additional questions, asking their favorite celebrity, television show and class; the question on spending money asked them how much allowance they received per month. In the later surveys, those three questions were removed, and the question on spending money asked how much money they used and on what.

In conducting the survey, the number of female respondents was kept artificially high to balance the limited reading patterns of young men: As mentioned earlier, young women tend to read a variety of titles across market segmented genres, whereas boys tend to read exclusively series aimed at boys or older men (Ogino (2001, pp. 141-146); Allen & Ingulsrud (2005, pp. 275-276)). The surveys in fact reflect this: Few boys mentioned reading any comic magazines for girls, and did not give girls-series as their favorite title, but many female students had read and chose as their favorite series titles for boys. Specifically, at the two co-ed schools in eastern Tokyo, out of the 44 boys who said they had a favorite series, only one boy gave a shoujo-manga series. In comparison, of the 55 girls who said they had a favorite title, 10 of them gave a shounen-manga title, and 16 a

shoujo-manga title. This trend of girls reading titles for both boys and girls but for boys to only read titles for boys is even more obvious when asking what magazines they had ever read: Of the 64 boys who had read comic magazines before, seven had read ones aimed for girls, whereas of the 72 comic magazine reading girls, 44 had read series for boys.

Using students' responses to what their favorite manga title was, I created a list of favorite series noted by multiple students. While there were a wide variety of answers, the majority of which were only raised by one or two students, some patterns were seen, and titles selected eight or more times over both surveys were added to the candidate pool. As a point of restriction, I omitted a limited number of older series from the list which, while 'classic', were no longer current at the time of the survey (e.g., *Dragon Ball*, which ended over ten years previously).

The results of the survey were then compared with a list of top-ten series from the approximately one-year period of July, 2007 to July, 2008. As Nakano (2009) describes, it can be extremely difficult to get reliable sales figures on manga, because detailed sales ranking for manga are not fully released by publishers themselves. Data instead comes in from various sources such as distributors and bookstores, and as such, used on their own, sales ranking can be difficult to assess appropriately. They can, however, be thought to be demonstrative of how much a series has permeated popular culture, if not a fully accurate description of which titles are being read by whom. Sales ranking was thus selected as a complementary tool to reassess the results of the surveys above in reflection of the overall market trends. The particular figures used here were those made public by PHP Interface, which announces weekly top ten rankings released by Touhan Shirabe, the research section of the major publication distributor Touhan (such data, including current and previous rankings, can be found on PHP's website, <http://www.php.co.jp/>). Final selection was done by comparing the candidate list from the surveys, and selecting those titles from the top of the list which also appeared on the top-ten list, after first dividing titles into shoujo-manga and shounen-manga.

Ten series were selected as a result of that process: from shoujo-manga, Tomoko Ninomiya's *Nodame Kantaabire* (*Nodame Cantabile*, below, *Nodame*), Ai Yazawa's *Nana*, Karuho Shiina's *Kimi ni Todoke* (*Reaching You*, below, *KimiTodo*), Aya Nakahara's *RabuKon* (*Lovely Complex*) and Yuki Tabata's *Bokura ga Ita* (*We Were There*, below, *Bokura*), and from shounen-manga, Gôshô Aoyama's *Meitantei Konan* (*Detective Conan*, below, *Conan*), Masashi Kishimoto's *Naruto*, Eiichiro Oda's *One Piece*, Tsugumi Ohba and Takeshi Obata's *Death Note* and Hideaki Sorachi's *GinTama* (Table 1). Note that while the series all began at different times, with some of the series already having been completed, they were all in publication at the time of the original surveys, and for an equal comparison the first three volumes of each series were chosen. Including the first three series secures similar data in the sense that first, we can anticipate that they all show both the initial setting-ups of the stories,

and second, they will also show some maturity, with the third volume usually having been published between two or three years after the introductory volume. In addition to being highly popular as comics, all of these series have either become animated television series or movies, live action movies, or novels, or have had video games and other related products produced. They can be thus assumed to have been well absorbed into popular culture.

It should be noted in passing that for many manga researchers, selecting individual titles as the sample pool rather than magazines may seem unnatural. Many, such as Ogino (2001), claim that serial magazines were the original core of the manga market, but as Nakano (2009) points out, entering the 1980s, most of the single-shot titles disappeared from the major magazines, only to be replaced by long-continuing serials. For such serials, republishing as tankoubon became a given. In reality, many magazines today are actually losing money, with publishing houses relying upon the tankoubon to make up the difference; the number of tankoubon published has in fact reportedly doubled in the last ten years—whereas magazine sales have gone down (Nakano, 2009). It seems safe to think, then, that reading manga series individually is becoming the main reading habit, hence the focus here on series over magazines.

**Table 1: Series in corpus**

Genre	Japanese title	English title	Author	Magazine	Publisher	Start date	Finish date
Shojo-manga	<i>Bokura ga Ita</i>	<i>We Were There</i>	Obata, Yuki	BetsuKomi	Shougakukan	10/2002	2/2012
	<i>Kimi ni Todoke</i>	<i>From Me To You (Lit: Reaching You)</i>	Shiina, Karuho	Bessatsu Margaret	Shuueisha	9/2005	—
	<i>Nana</i>	<i>Nana</i>	Yazawa, Ai	Cookie	Shuueisha	10/1999	—
	<i>Nodame Cantabile</i>	<i>Nodame Cantabile</i>	Ninomiya, Tomoko	Kiss	Kodansha	7/2001	10/2009
	<i>Rabū ★ Kon</i>	<i>Love Com</i>	Nakahara, Aya	Bessatsu Margaret	Shuueisha	9/2001	12/2006
Shounen-manga	<i>Death Note</i>	<i>Death Note</i>	Ohba, Tsugumi; Obata, Takeshi	Shounen Jump	Shuueisha	12/2003	5/2006
	<i>GinTama</i>	<i>Gin Tama</i>	Sorachi, Hideaki	Shounen Jump	Shuueisha	12/2003	—
	<i>Meitantei Konan</i>	<i>Case Closed (Lit: Detective Conan)</i>	Aoyama, Gôshô	Shounen Sunday	Shougakukan	1/1994	—
	<i>Naruto</i>	<i>Naruto</i>	Kishimoto, Masashi	Shounen Jump	Shuueisha	11/1999	—
	<i>One Piece</i>	<i>One Piece</i>	Oda, Eiichiro	Shounen Jump	Shuueisha	8/1997	—

### 1.2.3 Structural design

As noted above, linguistic data is presented in comics in a number of different ways. For example, *Lines* may appear in speech bubbles, or on occasion be written directly on the background, whereas *Thoughts* may be seen in square bubbles, or on whited-out areas of the background. There is good reason to think that readers process and deal with such information differently. In a suggestive informal experiment, Allen & Ingulsrud (2007) had readers mark with a felt pen what they looked at on the page as they read manga, as a result of which they found



that the most common reading pattern was for readers to search out conversational lines, paying less attention to other types of text and only secondarily at facial expressions, etc., as opposed to Rommens's (2000) suggestion that the graphic data is more important. Whatever the case may be, clearly any useful corpus would need to distinguish between those cases.

In defining the categories, I took first and foremost their presentational differences into account, such as whether they were in speech bubbles, or on the background. Whether they were written by hand or typed was also taken into consideration, as per the hints of Fuse (2004). In some cases, however, presentational differences were not deemed sufficient to make accurate distinctions. In those limited cases, other information, such as the voice or tense of the utterance were taken into consideration, which I make note of in the discussion of the categories below. With those points in mind I determined a total of eight categories that all the text could be categorized into: *Lines*, *Thoughts*, *Narration*, *Onomatopoeia*, *Background Text*, *Background Lines/Thoughts*, *Comments* and *Titles*. In line with the points above, the definition of each category is as follows, with figures for unclear categories. Note that the order below forms the category number for each type tagged in the corpus (e.g., *Lines* = category 1, *Onomatopoeia* = category 4, etc.).

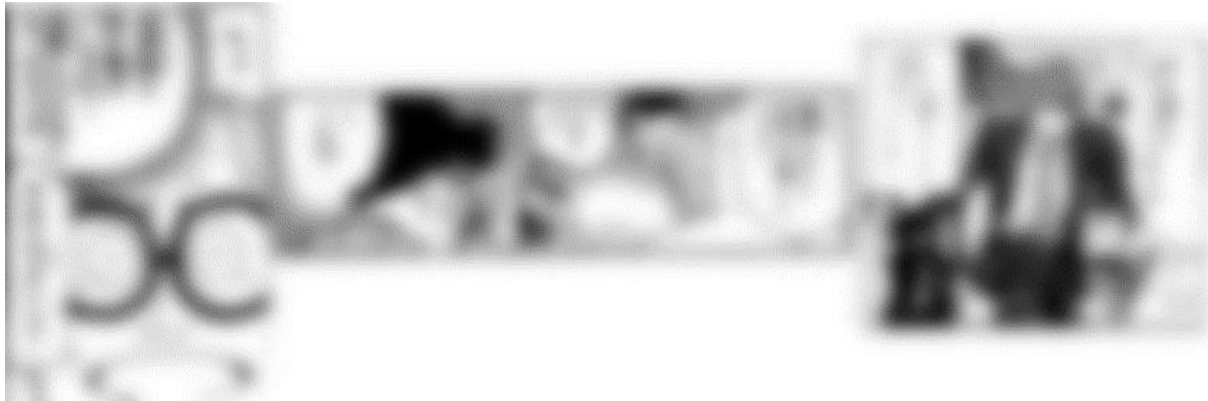
**1. *Lines*** (Figure 4, Figure 5, Figure 6): Any data found in unbroken speech bubbles representing audible information. Usually in type, *Lines* include anything from parts of conversation to radio announcements the characters hear. Their common characteristic, however, is that they are presumed to be audible within their manga worlds. Not all *Lines* are human or even voiced, however: *Lines* spoken by animals and onomatopoeia in speech bubbles are also grouped as *Lines*. As Groensteen (2007) describes in detail, the different shapes of speech bubbles are often assumed to have meaning, such as jagged edges representing anger, but here such differences will not be assessed given their comparative abstractness.

**2. *Thoughts*** (Figure 4, Figure 5, Figure 6): Any linguistic data representing characters' inner voices. They generally appear in type, in either speech bubbles with dots going towards the characters, or in squares or whited-out space in the background. As opposed to *Lines*, they are never said aloud, and represent characters' private thoughts. Put differently, other characters are not privileged to these utterances, and they do not directly address the reader.

**Figure 4 (Left):** Examples of *Lines* and *Thoughts* from *One Piece* (1:178)

**Figure 5 (Middle):** Examples of *Lines* and *Thoughts* from *Detective Conan* (2:91)

**Figure 6 (Right):** Examples of *Lines* and *Thoughts* from *Death Note* (3:62)

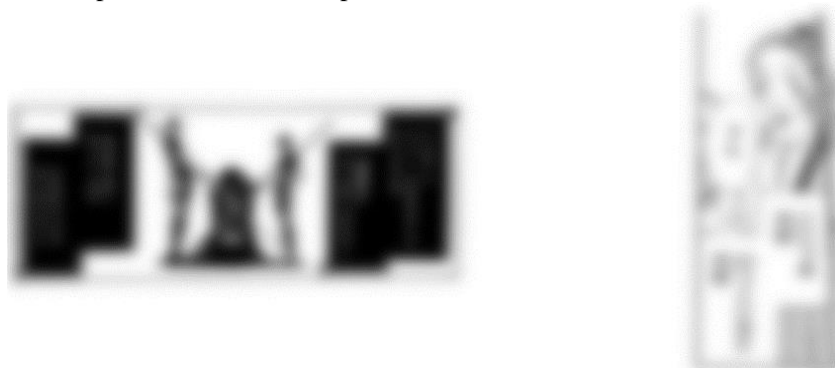


Here, Lines are in the unbroken speech bubbles; Thoughts are in the broken, whited-out areas, or in the third example, the one with the tail of circles.

**3. Narration** (Figure 7, Figure 8): Any data which supplies information on plot development, location, etc. directly to the readers. *Narration* usually appears in type and in square captions, and although it is commonly in the third person, it can be in any person. While its presentational style is similar to that of *Thoughts*, in opposition thereof, *Narration*'s primary role is descriptive, and it is often features privileged information that is unknown to the characters; it is also directed out towards an audience. Given their presentational similarities to *Thoughts*, linguistic hints, such as using a past tense, and who the utterance is being directed towards are sometimes used to distinguish between the two types.

**Figure 7 (Left):** Example of *Narration* in captions from *One Piece* (1:9)

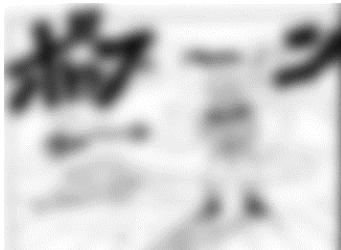
**Figure 8 (Right):** Example of *Narration* in captions from *RabuKon* (1:9)



4. *Onomatopoeia* (Figure 8, Figure 9): Any words which are written directly on the background to onomatopoeically describe the scene. As a general rule, onomatopoeic words do not form full sentences, and they are assumed to not be spoken by somebody, but rather mimetic of real world sounds or describing the nature or atmosphere of the scene.

**Figure 9 (Left):** Examples of *Onomatopoeia*, from *Naruto* (1:19)

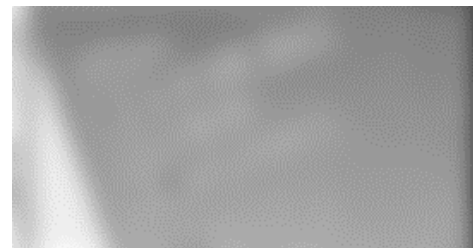
**Figure 10 (Right):** Examples of *Onomatopoeia*, from *Nodame Cantabile* (3:118)



5. *Background Text* (Figure 11, Figure 12): Any linguistic data that was written as a part of the scene in the comic which is not actually vocalized, such as advertisements written on billboards, building names, etc. Letters and other text that the characters read are also included as *Background Text*.

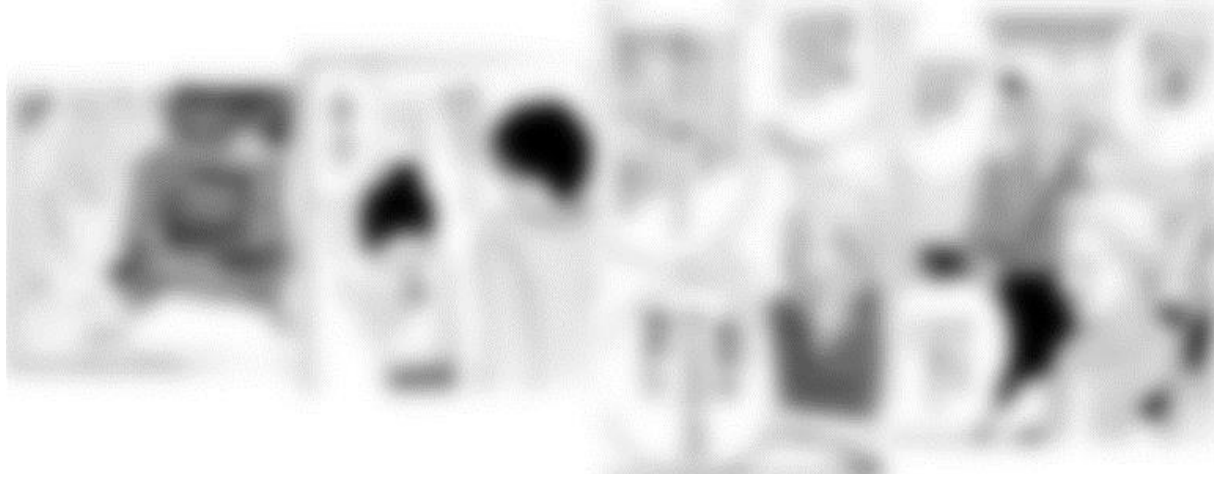
**Figure 11 (Left):** Examples of *Background Text* from *Bokura ga Ita* (3:25)

**Figure 12 (Right):** Examples of *Background Text* from *GinTama* (2:47)



6. *Background Lines/Thoughts* (Figure 13, Figure 14): Lines or thoughts which are written by hand directly on the background. Because they are not written in speech bubbles, it is impossible to tell whether they actually vocalized or not, and thus they must be differentiated from regular *Lines* or *Thoughts*. They are also usually written smaller than utterances in speech bubbles.

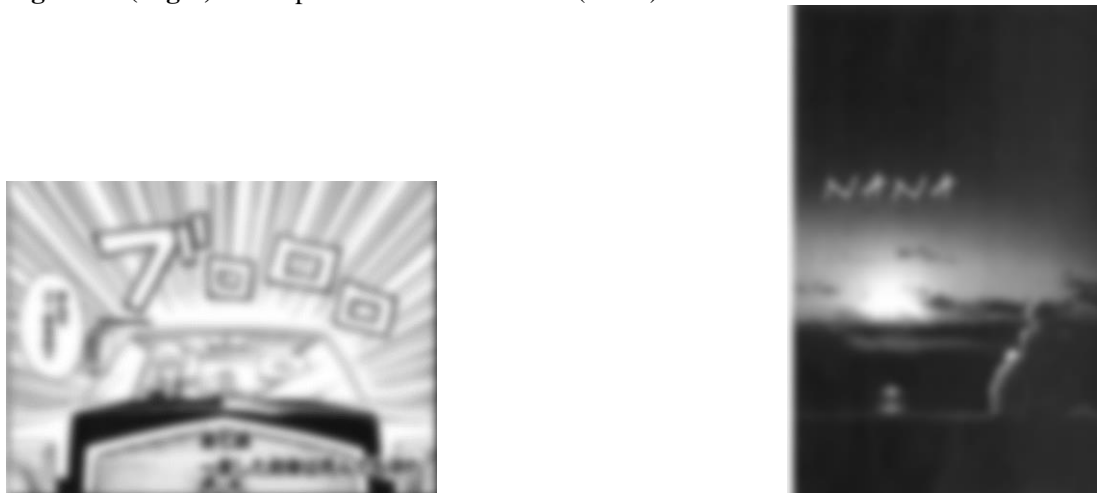
**Figure 13 (Far left):** Examples of *Background Lines/Thoughts* in *Naruto* (1:18)  
**Figure 14 (Mid-left):** Examples of *Background Lines/Thoughts* in *Kimi ni Todoke* (1:84)  
**Figure 15 (Mid-right):** Examples of *Comments* in *Nodame Cantabile* (3:128)  
**Figure 16 (Far right):** Examples of *Comments* in *Kimi ni Todoke* (1:134)



**7. Comments** (Figure 15, Figure 16): Notes or jokes which are written by hand directly on the background about the characters or the items appearing in the scene. Such *Comments* may or may not include arrows pointing at the items in question, but they are often full sentences, and supply privileged information about the scenes that has not been made available to the readers through the characters' *Lines*, *Thoughts* or *Narration*.

**8. Titles** (Figure 17, Figure 18): Data presenting the title, subtitle, author's name, etc., of the series.

**Figure 17 (Left):** Examples of *Titles* from *GinTama* (2:10)  
**Figure 18 (Right):** Examples of *Titles* from *Nana* (1:106)



All linguistic data was originally pure text in CSV form for ease of input, and have been organized using Microsoft Excel for its universality. Each individual block of text constitutes one entry in the corpus, and each individual entry is preceded by a set of tags noting where and how it appeared in the text. Respectively, the tags are for the title, volume, frame number (by logical reading order), category number, the occurrence number (as multiple entries of the same category may appear in the same frame), and the name and sex of the speaker (in the case of those categories which have a voice, i.e., *Lines*, *Thoughts*, *Narration*, *Background Lines/Thoughts* and *Comments*). Sex was selected out of five groups: male, female, both/unclear, onomatopoeia, and animal. By noting the category, page and frame number of each item in this way, it becomes possible to keep track of their occurrences in-context. It also, however, allows one to select just the individual category necessary—such as just *Lines* or just *Thoughts*—so as to analyze and compare the different categories as well. Below is an example of how the first four frames of one page from *Death Note* were input into the corpus (Example 1).

**Example 1:** Basic input style of text into corpus

Upper left: Target frames from *Death Note* (3:62). Upper right: Actual screen shot of data (includes more in-depth data). Bottom: Text-only version of target frames.



Title	Volume	Page	Frame	Category	Occurrence	Sex	Name	Text
DeathNote	3	62	1,2,1,1	月	1			なんとしても&あいつを殺してやりたいが&殺せば足がつく
DeathNote	3	62	2,1,1,1	デスノート	1			では&名前を書いた&その人間一人しか&殺せない
DeathNote	3	62	2,1,2,1	月	1			誰かを操って&殺させる事は&できない・・・
DeathNote	3	62	3,1,1,1	月	1			全く不便だよ&デスノートってやつは
DeathNote	3	62	3,4,1,0	0	0			キィ・・・
DeathNote	3	62	3,2,1,1	リューク	1			いつも冷静なライトが&逆切れか・・・&そうとう きてるな・・・
DeathNote	3	62	4,1,1,1	月	1			最初は名前さえ わかれば&事故死や自殺でいいと考えた・・・
DeathNote	3	62	4,1,2,1	月	1			しかし それは&あいつが100%L 本人だと&わからなければ駄目だ

## 1.3 Current State of the Corpus

### 1.3.1 Data summary

Table 2 below organizes data into how many characters, pages and entries are seen in total overall for each category, as well as by how many entries and characters appear on one page on average and the average length of entries by character. One entry here refers to one entry in the corpus; this is defined as any block of text, such as the text within a speech bubble or a string of onomatopoeic expressions, and does not correspond directly to sentences. Characters, on the other hand, refer to an individual ‘letter’ of text, including orthographic symbols and blank spaces. Overall, the 30 volumes included in the corpus consisted of 5,289 pages, or approximately a little more than 176 pages per volume on average.

**Table 2:** Overall data totals

Data	Genres	Series	Volumes	Pages	Characters	Entries
Total	2	10	30	5,289	688,341	55,480
Average/Pages	2,644.50	528.90	176.30	1.00	0.01	0.10
Average/Characters	344,170.50	68,834.10	22,944.70	130.15	1.00	12.41
Average/Entries	27,740.00	5,548.00	1,849.33	10.49	0.08	1.00

In total, there were 688,341 characters of text distributed over those pages, or 22,944.70 characters per individual volume. That text was dispersed over 55,480 entries in the corpus, with an average of 10.49 entries per page. Table 3 also shows the data summary for each individual category. While the number of entries is naturally always less than or equal to the number of characters as characters are what make up entries, the average number of characters per entry also differs from category to category, with *Narration* (15.92 characters per entry) the longest category on average and *Onomatopoeia* (3.69 characters per entry) the shortest category on average. Such differences point to the characteristics of the categories themselves: Where *Narration* is generally made up of full sentences, *Onomatopoeia* is individual words or strings of words. As a result, a gap is naturally expected here, and tells us about the different nature of the two categories. Looking next at the amount of text per page, there is an average of 130.15 characters per page, with the majority of those made up by *Lines* (94.57, or 72.63%). The average number of *Lines* per page is 6.48, making the average length of any given *Line* on a page 14.58

characters. After *Lines*, *Thoughts* were the category with the most characters per the page, or 17.49, but *Lines* clearly make the majority of linguistic information readers flipping through manga would find.

**Table 3:** Summary of data by categories, characters and entries

Genre	Data type	Category								
		Lines	Thoughts	Narration	Onomatopoeia	Background Text	Background Lines/ Thoughts	Comments	Titles	Grand total
Shoujo-manga	Characters	219,513	52,427	8,839	15,904	10,115	21,622	2,394	682	331,496
	Entries	16,239	4,285	582	4,164	977	2,310	342	73	28,972
	Entry Length	13.52	12.24	15.19	3.82	10.35	9.36	7.00	9.34	11.44
Shounen-manga	Characters	280,455	40,056	4,819	15,539	12,584	1,209	162	2,022	356,846
	Entries	18,052	2,276	276	4,367	1,212	142	20	163	26,508
	Entry Length	15.54	17.60	17.46	3.56	10.38	8.51	8.10	12.40	13.46
Total	Characters	499,968	92,483	13,658	31,443	22,699	22,831	2,556	2,704	688,342
	Entries	34,291	6,561	858	8,531	2,189	2,452	362	236	55,480
	Entry Length	14.58	14.10	15.92	3.69	10.37	9.31	7.06	11.46	12.41
	Chars./Page	94.57	17.49	2.59	5.94	4.29	4.32	0.48	0.51	130.20
	Entries/Page	6.48	1.24	0.16	1.61	0.41	0.46	0.07	0.04	10.49

*Characters: The number of "letters"; Entries: The number of blocks of text; Entry length: The average number of characters per entry; Chars./Page: Average number of characters per page; Entries/Page: Average number of characters per entries*

### 1.3.2 Potential uses and interests of the corpus

Given the underestimation of manga as linguistic texts and the general tendency to view them as visual works, the potential uses of this corpus may seem limited. Yet as I have hinted above, there are ample reasons for both manga specialists and non-specialists alike to be interested in such data. Of course, one must make certain decisions in order to use the data appropriately. In particular, one must decide what part of the corpus (i.e., which text types) to use and whether or not to look at entries or characters as the unit. In general, entries show how often that type of text appears, whereas characters tell how much data, i.e., the weight of a given category, can be found. Which statistics to utilize will depend on the purpose of the study. In a structural analysis of the types of texts found in manga entries may be more informative, whereas when using counts of certain words or grammatical structures, characters may be more appropriate. Note that some of these are points which are actually typical issues for corpus-based studies in Japanese because of the nature of Japanese text: As Japanese does not utilize spaces, defining the number of words in a text can be difficult, and corpora are often described by the number of characters in them instead (See Gotoo (2007) for more details on working with Japanese corpora). In either case, these points add to the corpus's versatility, and can be thought of as one of its interesting

points. While the rest of this thesis will look in detail at some questions more relevant to manga directly, here I shall demonstrate three other points of interest to a more general linguistics audience: the possible productivity of unique tags, the mass of specialized data, and the potential for comparing patterns in different linguistic environments.

First, there is the issue of tags. Given the basic structure/format described above, it should be comparatively easy to add additional notation for different projects. While this is the case for any corpus, the nature of manga is such that, given the drawings as references, one can create a wide variety of tags, from basic information such as the sex and name of the speaker/listener to much more complex ones, opening up wide potential fields of analysis. A good example of this would be the emotional tags created by Tokuhisa, Murakami & Ikehara (2006) for their corpus of the lines from a popular manga series. Using the expressions on the characters faces, they created a notation system that detailed the emotional status of characters. Not only could this be utilized in this corpus, but it could also be expanded on given the additional categories of text included here.

Secondly, the corpus also offers a large amount of specialized data. Of particular interest is the case of *Onomatopoeia*. One of the major characteristics of Japanese is its rich vocabulary of onomatopoeia, and manga actively use onomatopoeia in order to create a sense of action and atmosphere that would otherwise be static and lacking. Onomatopoeic words are written on the background (“ボッフ〜〜〜ン” (*boffu~~~n*), “フニユ〜〜〜ン” (*funyu~~~n*) in Figure 9) as a kind of special effect, and their abundance has been given varied explanations. Very common in manga, Pollman (2001, pp. 12-13) attributes stylized onomatopoeia to a “graphic” element of Japanese writing, whereas Schodt (1983) points towards how manga are read by quick scanning, and Peterson (2009) notes that the use of onomatopoeic expressions like this picks up on similar usages in older forms of story-telling. A perhaps more important point is that Japanese simply has more onomatopoeia than many other languages (Lee & Shaw (2006), Peterson (2009)).

Where *Onomatopoeia* are the second most commonly seen text type on any given page, however, they appear less frequently in actual speech, and collecting a variety of examples can be difficult. While such stylized *Onomatopoeia* appear in non-sentence forms in manga and thus cannot tell us about their usage in sentences, some researchers have already started to approach them both for the large quantity of data they present. Inoue (2009), for example, has explored how the patterns seen in manga fit into the larger phonological and syntactic structure of onomatopoeia. A corpus of manga that includes all of the linguistic data found in them would in itself represent a mini-corpus of onomatopoeia that would otherwise be difficult and time-consuming to collect. At 31,443 characters, the *Onomatopoeia* category represents a small portion of the overall data, but a large amount of onomatopoeic data. One promising line of research might be to examine how well onomatopoeia in



manga conform to the phonological structures described elsewhere, which a corpus would enable to do efficiently through searching for structural patterns: Onomatopoeia in Japanese typically follow a limited number of well-described phonological structural patterns, such as the repetition of a basic morpheme. (See Inoue (2009) for a review of these structures with particular light on how they appear in manga.)

*Onomatopoeia* is not the only category to form an interesting mini-corpus, however. Categories such as *Lines* also offer many opportunities to researchers. While fictional in nature, *Lines* in manga form the majority of the text in the corpus (499,968) and offer large quantities of simulated conversational spoken language. Lines from manga have in fact been used by some researchers such as Maynard (2000) on Japanese discourse for precisely those reasons, and a corpus such as this may be of interest to anyone researching Japanese speech. This point, of course, could be extended to the other categories as well, particularly *Thoughts* and *Narration*.

Thirdly, the categorical tags make it easy to collect and compare data from different contexts. With all text grouped into one of the above eight categories, an inherent part of the structure of the corpus, and manga itself, is such that it is comparatively easy to distinguish between different ‘narrative’ environments. Whereas in a novel one would have to look for markers such as quotations or sometimes difficult to ascertain linguistic elements (e.g., verbs of thought vs. verbs of narrative) to distinguish between narratives, spoken lines and inner voices, for example, in a corpus of manga all of these data are clearly divided into easily distinguished categories. This makes it extremely easy and viable to compare the contents of *Thoughts* versus *Lines*, or what appears in secondary environments such as *Background Text* versus more narratively-central conversational *Lines*. One possible point to consider would be whether characters are consistent in their use of gendered speech patterns in *Lines* and *Thoughts* in order to assess the role of having a listener in linguistic selection, but a wide variety of research topics are possible.

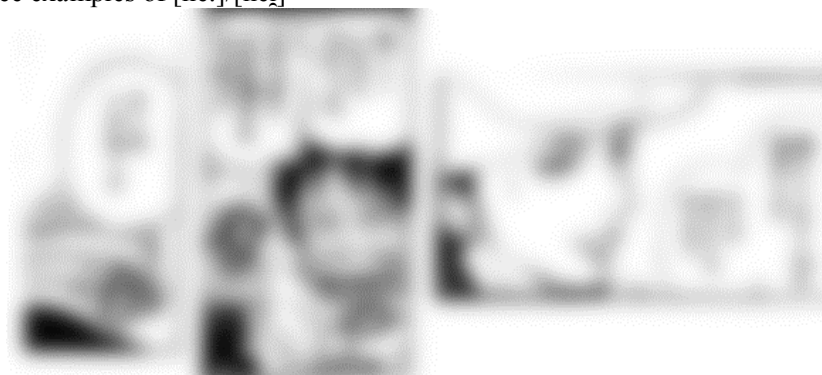
### 1.3.3 Unique issues

While the data has many interesting potential uses, the special structure of text in manga has meant that there are still some issues which must be properly assessed in order to use the corpus to its full potential. Three particularly critical problematic points are noted here: the use of nonstandard orthography; unusual structures to sentences; and comparing the data within the corpus with other text types. While the earlier two points primarily have to do with trying to effectively search and parse the text, the latter point is concerned with how to locate the linguistic characteristics found within the corpus within a greater frame of reference.

To begin, the use of non-standard orthographic conventions in manga makes it difficult to break down the text with morpheme parsers. This problem is first characterized by manga's distinct overuse of unusual punctuation. Approximately 22% of all the text is actually orthographic symbols, including but not limited to the question mark (“?”), the exclamation mark (“!”) and white space (“ ”). (All symbols intermixed in Japanese text were input as full-width characters, hence the use of full-width symbols over the more orthodox half-width symbols found in English language texts.) This is well above the 15.58% average found in newspapers (Nozaki & Shimizu, 2000) or the 17.6% found in magazines (Kashino, 2008). Note that Kashino includes numbers with symbols, so the number of orthographic symbols only should actually be considerably lower, making the percentage difference between them and manga higher. Likewise, many of the symbols found are in fact *not* commonly used in other texts—in particular white space between words—whereas some common orthographic symbols, in particular the period (*kuten*, “。”) and the comma (*touten*, “、”) are found very rarely.

The second part of the problem is that authors often use non-standard choices in writing out individual words. Japanese normally utilizes three systems of writing: *kanji* (Chinese characters), *hiragana* (a syllabary used primarily for function words) and *katakana* (a syllabary used primarily for foreign loan words). While most words are written in a more or less accepted way, they are less certain for non-standard spoken language, and it is not uncommon to see great variation and creativity in script choice in manga. For example, in male spoken speech, the negative ending marker, *nai* [nai], is sometimes pronounced [ne:] or [nej]. Normally *nai* would be written in hiragana as “ない”, but the masculine form, which is commonly seen in manga, is primarily a spoken form and is not as standardized. Thus, one sees versions of [ne:] and [nej] written in a variety of combinations of hiragana and katakana, such as “ねえ,” “ねえ,” “ねー,” “ねエ,” “ネエ,” “ネー,” and “ネエ” (Example 2).

**Example 2:** Three examples of [ne:]/[nei]



Left: *Naruto*, 3:31

Naruto,3,31,3,1,1,1,1,そーは&いかねーよ

Middle: *One Piece*, 1:10

One Piece,1,10,3,1,2,1,うそつけ！！&バカな事すん&じゃねエ！！

Right: *Nana*, 2:93

Nana,2,93,3,1,1,1,確かにこれで7万は&激安だよな&&なんか&出るんじゃ&ねえの？

(Emphasis added by author)

These related points both ultimately affect how easily the text can be effectively searched and parsed. In relation to the former point, Groensteen (2007) notes something similar in some Western comics, suggesting that the overuse of certain kinds of punctuation is meant to give the texts a sensation of orality. Whatever the case may be, without a better understanding of how punctuation is being used in manga, the overuse of punctuation marks, some of which are not featured in their expected environments, makes it difficult to appropriately parse text into smaller sentences and units. Likewise, the variability in spelling conventions means that not only are none of the more unusual patterns recognizable by morpheme parsers, which are currently oriented towards more conservatively written texts, but that words that are actually derivable to the same base may not be associated with the others. This freedom of script choice is one of great importance to Japanese morpheme parsing in general, but it is one that is especially relevant for a spoken-speech oriented corpus such as this. These issues are specifically brought up again in Chapter 3, where I look at the orthographic make-up of manga, and describe one attempt at dealing with those characteristics in conducting a morpheme analysis of the vocabulary.

Secondly, the special structure of linguistic data in manga can also affect their usefulness. Utterances are restricted by the size of their environment, whether that be a speech bubble or directly on the background, and their presentation is greatly affected by this. A good example of this would be line breaks. In order to properly represent line breaks within the same utterance, I have utilized the half-width ampersand (“&”), which otherwise does not appear within any of the texts. While I have omitted the ampersand from the character data listed earlier

since it is a place marker, if included, it would come to close to 10% of all text (75,824 out of 764,411 characters). Thus, individual sentences, even when in the same bubble (= linguistic environment), are divided into much smaller strings. It is also not uncommon to find sentences broken up into multiple utterances, thereby making each individual entry only fragments without the others. As one limited option, it is possible to remove the ampersand for analysis purposes in the case of line breaks, to create undisturbed text strings. This is less viable for utterances which have been made into separate entries because they were divided into separate blocks in the original manga, but it may be possible in the future to connect continuing entries by adding a name-tag to identify the speaker in combination with the occurrence number, to determine the order of utterances. This is a point I will bring up again in Chapter 3, where I look at the number of sentences in *Lines* as based upon the number of predicates and their forms of conjugation.

Finally, an important task will be to determine just how unique the patterns found in the corpus are to manga, and, in that respect, how manga compare as a linguistic medium to other types of texts. For at least some categories, such as onomatopoeic expressions, it seems that the patterns found in manga differ in some key ways from those seen in other types of texts (Inoue (2009)). This not only affects one's understanding of how language functions in manga, but also changes the viability of using manga as a resource for more generalized research. At this point, however, it is also unclear to what degree the patterns in manga are unique, and it is difficult to determine what would be appropriate parallel forms to look at, e.g., novels? plays? or animation? These are issues which still need to be addressed, but ties-in with other major industries, so-called 'light novels' (novellas based upon manga) or animated television series and movies of the series used in this corpus are some potential resources to look at. In general, this thesis has by-stepped these issues by using comparisons with the data from other resources, and, in large part, because the goal has been to assess first and foremost the characteristics of manga itself, and was not aimed to be a form of comparative research. However, this is a question which I will bring up from time to time within the text itself.

## 1.4 Overview of the following chapters

At its heart, this thesis is an attempt to examine the role and nature of language in manga, thereby adding to the understanding of how manga functions as a medium and how language affects how it is read and its readers. One major theme that will repeat throughout the thesis is the question of genre in manga. As has been described previously, genre in manga has increasingly been seen as an issue that can no longer be taken for granted, and I will argue here that language may be one way of reassessing genre in manga. The linguistic characteristics of texts have proven to be fruitful resources for genre identification in a variety of languages and media (e.g., Stamatatos, Fakotakis & Kokkinakis (2000); Kessler, Numberg & Schütze (1997); Biber (1993); Henry & Roseberry (2001)), and it would not be unreasonable to think that these would be valid for manga as well. As I will show, there appear to be many points on which shoujo-manga and shounen-manga differ, from their use of text types (Chapter 2) and vocabulary (Chapter 3) to characters' language patterns (Chapter 4).

Another theme of this thesis will be the question of what kind of a reading experience manga offers. While other researchers such as Ingulsrud & Allen (2009) and Nakazawa (2005) have noted the special skills—socially, in the earlier, and cognitively, in the latter—required to read manga, most of these questions have not dealt with the linguistic content of manga themselves. I would argue that looking at the linguistic content of manga could offer invaluable input on the question of *manga literacy*, as it has been dubbed, both from the question of the direct linguistic skills involved (what kind of vocabulary is seen, and how does that compare with their target age groups?—one of the themes of Chapter 3), but also from some of the even less clear points, such as how language is used to create characters that are recognizable as fitting certain molds (e.g., the question of stereotyped speech patterns and yakuwari-go, one of the themes of Chapter 4). While keeping the problematic points described in (1.3.3) in mind, the following chapters approach these uniting themes by looking at individual aspects of language in manga.

Finally, as one way of describing how the data collected and described here could be utilized outside of manga studies themselves, I will touch upon the possible impact of the language patterns seen in manga on their usefulness as resources for students of Japanese as a second language. As I mentioned above, manga are increasingly being seen as a way of connecting with students in the foreign language classroom, motivating them by using media that they—or at least, many of them—are interested in. Yet, with the lack of quantitative data on the linguistic characteristics of manga, it is not always clear how they can be utilized effectively. This will largely be the focus on my conclusions in Chapter 5, and I will give in particular the character's usages of speech patterns and manga's unusual orthographic qualities as issues to be aware of.

In dealing with these themes, I will look at language in manga at three, interrelated levels. First, I will look at the *structural* characteristics of manga through an examination of the distribution of the text types that were described above (Chapter 2). I will particularly look at how two text types—*Background Lines/Thoughts* and *Comments*—seem to differ in usage in shoujo-manga and shounen-manga, arguing that they are the result of different reading experiences the two genre offer. Then, I will move to the *orthographic* and *lexical* characteristics of manga, particularly those found in *Lines* (Chapter 3). To do so, I will show the results from three studies on the use of scripts and orthographic symbols; the readability of manga; and a morpheme analysis. The results of these studies will help to understand how manga is located within a spectrum of different texts, showing that it has many similarities with other new media as well as to the spoken speech it emulates. Finally, I will discuss how language is specifically manipulated to create characters, focusing in on the use of gendered language in particular (Chapter 4). By looking at how characters use personal pronouns and sentence final expressions, I will show that there are shounen-manga uses more stereotypical speech patterns, particularly in regards to female characters. I will argue that these differences are caused not only by author interference, but also by the roles of female characters in the text and the interpersonal distance between them and other characters. In this way, the discussion will move from more general linguistic characteristics to those more in-depth, creating a fluid transition towards more content-oriented analyses.

This thesis is built up of the research conducted since I first entered the Master's program in 2008, up until now, in January, 2013, and a large portion of the text itself is made up of previously published articles and presentations. In particular, the introduction to the corpus in Chapter 1 is based upon Unser-Schutz (Developing a text-based corpus of the language of Japanese comics (manga), 2011), and the discussion of text types in Chapter 2 is based upon (Language as the visual: Exploring the intersection of linguistic and visual language in manga, 2011) and (Exploring the role of language in manga: Text types, their usages and their distributions, 2010). The analysis of text characters in Chapter 3 was first introduced in (Kakikotoba toshite miru manga to sono hyouki-jou no tokuchou, 2012), and part of the morpheme analysis was previously introduced in (Reconsidering visual context through comics, 2012). Chapter 4 is largely based upon (Speaking out of character: Consistency in yakuwari-go 'role-playing' speech in a corpus of Japanese comics, 2012), (Balancing the scales: Analyzing the distribution of text amongst speakers in popular Japanese comics for girls and boys, 2011), and (Personal pronouns and gendered speech in popular manga (Japanese Comics), 2010). Chapter 5 includes large parts of (Reconsidering visual context through comics, 2012) and (Manga as a linguistic resource for learning, 2011). A full list of the presentations and articles which I have presented on these topics may be found in Appendix 6.3.

Note in closing that the following guidelines will be followed when dealing with certain potentially confusing orthographic and presentational issues.

Japanese words which are used repeatedly throughout the text, such as ‘manga’ or ‘hiragana’, will be written in italics the first time, and then in plain text following that. Japanese names will follow the western order of first name first, last name last, to maintain consistency with the other names used within the text. Romanization will generally follow the Hepburn system, with long vowels written out. Exceptions are made in cases where other Romanizations are more common or accepted.

All translations from Japanese into English—as well as any mistakes to be found therein—were by myself, except where otherwise noted.

When noting the page and volume numbers for images and other references to manga themselves, the basic key will be volume:page, so that the citation 3:62 will refer to *volume 3, page 62*.

In referencing the formal categories described within this chapter, I will always use italics and capitalize the first letter. Thus, *Lines* refers to the category of text within speech bubbles, whereas lines would not be used with this specific connotation.

A sample of texts, including actual drawings, the corpus data and the morpheme analysis data, are included in the appendix.

Finally, a gloss has been included in the appendix to reference Japanese words that may be unfamiliar to readers.

I would also like to add my deepest appreciation to all of those who have helped me along the process, from my long-time advisor Professor Yumi Nakajima and my parents, Margaret and Randolph Schutz, to my friends, colleagues and all the other people along the way who have offered advice, comments, and an ear (or two).

## 2. The Structural Characteristics of Language in Manga

### 2.1 Chapter themes and overview

#### 2.1.1 Introduction

In this first chapter of analysis, I will be looking at the role that different types of texts—the categories described previously in Chapter 1—have within manga. The variation of text types within manga, I will argue, is one of the key characteristics differentiating manga from other genres, and its analysis is essential for understanding the structural roles of text in manga. As I will discuss below in detail, language in manga is not static, but is dynamically non-linear in presentation, allowing for different forms to express a variety of information. This may be found in forms such as lines through speech bubbles, narration in square captions, onomatopoeia embedded directly in drawings, etc., and they each provide something different to manga and change the reading experience accordingly. There is good reason to think that text types are treated by readers differently. In Allen and Ingulrud's (2007) previously mentioned experiment, it was found that while readers searched out for lines, they often skipped other types of text like narration, in this sense supporting Rommens's (2000) assertion that narration does not play an important role in manga. This suggests that an important part of analyzing how manga functions may be to examine the role of different text types, yet most previous research on linguistic aspects in manga have tended to focus in on one set category such as *Onomatopoeia*, or to look at them anecdotally, without detailed distributional data.

To truly understand the role that *linguistic* language—in the sense of spoken language, as opposed to *visual* language—plays in manga it is necessary, then, to take a step back and consider its inherent interrelationship with visual structures more comprehensively. As Nakamura (2006) discusses, the complex relationship between linguistic language and visual language in manga has actually resulted in a lack of thorough attention to how linguistic language functions in manga, as there has been a tendency to assume that because language must be mediated by visual structures, it is always a secondary element subsumed to the visual. However, for precisely that reason it is essential that this relationship be reviewed, and I would argue that by looking at the way that linguistic language and visual language interact in manga, we might obtain a better idea of how they both function together to form a complete work. This chapter works to deal with these most basic



of questions, e.g., how much of the different text types are really seen in manga, and are these consistent across genres and series by examining the distribution of the different text types within the corpus. I anticipate in particular that the quantitative data offered here could give insight into language in manga that is not immediately obvious, as well as help establish the verity of claims previously made. The goal, as such, is to shed light on the way that language is distributed in the corpus and to give an analysis of how to think about the meaning of those distributions.

In proceeding with this examination, I will specifically look at two special types of text in manga, handwritten lines and thoughts (which I will consider to be one text type), and authorial comments. While the role of visual structures has been well covered for major categories similar to what I have defined as *Lines* and *Thoughts*, such as in Groensteen (2007), language in manga actually appears in many other forms which have not been as well described. Handwritten lines, thoughts and authorial comments in particular have been given more limited attention in previous research on language in manga, and they may offer a unique chance to go back to the most basic point of the meaning of the visual presentation of text—i.e., *how and why* language in manga is visually encoded in these different ways. They may prove to be of interest in at least two regards. First, they have many similarities with the more orthodox categories of *Lines*, *Thoughts* and *Narration* which appear in well-established environments like speech bubbles or captions, demanding one to consider why authors would choose to make such distinctions. Second, variations in their distributions across series suggest that they are a point of difference between genres: While infrequent in general, handwritten lines, thoughts and authorial comments are found with regularity in shoujo-manga, but very infrequently in shounen-manga.

This examination will proceed as follows. First, I will describe the nature of text in manga and comics, arguing that one of the key characteristics of text in manga as compared with other textual genres is not simply their visual aspects, in so far as they are integrated into the drawing, but their non-linearity, which makes unique demands upon readers. Next, I will go over the broad characteristics of the different categories, looking particularly at how they differ between the two major genres of shoujo-manga and shounen-manga. In identifying the distribution of text within the different types, it will become obvious that two text types, *Background Lines/Thoughts* and *Comments*, the former corresponding with handwritten lines and thoughts and the latter with handwritten authorial comments, vary by genre, which is reminiscent of the genre-differentiation seen in the use of highly visual monologue-like texts which characterized the so-called psychological 1970s/1980s shoujo-manga, styles which eventually fell out of popularity (Ôtsuka, 1994). In interpreting their meaning, I will consider how *Background Lines/Thoughts* and *Comments* pick-up and differ from these styles, eventually suggesting that *Background Lines/Thoughts* and *Comments* might function similarly to Takeuchi's (2005) mediator and spectator characters. As styles that seem to emerge from a new authorial presence in shoujo-manga

magazines such as described by Nanba (2001) I will also argue that these types of handwritten text offer a sense of personality and closeness while also functioning to visually-code the data in terms of relevance, thereby helping readers search for information more efficiently while also allowing interested readers to feel a sense of connection with authors. Through this analysis, I will suggest, finally, that using different graphic formats for text allows authors to switch between styles and create dynamic atmospheres. In this way, I hope to lend clarity not only to language's role in manga, but also to the interaction of language with the structures which make up manga's visual language.

## **2.1.2 Text in comics and manga**

While the most obvious difference in reading comics from more traditional forms of text is the negotiation between drawings and texts, one thing that any reader will have experienced feeling is a certain sense of uncertainty at times as to where they should look next. With some major exceptions such as visual poetry—defined by Bohn (2001, p. 15) “. . . as poetry that is meant to be seen,” and which places words within in the page in a way which prioritizes visual effect—it is the rare literary work that does not continue on in the accepted movement of text for whatever language it might be. A work of English will take for granted that the next word will come to the right or below the last; although there may be a double-carriage return between them, the next paragraph will come below the first—and if not, then surely on the page to the right. Sometimes these rules may be complicated, such as in Japanese, which is identical in order to English for horizontal text, and left-to-right, top-to-bottom for horizontal text; yet as soon as this is determined, the text will remain consistent, at least for that section. While there are on occasion times in when one sees volumes which mix vertical and horizontal text in Japanese, the same basic principle stands: The vertical texts are compiled together on one end of the book, opening from the left, and the horizontal texts together on the other end, opening from the right. When one reaches the middle, the book may be turned over and started again from the other side. There is no experience of moving around throughout the page: It is a *linear* experience, where one moves throughout the text following a set course that is determined almost immediately with the appearance of the first word on the paper. Comics, however, break this golden rule, as text is embedded into the drawings themselves, partitioned into blocks of text that are interspersed through the pages and frames.

While comics still follow some basic principles of reading in terms of how to move through pages and frames in a mannered order—this order being paramount above all else, at least according to Scott McCloud,

who put the sequentiality of images as the definition of comics (1994)—text within a given frame does not necessarily follow that path; it rather requires readers to make judgments about orders based upon what they determine to be a viable narrative. An example might be found in Figure 19, from Tomoko Ninomiya’s *Nodame Cantabile*. Here, the main female character, Nodame, and a secondary character, Masumi, have been competing to get the main male character, Chiaki, to go on a date with them on Christmas. Both characters are very upset at having not been able to get Chiaki to go out with them; given what Groensteen (2007) and others have written about the shapes of speech balloons, we can assume that this is being expressed not only by their having their heads in their arms, but by the use of jagged-edged speech balloons. A disadvantage of jagged-edged speech balloons, however, is that they generally do not feature tails to demonstrate who is speaking. As a result of this, there is some ambiguity in terms of assigning the bubbles to speakers. Assuming that the speech bubbles are assignable to the speakers to whom they are closest, the right most two would go to Nodame and the left most two to Masumi, and the content of the speech bubbles themselves seem to correspond with this interpretation, as does the use of language within the two frames, with the one on the top right calling Chiaki *sempai* (senior/upperclassman; see glossary) as Nodame generally does, and the top left using the strongly feminine sentence-final particles *no-yo* that Masumi generally does (gendered speech and character development will be brought up again in Chapter 4). However, what remains questionable is the order in which these should be read. Clearly, the content of the two lower bubbles must be read *after* the corresponding upper bubbles, as they are post-position, inverted subordinate clauses. That is to say, they are dependent sentences which normally would appear in the beginning of the sentence, but were tagged on afterwards, a common phenomenon in Japanese (Eguchi, 2000). Here, the use of *no-ni* (even though) in Nodame’s second sentence clearly tags its relationship to the sentence above it. Masumi’s second sentence also ends in *no-ni*, and parallels Nodame’s structurally. Yet we might ask the question whether it should be Nodame-Nodame-Masumi-Masumi, or should it be Nodame-Masumi-Nodame-Masumi. Both seem viable options, especially given the structural parallels in the language. It would even not be impossible to think that theses were not post-position subordinate sentences at all, if one accepted reading from the bottom up.

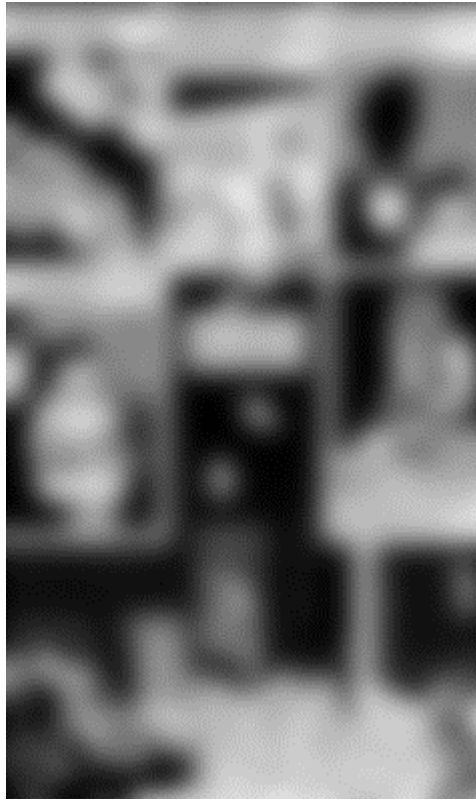
**Figure 19:** Example of speech bubbles where it is difficult to determine the speaker and natural order in *Nodame Cantabile* (2:47)



Nodame:	Doushite & senpai ha kyu ni & taido wo hirugaeshita & n deshoukaa <i>Why did Chiaki suddenly change his attitude?</i>
Masumi:	Doushite & watashi ga & taijou saserareru & no yo---!! <i>Why am I forced out?</i>
Nodame:	Yakusoku shita & no ni! <i>Even though (he) promised!</i>
Masumi:	Ai wo hyougen & shita dake <i>I was just expressing my love!</i>

Under some circumstances, readers may also find themselves moving back into a previous frame to access the linguistic text necessary to string together the story. In an example of the universality of such issues for all comics, one may look at Figure 20 from Karasik and Mazzucchelli's 2004 adaptation of Paul Auster's *City of Glass* (Auster, Karasik, & Mazzucchelli, 2004). Looking at the overall page, we see that it has been broken into two panels with six frames in total, three on top, two in the middle, and one on the bottom. However, the bottom frame is peculiar in that it is shaped like a pyramid, with its top breaking up the two frames in the middle panel. Visually, too, the top part seems to almost form a third frame, with the background design gradated so that it only starts to blend in graphically with the lower part of the frame as it approaches the bottom part of the middle panel. Thus, when reading the text in what would seem the most natural order, we find ourselves inclined to read the text featured in the top part of the bottom most frame between the two panels in the middle in a natural left-to-right, top-to-bottom order: "... Quinn himself was & the dummy ..." / "... and work was & the voice that & gave purpose to & the enterprise." / "Little by little, work & had become a presence in Quinn's life ..." In this sense, the bottom frame becomes a part of the middle panel; in a linear sense, it should thus be 'spent', inaccessible now to readers. Yet the most natural reading, both in terms of content and in the continuity created by the ellipsis "... " is to link it here with the bottom-rightmost "... his comrade & in solitude."

**Figure 20:** Example of speech bubbles allowing for movement in *City of Glass* (8)



In both the cases of Figure 19 and Figure 20, readers, motivated and searching for narrative, are forced to make judgments. While research has shown that manga reading skills are *developed*, meaning that they are acquired over time with manga reading experience, the results being more consistent interpretations of reading orders for panels and other aspects (Nakazawa, 2005), readers will nevertheless find themselves moving between them in a way that is not possible in linear texts. Whatever readers feel makes the most sense as a narrative, there are many cases where they will not be able to achieve a viable interpretation without moving back and forth between the frames. Research on the experience of reading manga has shown that rereading is a crucial part of the process, and it may come either immediately, before one moves through to the next page, or after one goes through the chapter or book to completion (Ingulsrud & Allen, 2009). These kinds of examples clearly speak of the importance of that process: Initial readings may not lead to appropriate interpretations, thus literally demanding readers go over the passages again. Yet as with the example above from *Nodame Cantabile* with the use of jagged-edged speech bubbles, text in manga, incorporated into the drawings as it is, often borrows graphic elements in order to add in extra information that can help with these processes. The role of visual representations



and katakana, there are no obvious clues telling you that this first line would just be free-standing onomatopoeia, until you finish the line and realize that it contains no verbs.

Similarities, of course, can be found within certain genre of literature; the stylized distinction between text types may be comparable in part to a script or a play in the sense that certain rules have been adopted in order to distinguish between what is meant to be spoken—and by whom—and what are just notes. Speakers' lines are identified by the inclusion of a name at the beginning of an utterance, separated from the body of the text by tabs or colons; set notes or information about how a certain line is to be performed are usually either in different font, font size, or style such as italics, and naturally lack an identifying speaker. Yet even in plays, the visual categorization of text serves a fundamentally different purpose from what it does in comics: Ultimately, plays are meant to be read aloud, taken up and interpreted by the spoken voice, and proper formatting allows the clarity necessary to do so; in manga, however, the visual presentation of text is another tool of narrative. Sam Smiley (2005), writing on the art of playwriting, states that “(s)tating a play in a professionally typed format is an essential aspect of the playwright’s craft (295),” and offers four “vital functions” of the playwriting format: the psychological effect of format on writers (in the sense of accomplishment); giving the play a professional appearance (thus helping its marketability); providing a way of estimating a script’s performance length (in that pages following the format can be accurately estimated to take up a certain amount of stage time to speak); and readability (to add clarity to reading) (295-296). While the first two are largely surface-level and thus secondary, the third may seem comparable; after all, both Takeuchi (2005) and Yamaguchi (2005) have argued that certain aspects of language create a sense of time. Yet in manga, this is largely a tool used to the advantage of narrative, less concerned with actual minutes and seconds than with a *sensation* of time and rhythm in the reader, whereas Smiley’s time is more of a practical matter of real-time concerns for scheduling and acting. And while the visual presentation of text certainly affects readability, in manga, as some of the previous examples demonstrate, it can also complicate readability, making text in fact *less* clear for narrative purposes. Needless to say, the structure of the text in plays still follows a clearly linear order, forcing an order into the flow of reading to follow the relevant rules of the language.

As a result of this non-linear, visually-distinguished use of language in manga, there exists—in exaggeration, perhaps, but not to a strong degree—a cacophony of sounds and voices in any given panel, with tension existing between interflowing narratives, *Lines*, *Onomatopoeia* and other forms of auditory data. Manga have not surprisingly been compared to cinema, particularly in their use of frames (Rommens, 2000), but one might say that they are also similar in regard to this point: Movies, too, offer a multiplicity of sounds, quite literally, with background music, outside noises, narration and dialog sounding off at once, each having their own role in the narrative, each playing their own part. While other written literatures may certainly achieve a

sensation of sound and multiplicity of voices, few offer it on the same, non-linear, simultaneous way that manga and comics do, making this an essential part of the manga reading experience.



## 2.2 Data analysis

### 2.2.1 Text types in the corpus: The distribution of categories

If these issues sound familiar, it is because they are part of one of the key questions approached by this project, that is, what is the nature of the different types of text in manga, and what are they made up of? In particular, this issue was at the heart of the categorization of the text that was the first fundamental step in corpus design described in Chapter 1. Thus, to answer this question simply, I will reiterate here that they are made up of generally eight distinct kinds of text: *Lines*, *Thoughts*, *Narration*, *Background Text*, *Background Lines/Thoughts*, *Comments* and *Titles*. While the more detailed definitions of the categories may be found in Chapter 1, Table 4 summarizes their differences by environment and role. The issue that is perhaps more important, however, is how are these actually used in the corpus: Which ones are most central to the narrative? How do they differ between genres and individual series and genres? As noted in the initial introduction in Chapter 1, there appears to be some reason to think that there are differences in the usages of the categories dependent upon genre and individuals series. In answering these questions, I will take a largely quantitative approach, as I call upon some of the most basic data from the corpus concerning the distribution of text across those eight categories.

The basic data at hand here are listed in Table 5, a more detailed version of Table 3 seen in Chapter 1, which organizes the data by how many characters and entries are seen in total overall for all series and for each category. Note again that one entry here refers to one entry in the corpus; or one block of text within the original manga, whereas characters refer to the number of individual text ‘letters’. Focusing attention on the categories themselves, as has been hinted already, *Lines* made up the majority of all text by far, at 499,968 characters. *Lines* also formed the majority of entries, but the percentage that they comprise is considerably lower (61.81% vs. 72.64%). Given what we know about entries, this suggests that the average length of individual entries for *Lines* is considerably higher than that of the other categories; put differently, the other categories tend to be very short in length but are not infrequent *per se*. This is undoubtedly affected by the nature of categories such as *Onomatopoeia* and *Background Text*, since they tend not to be made up of full sentences, but rather, in the case of *Onomatopoeia*, individual words, and in the case of *Background Text*, snippets or small sections of greater objects. Thus a gap is found in the second place number of entries and characters: *Thoughts* (92,482 characters) came to be the second largest collection of characters where *Onomatopoeia* formed the second highest number of entries (8,531 vs. 6,561). This can be interpreted as showing that while *Onomatopoeia* appear more frequently on the page, when *Thoughts* do appear, they show up as a larger block of text, affecting how they are read.

**Table 4:** Summary of the roles and environments of the major text categories in manga

Category	Environment	Characteristics	Text
Lines	Unbroken speech bubbles	Audible information; primarily dialogue but some onomatopoeia as well	Generally type
Thoughts	Dot-tailed speech bubbles; squares or whited-out space; directly on the background	Characters' inner voices; are not audible to other characters; do not directly address the reader	Generally type
Narration	Square captions; directly on background	Text informing readers of plot development, location, etc.; primarily descriptive in nature, often featuring privileged information unknown to characters; can be in any person; is not audible or accessible to characters	Generally type
Onomatopoeia	Directly written on background	Do not form full sentences; are not spoken by anyone; are mimetic of real world sounds or describe the nature or atmosphere of a scene	Stylized; graphically handwritten
Background text	Part of drawing	Text written as a part of the scene; is not actually vocalized, such as advertisements, building names, etc.; appear as text to characters	Graphically incorporated into drawing
Background lines/thoughts	Directly written on background, sometimes marked by straight lines	Text representing secondary lines or thoughts; it is impossible to tell whether they vocalized; often jokes, criticisms or other non-essential information	Handwritten
Comments	Directly written on background, sometimes marked by arrows or stars	Notes or jokes about characters or items; supply privileged information about the scenes that has not been otherwise made available to the readers; generally non-essential information	Generally handwritten
Titles	In captions; directly on background	Titles or subtitles of the chapter or series name; authors' names.	Generally type

On any given page, one sees on average of 130.15 characters over 10.49 entries; most of those entries (6.48) are *Lines*, with 1.24 *Thoughts* and 1.61 *Onomatopoeia* (Table 5). All of the other text types appear on one in two pages or less. When looking at the average length of entries, one finds that *Lines* were not actually the average longest type (14.58 characters/entry): *Narration* (15.92 characters/entry) have somewhat larger chunks on average. This is probably partially the result of the greater variability of *Lines* in terms of their content; everything from a single exclamation point to long, monologue-like forms can exist as *Lines*, whereas *Narration* may be more restricted in what is permissible or likely given their relative infrequency.

Naturally, there is also some variation between series. Part of this is clearly due to differences in the number of total pages for each series, as the number of pages allocated to manga when originally published in chapter form differs both monthly and by magazine. However, the distribution of entries and characters is also dependent upon the structure of the series and stories they express. The number of *Lines* ranged from between 44.04% of all entries (*KimiTodo*: 57.87% of characters) to 73.01% (*One Piece*: 85.33% of characters), and while *Lines* thus still make the single most major category in any series, it is the variation between series that characterizes them. Other issues such as story lines may also be worth examining, and are picked up within the following chapters: Where the series with the most *Lines*, *One Piece*, is a fantasy-action story, the series with

the least, *KimiTodo*, is a high-school drama. It may be that *One Piece* is more oriented towards conversation and fighting scenes, typified by many *Lines* and a high percentage of *Onomatopoeia* (21.37% of entries, 8.29% of characters—the second highest percentage of *Onomatopoeia* entries, and the highest percentage of *Onomatopoeia* in any series) so as to describe the action, whereas *KimiTodo* focuses more on the psychology of the individual characters, characterized by a lower percentage of *Lines* in favor of a higher percentage of *Thoughts* (17.71% of entries, 19.80% of characters—the second highest percentage of *Thoughts* entries, and the highest percentage of *Thoughts* in any series). These two questions will be specifically addressed in Chapter 3 while looking at the vocabulary seen in manga using a morpheme analysis.

Table 5: Distribution of text categories by genre and series

Genre	Title	Pages	Lines		Thoughts		Narration		Onomatopoeia		Background Text		Background Lines/Thoughts		Comments		Titles		Total
			Entries	Characters	Entries	Characters	Entries	Characters	Entries	Characters	Entries	Characters	Entries	Characters	Entries	Characters	Entries	Characters	
Shounen-manga	Beerus	540	3464	33,538	1,072	9,324	47	0.846	536	512	9,156	3,216	1,712	1,712	179	3,206	2,416	4,486	43,792
			6.44	62.11	1.99	19.27	0.09	0.09	0.95	0.95	3.17	3.51	0.33	0.33	0.33	0.46	4.376	4,486	100.00%
	Kami	526	2,853	45,408	1,140	15,541	57	0.872	872	1,235	5,442	6,916	1,896	1,116	119	948	948	3,939	90,317
	Tobito		5.38	86.13	2.17	29.55	0.11	1.06	2.25	2.25	9.61	12.16	0.22	0.22	0.22	0.46	3,939	92	100.00%
	Nano	554	61,386	69,976	11,796	11,536	5,376	5,626	8,346	2,056	3,156	3,896	8,976	6,516	0.476	6,516	0.476	0.236	100.00%
			5.94	95.05	1.14	15.66	0.52	7.64	0.84	2.79	0.34	5.29	0.87	8.85	0.05	0.30	0.02	0.27	9.68
	Medusa	519	3,247	40,131	735	9,540	118	2,016	945	3,696	252	3,109	357	3,070	139	1,071	18	156	62,789
	Corde		6.26	77.12	1.42	18.38	0.22	3.88	1.65	7.12	0.49	5.99	0.09	3.92	0.27	2.06	0.03	0.12	120.98
	Reborn	506	3,802	17,772	706	9,342	1,382	1,482	1,052	5,512	4,392	1,552	722	3,112	33	1,112	12	112	66,176
			6.72	34.42	1.40	12.96	1.38	2.34	2.08	6.88	0.47	3.86	0.47	3.86	0.07	0.32	0.02	0.22	11.39
Shounen-manga	Total	2,645	16,239	219,513	4,285	52,435	582	8,839	4,164	16,504	977	30,115	2,330	21,622	342	2,394	73	682	331,495
			16.239	219.513	4.285	52.435	0.582	8.839	4.164	16.504	0.977	30.115	2.330	21.622	0.342	2.394	0.073	0.682	100.00%
	Depth	557	3,080	64,579	621	15,340	36	636	430	1,497	468	830	1	5	-	-	-	0.46	90,492
	Note		66.016	713.906	11.316	16.956	0.776	0.726	9.226	1.656	10.086	8.686	0.026	0.016	-	-	-	0.046	100.00%
	Gift	476	72,866	88,416	2,186	2,506	1,016	1,346	15,526	3,956	6,546	2,356	1,307	2,47	-	-	-	0.866	100.00%
			7.28	130.51	0.24	3.12	0.10	1.47	1.59	4.54	0.65	2.93	0.04	0.52	-	-	-	0.09	125.00
	Derivative	547	7,602	127,102	1,142	18,112	0.12	2.49	1.16	4.13	0.35	2.59	0.00	0.01	-	-	-	0.06	100.00%
	Canon		7.602	127.102	1.142	18.112	0.12	2.49	1.16	4.13	0.35	2.59	0.00	0.01	-	-	-	0.06	100.00%
	Nano	548	57,386	69,316	11,296	18,076	0.986	2,096	22,136	6,376	3,216	1,856	1,856	1,416	0.386	0.286	0.656	0.656	100.00%
			5.54	72.01	1.28	18.78	0.09	2.17	2.25	6.62	0.31	1.92	0.18	1.47	0.04	0.30	0.05	0.65	30.91
Grand Total	One Page	516	4,312	59,256	121	2,488	70	869	1,262	5,802	91	588	21	149	-	-	-	29	5,965
			73.016	85.336	2.056	3.556	1.396	1.246	21.376	8.296	1.546	0.846	0.386	0.216	-	-	-	0.496	100.00%
	Total	2,644	18,652	288,555	2,776	42,056	276	4,619	4,467	15,539	1,232	22,588	142	3,209	20	162	163	202	25,508
			18.652	288.555	2.776	42.056	0.276	4.619	4.467	15.539	1.232	22.588	0.142	3.209	0.020	0.162	0.163	0.202	100.00%
	Grand Total	5,289	61,816	772,036	11,836	13,446	1,596	13,658	15,504	4,576	0.416	4.29	4.26	4.32	0.07	0.48	0.04	0.51	130.15
			6.48	94.53	1.34	17.49	0.35	2.58	1.64	5.94	0.416	4.29	4.26	4.32	0.07	0.48	0.04	0.51	130.15
*The first number for each item represents the number of entries or characters. *The second number for each item represents the percentage they take up for that series or genre. *The third number for each item represents the average number of that type per page.																			

### 2.2.2 Textual categories and genre

In the meantime, while one sees some obvious major trends as noted above, there is also some real variance between series and genres. In particular, when one divides the sample into shoujo-manga and shounen-manga there are several interesting trends to be noted, starting with the fact that the percentage of *Lines* is significantly higher in shounen-manga titles than in shoujo-manga titles. They make up a much higher percentage of characters (72.63% vs. 66.22%;  $t(8) = 2.73$ ,  $p = .0262^*$ ), and there is also a significant difference in the percentage of entries that *Lines* make up (61.81% vs. 56.05%;  $t(8) = 2.62$ ,  $p = .0306^*$ ). As should be obvious, shoujo-manga have higher percentages of most of the categories, with the exception of *Onomatopoeia* (entries only; shoujo-manga actually rank higher in percentage of characters, betraying any expectations such as argued in Ôtsuka (1994) that action-oriented shounen-manga should use more *Onomatopoeia*), *Background Text* and *Titles*, all three of which are minor in difference, however.

Where is the rest of the text distributed, then? *Thoughts* take up a higher percentage of text in shoujo-manga on-average (15.82% of characters and 14.79% of entries vs. 11.23% of characters and 8.59% of entries), superficially supporting claims stressing the psychological and interior aspects of shoujo-manga (Schodt (1996, p. 155), Natsume (1997, pp. 165-166), etc.). Yet this does not hold up when looking at individual series: *GinTama* and *One Piece* showed almost no *Thoughts* (2.50%, 3.55% of characters; 2.36%, 2.05% of entries), whereas the other shounen-manga had a range between 13.09%~18.07% of characters (12.20%~13.31% of entries), similar to the ranges found in to shoujo-manga (11.53%~19.80% of characters, 11.79%~19.15% of entries). Instead, this tends to support Ôtsuka's (1994, pp. 56-72) assertion that while the shoujo-manga which came out in the mid 1970s to the early 1980s were interior-oriented, expressed in great part by their use of special types of narration, thoughts and monologues, such highly interior shoujo-manga actually fell out of favor in short time, creating a gap between the characteristics of shoujo-manga before, during and after the early 1980s. That current series should show little difference in the percentage of *Thoughts* appearing in genres may be a natural reflection of Ôtsuka's claim, then.

### 2.2.3 Defining handwritten lines, thoughts and authorial comments

It is at this point that it is worth considering two of the less commonly recognized forms of text within the corpus, those of *Background Lines/Thoughts* and *Comments*. These two categories were one of the major points of

difference in my categorization of the text in manga and those in the (albeit limited) previous studies. In fact, most of the definitions I propose for the more traditional categories—*Lines*, *Thoughts*, *Narration* and *Onomatopoeia*—follow the discussions found in Groensteen (2007) and Kai (1989). However, in compiling a corpus it was necessary to obtain a level of objectivity about what texts were relevant to the manga reading experience, and thus a major departure point from more traditional categorizations is that I also differentiated and included non-onomatopoeic text seen outside of speech bubbles and captions, etc.; *Background Text*, *Background Lines/Thoughts*, *Comments* and *Titles* are all four such speech types. As mentioned previously, of these four, *Background Lines/Thoughts* and *Comments* are the two categories which encompass handwritten lines, thoughts and authorial comments. While *Comments* refer directly to handwritten authorial comments, *Background Lines/Thoughts* encompass both handwritten background lines and handwritten thoughts, a point which will be explained below. As should be clear from Table 4, both of these categories have much in common with *Lines*, *Thoughts* and *Narration*: *Background Lines/Thoughts* represent characters' conversational lines or their inner thoughts, just as *Lines* and *Thoughts* do. Likewise, *Comments* present privileged information about characters or scenes in much the way of *Narration*. However, they also differ in several key ways, and they do not appear to be completely interchangeable. Three major factors can be noted here: (1) their visual presentation; (2) their relationship with characters; and (3) their roles in the narrative. It should be noted, however, that these three points are closely interrelated, and thus overlap to a degree.

The first factor, their visual presentation, is the most important difference between *Background Lines/Thoughts* and *Comments* and *Lines*, *Thoughts* and *Narration*. As noted already, *Background Lines/Thoughts* and *Comments* are written directly on the drawings themselves. In this sense, they seem to float within the drawings, not anchored to anything in particular; at most, there may be a set of straight lines on the sides of *Background Lines/Thoughts* (Figure 21) or an arrow pointing to an object or character with *Comments* (Figure 22), but these are optional and inconsistent. More often than not, the text appears bare. In comparison, *Lines*, *Thoughts* and *Narration* appear in a set number of clear environments. *Lines* as defined here only appear in unbroken speech bubbles. *Thoughts* tend to appear in special cloud-shaped or broken speech bubbles, with tails formed by linked dots; they may also appear in areas of whited-out negative space in the screen tone. *Narration* generally appears in square captions, sometime stylized but often just a rectangular box superimposed on the drawing. Both *Thoughts* and *Narration* occasionally appear written directly on the drawings, but in such cases they abide by the second major visual difference between them and *Background Lines/Thoughts* and *Comments*: They appear in type, with *furigana* (small phonetic gloss readings of kanji, placed next to the relevant character between lines of text) when applicable. (The role of *furigana* and what series and words receive them are described in more detail in Section 3.2 of Chapter 3) *Background Lines/Thoughts* and *Comments*, however,

are generally written in very casual, handwritten text. The text is often very small, and by no means neat: Readers may find that they must strain their eyes to correctly read some of them, especially when they are reading them reprinted in the *bunko* (A6) or tankoubon (B6) sizes, which are considerably smaller than the size they are when originally printed in magazine-form (See Couch (2000) for a comparative overview of how manga are published). Of course, there are some exceptions to this, and one may occasionally see handwritten text within speech bubbles or captions; in such cases, however, the fact that they are in the appropriate environment for *Lines*, *Thoughts* or *Narration* overrides that they are handwritten.

**Figure 21 (Left):** Example of *Background Lines/Thoughts* surrounded by lines in *Naruto* (1:121)  
**Figure 22 (Right):** Example of *Comments* with an arrow from Yazawa Ai's *Nana* (2:135)



The second factor is that while it is generally very easy to differentiate between these text types by how they appear visually, they also have a peculiar relationship with characters. In regard to *Background Lines/Thoughts*, which generally do not feature any identifying elements such as the tails on speech bubbles for *Lines*, it is difficult to determine who they are associated with, usually forcing the reader to guess that they belong to the character they are closest to. But this is not always possible; in many cases, they are separated from characters completely, or the closest characters are essentially indistinguishable “extras” (Figure 24). Naturally, as I mentioned with the example of Figure 19, there are times when it can also be difficult to tell who a particular speech bubble is associated with, but the inclusion of tails, whether *Lines*, triangles or dots, makes it is easier for readers to interpret the text properly. Yet combining the difficulty in ascribing *Background Lines/Thoughts* to individual characters with the fact that they are generally short, snippet-like texts makes it is often nearly impossible to tell if they are even meant to audible: While they sometimes appear to form part of a conversation, they are more often than not *not* responded to, leaving it a mystery as to whether or not they were meant to be

heard by other characters or simply personal remarks made only for the *speakers* themselves. The reason that *Background Lines/Thoughts* are compiled into one category derives directly from this observation.

**Figure 23:** Example of *Narration* using past tense from *Nana* (2:60)



*In Figure 23, Nana Komatsu—not featured on the page—is looking back at events using the past tense. The scene observed here comes to life in the next page, thus showing us the gap between the narrator’s sense of the presence and the events we witness.*

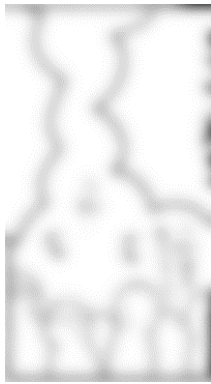
*Comments* also have an unusual relationship with characters which differentiates them from *Narration*. As Rommens (2000) discusses, narration in manga is kept to a minimum, usually only to explain what is not obvious by the drawings themselves. As a result, much of the narration we see simply acts to provide referential information to the scene such as the date, time or location. Even when narration is longer or more expressive than this, it is generally a part of an interior narrative, such as with the first-person narrator found in *Nana*. This narrator is, we come to understand, Nana Komatsu, and there is consistently a gap between the tense she uses in her narration and the scenes readers encounter: While the scenes occur right before our eyes, to Nana Komatsu, they are all things of a time gone by, as noted by her consistent use of the past tense (Figure 23). As a result of this gap, readers come to understand that something has happened, to which we will eventually become privy. Nana Komatsu’s narration in this sense is important to setting up the story, and locates her directly within the events. Thus, even such identifiable narrators do not know about the production of the manga, or even all of the scenes. Even if they have knowledge about the events by being in a different tense from the scenes, they do not



know everything about all of the characters. On the other hand, the speakers of *Comments* know things about characters that other characters should not know, and that other characters themselves may not know, either. They may be critical, complaining about characters' actions or about some gap in reality, and even sometimes about the drawings themselves (Figure 25). In this sense, *Comments* seem to be motivated differently from *Narration*, and they exist outside of the interior narrative; while they are closely familiar with the characters, the speakers of *Comments* are not involved as participants themselves.

**Figure 24 (Left):** Example of *Background Lines/Thoughts* with shell-like, extra characters from Shiina Karuho's *Kimi ni Todoke* (1:128)

**Figure 25 (Right):** Example of a critical *comment* in Ninomiya Tomoko's *Nodame Cantabile* (2:136)



*In Figure 24, it is impossible to determine the identity of the characters. In Figure 25, the Comment is critically noting that what the male teacher is doing is inappropriate—“Sude ni hanzai . .”, or, “It’s already criminal . .”.*

The third and final factor is that both *Background Lines/Thoughts* and *Comments* have a unique role within the narrative. As one might predict, in comparison with *Lines*, *Thoughts* and *Narration*, *Background Lines/Thoughts* and *Comments* are inherently secondary texts which offer readers non-essential information. Whereas *Lines*, *Thoughts* and *Narration* all feature information which can generally be assumed to be essential to understanding the story to some degree, what we see in *Background Lines/Thoughts* and *Comments* is not; they are jokes, notes, explanations or other types of information which, while perhaps interesting, are not necessary to understand the story. These types of text may be skipped over by the hurried reader, who would not miss anything except perhaps an extra depth or humor; they would still be able to successfully understand the goings-on of the narrative. Of course, this begs for a definition of “essential”, especially given that empty speech bubbles or ones only containing non-linguistic symbols are not uncommon. However, even empty speech bubbles would still play a role in creating time and interiority as Takeuchi (2005) describes; thus I would argue

that “essential information” does not necessarily mean that it supplies information about the narrative directly, but rather that it helps in the larger sense of creating a certain atmosphere, interiority or rhythm.

As hinted above, this secondary nature of *Background Lines/Thoughts* and *Comments* may be understood by differences in the use of furigana. In manga which feature furigana, *Lines*, *Thoughts* and *Narration*, as typed text, are all subject to having furigana, whereas *Background Lines/Thoughts* and *Comments* are not. Interestingly, this aligns them with *Background Text*—text which has been graphically incorporated into the drawings, such as advertisements and building names—which also do not feature furigana. While the lack of furigana on *Background Lines/Thoughts* and *Comments* may be partially attributed to the fact that, as handwritten text, they do not have the space to allow for it, I would propose that it is also because they are not essential reading; whether readers choose to fully read them is their own choice, but because readers do not need them to fully understand the text, it may not be necessary to make sure that all readers have access to them. Like *Background Text*, they may add to the atmosphere, but can be skipped over without harming one’s ability to understand the overall flow. On the other hand, *Lines*, *Thoughts* and *Narration* appear with the understanding that readers *will* and *must* read them to fully understand the text; thus furigana has to be given to such examples. In a similar vein, the sometimes messy, difficult to read, casual handwriting of *Background Lines/Thoughts* and *Comments* also suggests that authors are not as concerned with making sure they are comprehensible to all readers.

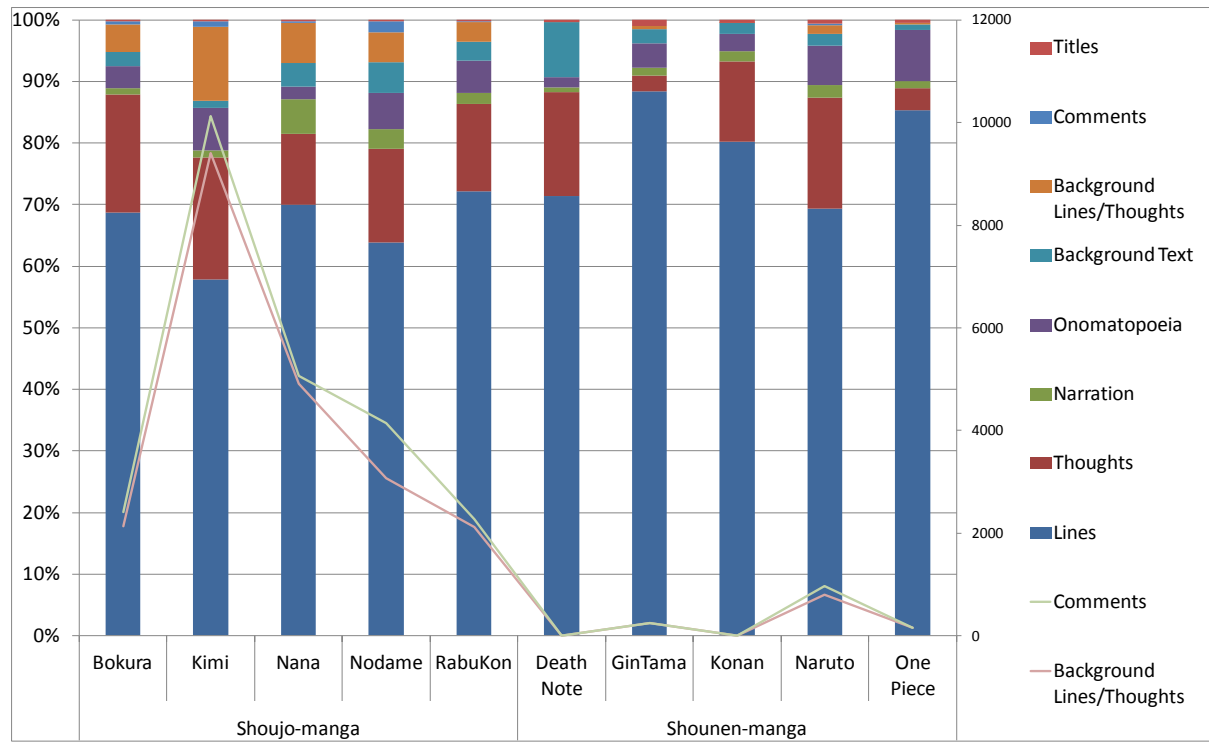
## 2.3 Analyzing *Background Lines/Thoughts* and *Comments*

### 2.3.1 The relationship of *Background Lines/Thoughts* and *Comments* with genre

Although their presentation within the text allows enough ambiguity in their status to prevent *Background Lines/Thoughts* and *Comments* from being wholly absorbed into the categories of *Lines*, *Thoughts* and *Narration*, one might still ask why they should not simply be treated as some variation of those types. Here, however, it is interesting to consider the fact that *Background Lines/Thoughts* and *Comments* actually show a very marked and distinct variation in distribution (Table 5; Graph 1). Although they appear infrequently compared with the other categories, both *Background Lines/Thoughts* and *Comments* appear at least occasionally in shoujo-manga, but minimally at best in shounen-manga. In shoujo-manga, *Background Lines/Thoughts* seem to appear on a regular basis, coming in at an average of 6.52% of all characters, and appearing a little less than once a page (0.87 entries/page), as opposed to shounen-manga, where they average 0.34%. While *Comments* are comparatively less common, they are still seen regularly irregularly in shoujo-manga (0.72% of all characters), but almost never in shounen-manga (0.06%).

These are also the only categories that one can say really show a consistent division along genre lines. As I discussed earlier, while when looking at averages, shounen-manga seem to have considerable more *Lines* than shoujo-manga (78.59% vs. 66.22% of characters), the variation within series is actually very extreme, going from as low as 69.31% (*Naruto*) to as high as 88.41% (*GinTama*) in shounen-manga, and from as low as 57.87% (*Kimi ni Todoke*) to as high as 72.20% (*RabuKon*) in shoujo-manga. Thus while the range still places the shoujo-manga with the least amount of *Lines* below the same-ranking shounen-manga—and vice-versa for the highest—there is also much overlap in their ranges. Similar observations may be made for both *Onomatopoeia* and *Thoughts*. In the case of *Background Lines/Thoughts* and *Comments*, however, the only shounen-manga to use either text type with any frequency is *Naruto* (*Background Lines/Thoughts*: 1.41%, *Comments*: 0.28%); their use in *Naruto* is, however, extremely restricted, showing up almost exclusively to describe the contradictions in the *burikko* (feigned-cute) character Sakura's outside demeanor and inner thoughts (Figure 26). All the other series showed *no* examples of *Comments*, and less than 0.40% for *Background Lines/Thoughts*.

**Graph 1:** Distribution of text across series by text type, with details for *Background Lines/Thoughts* and *Comments*



**Figure 26:** Example of comments and *Background Lines/Thoughts* in *Naruto* (1:111)



The explanatory text on the left Sakura's forehead ("Uchi naru Sakura" "The inner Sakura") categorizes as Comments, whereas the handwritten text above her ("Kou yuu no kekkou suki nano youu!!" "I really like these kinda things!") categorize as Background Lines/Thoughts

Interestingly, this is not the first time textual differences have been given for differences in genre. As Ôtsuka (1994) discusses, visually distinct text-types were found in the form of highly visual narrative/monologue-like texts characterizing the so-called psychological mid 1970s and early 1980s shoujo-manga, styles which eventually fell out of popularity and criticized as being difficult to read. Ôtsuka writes that “. . . in shounen-manga, aside from handwritten-onomatopoeia, the majority of text is in speech bubbles . . . , but shoujo-manga has much text outside of speech bubbles. And, to put it simpler, this is most likely the only standard by which to differentiate between shoujo-manga and shounen-manga (1994, p. 61).” This text that he describes as appearing outside of speech bubbles generally corresponded to interior thoughts or monologues, which, according to his explanation, were developed by female manga-writers in the mid 1970s of the 24-*gumi*, a group of shoujo-manga writers born in the 24<sup>th</sup> year of the Shouwa period (1949) with a distinctly new, literary-bent. Seeking to express characters’ psychological depth amongst the “discovery of the interior”, they developed new forms of text which lead to a new systemization of interiority in shoujo-manga.

However, Ôtsuka notes that with the emergence of (female) shoujo-manga readers who found it difficult to read such interior-oriented shoujo-manga, these types of text seemed to disappear or fall out of favor: The main difficulty for such readers—and male, non-readers of shoujo-manga, too—was to be found not in the drawings or narratives, but in their inability to get used to the “multi-layering” within which shoujo-manga finds its “essence”, and the inability of “their sense to follow the words written as the ‘interior words of consciousness’ that appear like poems (1994, p. 67)” in shoujo-manga. When such text outside of speech bubbles did get used, it was used in extremely restricted, patterned ways that differed from their original types: “. . . The characteristic use of words outside of speech-bubbles stopped at best at the level of ‘lines within their hearts’, and the form of expression found in shoujo-manga which objectified the self to infinity was abandoned (1994, p. 68).” In this way, Ôtsuka argues that there was a reversal in shoujo-manga trends characterized by a retreat backwards to the more traditional forms of shoujo-manga expressions.

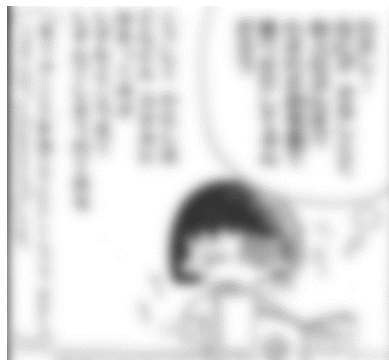
In some ways, the shoujo-manga collected here seem to superficially support Ôtsuka’s description of such psychologically-driven shoujo-manga. As noted above, they use less *Lines* than shounen-manga, and they also seem to have a slight edge on the number of *Thoughts* (15.82% vs. 14.79% of entries); however, the number of *Thoughts* varies radically between series. In fact, the shounen-manga *Naruto* actually uses more *Thoughts* (18.07%) than all but two shoujo-manga series, *Kimi Ni Todoke* (19.80%) and *Bokura ga Ita* (19.11%). Thus, not only have shoujo-manga lost some psychological depth, as Ôtsuka asserted, but shounen-manga may also have *gained* some. (Note that since I do not distinguish between some of the many different types of *Thoughts*, it is possible that they may not all be of the same, out-of-speech-bubble types that Ôtsuka describes; one might find somewhat different results with re-categorization.) On the other hand, however, *Background Lines/Thoughts*

and *Comments* seem in some ways similar to the patternization of Ôtsuka's interior monologues which came after their decline in the late 1980s. While *Background Lines/Thoughts* and *Comments* actually have some visual similarities with Ôtsuka's interior monologues, their roles differ greatly; *Comments* are clearly not associated with individual characters, and *Background Lines/Thoughts* also offer a sense of lightheartedness and humor where Ôtsuka's monologues offer reflection and depth. It is not hard to associate *Background Lines/Thoughts* with the "lighter" lines within their hearts that Ôtsuka says his interior monologues were reduced to, but he also describes a type of text very similar to *Comments*, writing that after the decline of the interior monologue, one began to see critical texts that appeared to be commenting from outside of the interior narrative. Finding these types in two popular series from the early 1990s, Sakura Momoko's *Chibi Maruko-chan* (*Little Maruko-chan*) and Sasaki Noriko's *Doubutsu no Oisha-san* (*The Animal Doctor*), Ôtsuka writes that

Both of these series do not use them [=text outside of speech bubbles] for "interior" expressions, but as a place to insert a kind of narration. But the distance between the narration and characters is very interesting . . . This narration looks at characters objectively, and adds accurate jabs in response to their actions and psychology. In the past words which corresponded to the "interior" were put in their place, through which readers were unified with characters to an even excessive degree. However, Sakura and Sasaki invite readers to the outside of the characters and the stories by placing narration there; characters are thoroughly objectified by this narration (1994, p. 71).

While the text is not accompanied by examples, and he still makes no differentiation between type and handwritten text, Ôtsuka's description here seems to correspond closely to the humor and critical stance which characterizes *Comments*; in fact, examining the first few volumes of both series, I found several examples of what would likely be classified as *Comments*, such as in Figure 27.

**Figure 27:** Comments-like text seen in *Chibi Maruko-chan* (1:17)



*From Momoko Sakura's Chibi Maruko-chan (1987, Tokyo, Japan: Shueisha)*

### 2.3.2 Interpreting the roles of *Background Lines/Thoughts and Comments*

While Ôtsuka's argument offers an interesting twist to the possible evolution of comment-like type texts, it must be admitted that the time line which is laid out here may not be entirely reflective of reality. One already sees examples of these kinds of texts in the works of authors like Keiko Takemiya, a member of the 24-*gumi* in the 1970s and 80s (Figure 28) placing them in the same time period and context as Ôtsuka's interior monologues, and even Ôtsuka himself is somewhat contradictory on this point. While in Ôtsuka (1989) he already notes the emergence of authorial notes in some early works, only to later describe the critical comments he finds in works like *Chibi Maruko-chan* as "new" in Ôtsuka (1994)—two types of examples which I would not only say were similar, but would classify under the same system here. In this sense, Ôtsuka left much to be considered. While I will touch upon the emergence of handwritten texts below, it is important to note here that Ôtsuka does not seem to value these kinds of texts as highly as he does his interior monologues. He goes on to emphasize that the drama of both series is very close to shounen-manga or *seinen-manga* (manga for adult men) in terms of the expressions used because they are primarily pushed forward by *Lines*, giving them a different kind of ease of reading not found in old-style shoujo-manga; this, he says, is one of the factors for their becoming best sellers (1994, p. 71). This is not something that can be said of the series included in the corpus, however; as has become obvious, *Background Lines/Thoughts and Comments* in fact do *not* seem to be found in shounen-manga, and while two of shoujo-manga, *Nodame* and *RabuKon*, could be described as being comedic, they are not, however, primarily, *episodic*; their narratives do advance, as do their characters, who gain depth and new aspects to their personalities as the stories progress. The other two series, *Nana* and *Kimi ni Todoke*, both are also primarily dramas which have the development of their characters at heart, and even though the percentage of "shounen-manga-like" *Lines* goes down, *Kimi ni Todoke* in particular has the second highest average-number of *Comments* after *Nodame Cantabile*.

**Figure 28:** *Background Lines/Thoughts* and comments as seen in Keiko Takemiya's *Kaze to Ki no Uta* (Song of the Wind and Trees) (1:67)



(Originally published in 1976; 2002 edition, Tokyo, Japan: Chuouokuron-shinsha)

*Background Lines/Thoughts* and *Comments* may play a larger role in the narrative than simply adding humor to the text. There are many similarities between *Background Lines/Thoughts* and *Comments* to Takeuchi's spectators (*boukansha*) and mediators (*baikaisha*), respectively, which he describes as techniques of expression which "draw readers into the manga (2005, p. 39)." Takeuchi argues that there are two types of narratives occurring simultaneously within manga, an internal narrative and an external narrative.

An interior narrative is the concept of a narrative which includes within itself the story (the development of the narrative in respect to a time-axis) and the plot (the networks of cause and effect related to the narrative). The external narratives are the elements of the narrative which, while indirectly participating in the development of the story, are in attendance within a separate, outside frame, and are in a supplemental location, so to speak. I would like to suggest that the complete narrative is constructed of both of these parts (2005, pp. 40-41).

Within this framework, mediator characters, which he states are essentially the placement of "authors' faces" within the text, are not deeply involved with the interior narrative, but rather look at it from the sidelines – i.e., they are a part of the external narrative. By introducing a character "labeled" as the author writers can hint at how the story will develop, and "fulfill the role of guide to the drama (2005, p. 40)." As he describes for Anno Moyoco's use of a "Moyoco character", they also, however, bridge the distance between readers and authors and make the manga more personal. On the other hand, spectators observe what is happening from the inside,



from a position on the sidelines, bridging the development of events. As non-main characters, the role of observers is to “bring a sense of sympathy to readers”, “encourage readers to identify with characters” and “help understand the situation (2005, p. 59).” (Note that although he does not use such terms himself, Takeuchi’s arguments may be thought comparable with discussions on diegetic layers. However, reconfiguring Takeuchi’s arguments in respect to such terms would be a task largely separate from the one set out here, and therefore must be left for some other opportunity.)

In the sense that they seem to represent the voices of authors, then, the parallel between *Comments* and Takeuchi’s mediator characters is clear; *Comments* are privileged in much the same way, being omniscient about the scenes being depicted, and having the same sense of distance between themselves and the stories as mediators do. Further, their critical touches offer much of the same parody-like humor that Takeuchi observes (2005, p. 49). While *Background Lines/Thoughts* may seem more distant from spectator characters as they are used by a wide variety of characters both central and not, it should be remembered that they are often used with shell figures, as mentioned earlier (Figure 24). In cases like these, *Background Lines/Thoughts* seem to act as noise, filling out the scenes and offering new points of view in much the same way as spectators. Even when they are used by main characters, however, the common gap between the main text (*Lines* and *Thoughts*) and what we see in *Background Lines/Thoughts* gives a different sense of perspective, filling out characters and offering readers a new way of looking at the text in a way similar to Takeuchi’s spectators.

### 2.3.3 On the need to use different textual styles

Although a possible context and role for these text types has been defined, it is not yet clear why they should be presented *visually* as they are. To bring the circle to a close, this final section deals with this key question of what graphically-different writing does for manga by examining the peculiar role of handwritten text. Handwritten text of course is nothing new to manga, *per se*. While not a thorough review of its usage by any means, if one goes back to the early one frame political cartoons found in *Tokyo Puck* from the Meiji period (1868-1912) one finds that handwritten text is quite common, and it can also be found in works like *Enoshima Kamakura Choutan Ryokou* by Beisaku Taguchi, which Shimizu (1991, p. 107) describes as one of the earliest examples of story-manga in the Meiji period. However, with the development of the manga market after WWII, a new norm seems to have developed, with type emerging as the major way to write most text. I should note that in certain genre, such as *4-koma* (frame) gag-manga, handwritten text seems to have maintained a consistent

presence over time; thus it should be remembered that, as mine is a corpus of popular story-manga, I am primarily making these arguments with similar popular series in mind.

Looking at *Shou-chan no Bouken* (Kabashima & Oda, 2003), a popular series that straddled the border of *e-monogatari* (picture stories) and manga published from the end of the Taishou period (1912-1926) to the first part of the Shouwa period (1926-1989), aside from the exclusive use of katakana (a Japanese syllabary), the earlier works from the late Taishou period differ primarily in that while narration—in the gutters and margins between panels—is written in type, the writing within speech bubbles is all handwritten. In the works dating from the post war period 20 years later, however, not only has all the text reverted to hiragana (a different Japanese syllabary now used as the standard for writing function words) and kanji mixed text, but it is also written exclusively in type. (This examination is based upon the republished edition of Kabashima & Oda's (2003) *Shou-chan no Bouken*). One can hypothesize several reasons for the switch to type as the major way of writing in manga: It is consistent in style and easier for readers to process; it can be manipulated by other types/fonts for more interesting styles; and the emergence of new technology may have made it easy to incorporate it into non-linear text such as it appears in speech-bubbles.

However, there seems to have been a reemergence of handwritten text starting in the 1970s coinciding with new changes in the roles of authors in shoujo-manga. As Nanba (2001) describes, the then-booming 1970s shoujo-manga differed from its 1960s counter-parts in several ways, including the emergence of *furoku* (special extra gifts with magazines) and the disappearance of male writers. Especially important, however, was the emergence of the “manga-writer as star” model. Shoujo-manga authors, who in the 1950s shared space within more-general girls' magazines, were originally not given much consideration, and even in the 1960s it was not uncommon to find shoujo-manga published without any author credits. Starting in those years, however, one began to see “author corners” or other spaces for authors to write about what was going on in their lives, give advice to readers and comment about their series, amongst other things. These forms of text really began to boom in the 1970s—and thus coinciding with the 24-group and the development of interiority—encouraging the formation of a sense of an “imagined community (*souzou no kyoudoutai*)” between readers and authors (Nanba, 2001, p. 202), through which readers became both interested and involved with authors. Writing on the connection between readers and authors as the same shoujos, Ôtsuka argues that this emergence was characterized by the use of handwritten text; it was around this time that he began to see authors write *Comments* not at all directly related to the text such as notes to readers, often in the cut areas outside of panels but also, occasionally, as he mentions in the case of one work by Hideko Tachikake, directly in the panels themselves (1989, pp. 61-62).

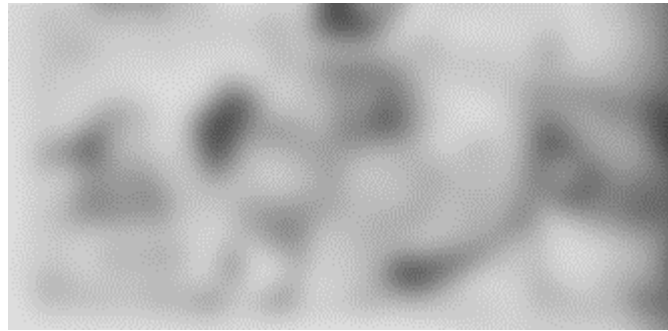
While just a part of the whole, with much of this text being handwritten, not only did the relationship between authors, readers and shoujo-manga change, but that change was largely mediated through handwritten text: Writing from authors was encoded through this usage, giving it a distinct flavor and space within the text. Thinking about how this relates to *Background Lines/Thoughts* and *Comments*, it is not difficult to hypothesize that *Comments* in particular might be considered a special form of this kind of text; they are a location for authors to insert their voice in the text, making readers aware of their relationship in the creative process. As Mealing writes, “hand-written text clearly speaks of its author . . . (2003, p. 47),” and thus readers who encounter it will, if only unconsciously, associate it with some person. Even with the retreat of interiority from shoujo-manga, this similar awareness of the existence of manga writers has not changed; thus in this sense, it is perhaps not surprising that these types of texts should still remain while Ôtsuka’s began to retreat.

At the same time, if one considers them to be similar to Takeuchi’s mediator characters, which have a long-established history, then they also have an organic place within the history of manga expressions. The same sense of personality and closeness that handwritten text gives to *Comments* may also be applicable to *Background Lines/Thoughts*. While the role of *Background Lines/Thoughts* may be a “light” form of Ôtsuka’s interior monologues, one might suggest that the use of handwritten text here is in fact a natural progression. As Ôtsuka himself writes, one might say that his special interior monologues really should have been written with *hentai shoujo-moji*, a special type of writing style popular with young girls in the 1970s and 1980s typified by how it round it out characters: Because *hentai shoujo-moji* were used as a writing style to express young girls’ (shoujos’) inner selves, they are, he argues, inherently linked to his interior monologues (1989, p. 64). Note that he does *not* state that his interior monologues are written by handwritten text. He does, however, point out the use of different fonts, stating that the Nar font used commonly with his interior monologues was the typographic equivalent of *hentai shoujo-moji*; he also notes that the switch to Nar from a gothic font for these kinds of texts coincided with the development of *hentai shoujo-moji* (1989, pp. 63-64).

In this sense, just as *Comments* are better expressed through handwritten text in that they give them a personalized flavor of the “author”, *Background Lines/Thoughts* might be considered to be better written by hand because they represent a personal side of characters. As Takayanagi comments while observing the usage of different types of writing in billboards, using handwritten text over type may give a sense of warmth and familiarity to writing (2005, pp. 234-243). Thus, even when *Background Lines/Thoughts* are used with extras and side-characters, the use of handwritten text may impart them with a sense of closeness or a presence that would not be possible with type. While *hentai shoujo-moji* have since lost their ground, we might still expect handwritten text to continue as other popular writing styles emerge. As Nanba (2001, p. 209) writes, while the writers of shoujo-manga for the now popular “teen magazines” are not users of *hentai shoujo-moji*, they are,

however, users of the *hetauma-moji* – good poor-writing – that have come to be associated with the teen readers of magazines like *Egg*.

**Figure 29:** Example of comment-like text within bubbles in Takemiya Keiko's *Watashi wo Tsuki made Tsuretette!* (*Take Me to the Moon!*) (1:56)



(Originally published in 1981; 1988 edition, Tokyo, Japan: Kadokawa-shoten)

Of course, handwritten text is not the only method of expressing the personalities of different characters or authors, nor does it exclusively serve that purpose. Switching between type and handwritten text may also offer authors a strategic way of dealing with text in that it allows authors to visually-code data as being more or less relevant, which is complimented by their appearance *outside* of speech bubbles and captions. As Gilreath reviews, there have been many studies showing that “graphic cues can greatly improve performance for virtually all reading strategies, ranging from careful serial reading to searching, surveying, browsing, skimming and other forms of selective reading (1993, p. 342).” We can expect that manga should be no different from other texts, and if anything they should utilize such points all the more actively as a mixed, visual media, and as Allen & Ingulsrud’s (2007) experiment showed, readers clearly look for different types of text, with the visual representation of text clearly an important factor in choosing what to read. This applies to *Background Lines/Thoughts* and *Comments* in two ways. First, by placing them outside of speech bubbles, authors are able to take away from its focus as a major item to be read: Readers searching for speech bubbles exclusively will automatically skip them. Second, by writing them by hand they are also able to distinguish such text from other, more central text types which can appear outside of speech bubbles such as *Thoughts*. In this way, readers can search important information more efficiently: Those who wish to skim can skip right through such text, but those who wish to read thoroughly may take the time to absorb all of the items, thereby not only getting more information which might give a new depth, humor or insight into the events being depicted. Allen & Ingulsrud (2007) have also emphasized the role of rereading and manga, noting that many readers recount how they go

over the series multiple times. It is easy to think that *Background Lines/Thoughts* and *Comments* may also offer enticing bits for such readers to come back to later.

Looking at earlier works that do use texts of a similar nature to the ones described here, one notices that the separation of these kinds of handwritten texts from speech bubbles and captions may not yet have been complete. For example, in the works of Keiko Takemiya, one notices that text similar to *Comments* and *Background Lines/Thoughts* continue to be seen both in speech bubbles and directly on the drawings, suggesting that their patternization was not yet complete (Figure 29). In this sense, the real separation of these kinds of texts may be the major difference between them and earlier types.

The visual encoding of these texts may be particularly important given the special role they may be fulfilling. On the one hand, not all readers may want a close sense of “community” with authors, so that changing the way that the texts are written may offer them the choice, i.e., to participate in it or not through its consumption (=reading). If, for example, the presence of the authors was made more obvious—such as through the use of actual “author characters”—then readers who felt that such authorial presences took away from the reading experience may not be able to avoid reading them. At the same time, such texts may still fill out the narrative through the roles of mediators and spectators visually; encoded as it is by handwritten text, the text alone, regardless of its contents, may still create a sense of multiple voices just by being visible to readers as it implies that there is someone there with something to “say”.

## 2.4 Conclusions

Writing on how typography effects how text may be read, Waller comments that

. . . (W)hile speech is on the whole spontaneous, text is a planned communication and makes a series of assumption and predictions about the reader. It is, therefore, the product of a *design process*; and whereas the production of a text involves predictions ranging from the reader's knowledge and purposes to his or her eyesight or the size of the bookshelf, clearly our concept of "design" must refer not only to the visual appearance of the document but to all aspects of information defined in the broadest manner. If 'literacy' may be taken to refer to the skilled use of the written word, *it is as much a design as a linguistic issue* (1980, p. 243) (emphasis added).

It is interesting to consider Waller's point here in light of the increasing attention given to the concept of a "manga (comics) literacy", or a set of skills necessary to read manga and other comics effectively in works (Ingulsrud & Allen (2009); Nakazawa (2005)). This is clearly a part of a larger trend of assessing the skill sets necessary to deal with many different kinds of media, such as Kress (2003) does for new media. In the case of manga, where linguistic language walks a fine line between being a purely graphic entity and one meant to represent words, the issue of design could be said to be particularly important, as authors make decisions about how to represent text in dynamic ways. The issue presented by the roles of different text types in manga is one that is inherently connected with this kind of literacy, and not just what texts present in terms of their contents but the way that those texts are presented *visually* is a key part of this.

With their main differences with other, more established text being how they are presented visually, handwritten *Background Lines/Thoughts* and *Comments* offer a particularly interesting view into how text, even as it acts as linguistic language, can be integrated into the visual structure. Unfortunately, it was beyond the scope of this chapter to fully explore the roots of these texts, and to further the discussion advanced here, it will be necessary to more fully trace their beginnings. However, as I have argued, there appears to be reason to think that these often over looked texts play a vital role in the reading experience. Visually encoded as they are, readers are able to choose how they want to read, creating a dynamic element to the manga reading experience which allows for both scan-reading *and* for the development of deeper, close readings. Yet regardless of their linguistic context, their existence, particularly as handwritten text, always implies a writer, offering a sense of depth which is largely created by visual structures. While only one aspect of the larger phenomenon, it is in ways like this that textual language and visual language intertwine to offer a deeper manga experience.

Finally, one important thing that this chapter established was the importance of linguistic structures in differentiating genre and manga more generally. In the arguments described here, I focused on the ways that language is mediated in text, hence the emphasis on the relationship between visual structures and text. However, this meant that I gave only passing attention to what the language was actually communicating. In Chapter 3, I will deal with exactly that question through an analysis of the orthographic and lexical characteristics seen. By doing this, I will bring the discussion one level deeper as I move from the structural to the contextual questions of language in manga.

### 3. The Orthographic and Lexical Characteristics of Manga

#### 3.1 Chapter Themes and Overview

Whereas in the previous chapter I dealt with language in manga from the perspective of its structural characteristics, in this chapter I will now move the focus down to the actual *content* of the language seen. As they make up the large majority of text, my analysis here focuses in primarily on *Lines*, and will be approached in two ways: first by exploring the orthographic characteristics of manga, and second by examining the vocabulary observed through morphological analysis. In between these two studies, I will complement the discussion by a study on the *readability* of manga—that is, how easy manga are to read—in order to consider how these issues impact the reading of manga.

As I will demonstrate, the orthographic characteristics of manga are unique in comparison with more traditional or conservative forms of writing, such as newspapers or novels. However, as Coates (2010) suggests, they also have some shared characteristics with other genre, such as *keitai-shousetsu* (cell phone novels) and *keitai-meeru* (cell phone emails), which in response also have many characteristics in common with Satake's (1980) *shin-genbun 'icchi-tai*, or new vernacular, conversational style of writing. These characteristics, thus, are one way of assessing how manga are situated as works of writing, placing them closer to some writing types while creating distance from others. They also say something about the quality of the vocabulary seen in manga, which is further associated with genre. As the different scripts are each associated with different modes—hiragana, for example, is often associated with women and softness, whereas kanji are associated with men and more formal texts—and also different vocabulary—hiragana being primarily for native-Japanese words, and kanji for Sino-Japanese words—the actual usage of script can tell us something about the impression the texts give, and the vocabulary being seen, even without looking at the vocabulary directly. (For more details on the impacts and usages of different script types in Japanese, see Shibamoto & Schmidt (1996) and Taylor & Taylor (1995).) More directly, however, the unique style of writing seen in manga has an impact on how it is processed by text analyzing programs used for morphological analysis. Because most programs are informed by more conservative, standard models of orthography, they work very poorly for texts which deviate from such norms (Kaalep & Muischnek, 2011). Understanding how manga manipulate orthography is thus crucial for their



morpheme analysis, and a large portion of this chapter will deal with manga's orthographic style, which could offer some insights to those working with similar orthographic styles.

More specifically, the chapter itself will advance as follows. First, after observing the relationship of orthography in linguistics studies in Section 3.2.1, I will go over the characteristics of Japanese orthography in regard to some of the particular issues in manga (Section 3.2.2). In doing this, it will become clear that the types of text used and the use of unusual furigana—glosses—are two of the more crucial issues for manga, and Sections 3.2.3 to 3.2.7 will summarize two studies on these topics. Through these studies, it will become clear that manga have a distinct orthographic style which imparts to them a sense of orality, making them closer to the spoken speech they emulate. In the next Section 3.3, I will look at the impact that these orthographic characteristics have on an important aspect of manga: their readability. Using newly developed tools made available by Sato, Matsuyoshi & Kondoh (2008) and the Readability Research Laboratory at the Nagaoka University of Technology (Readability Research Laboratory, 2010), I will show that they are generally rated as being very easy to read, primarily because of their low usage of kanji. However, I will also argue that the more *vernacular* writing styles may prove more difficult than assumed given that they are not forms formally taught in school. Finally, in Section 3.4, I will introduce a morpheme analysis I conducted using software designed by the National Institute for Japanese Language and Linguistics (NINJAL), detailing the most frequent word types and differences in genre. Through this analysis, the connection between manga and more informal texts will again become clear in the use of different word classes in particular, thus giving more clarity into manga's *style*.

## 3.2 Orthography in manga: Studies on writing types and furigana

### 3.2.1 Orthography in language research

Traditionally, the trend in mainstream linguistics has been to overlook orthographic aspects of language as being metalinguistic and thus outside the scope of *real* language study. As Aranoff (1985, p. 28) notes, “(l)ong ago we found a rationale for its banishment in the undoubtedly correct observation that spoken language is ‘true’ language, while written language is an artifact—the necessarily imperfect product of human intelligence.” As a human product, written language has been seen as not being purely linguistic, with too many artificial aspects to reflect the real state of language. Yet recent years have shown a new level of interest in the role of orthography. Research has gone on to show that orthographic issues affect how language is processed and perceived from a variety of positions (such as on the role of writing systems on second language learning (Cook & Bassetti, 2005); the role of visual forms (Lavidor, Babkof, & Faust, 2001); and the relationship between exposure to non-standard forms and recognizing mistakes (Powell & Dixon, 2011)), demonstrating that orthography is relevant to linguistic analysis. Deviations from standard orthographic practice have also proven to be not simply “mistakes”, but reflective of linguistic changes (e.g., Myslín & Gries (2010) on Spanish internet orthography), proving orthographic characteristics offer real input into linguistic processes.

In particular, however, one issue at the heart of much of this new found interest could be changes in writing practices with the advent of new media and genres such as internet message boards and texting. The particularities of such genres have led to changes in the use of spelling and punctuation symbols across a variety of languages. The language found on message boards and in text messages, for example, is widely reported for many languages as featuring new abbreviations, more phonetic spellings, and orthographic symbols for non-linguistic usages, e.g., smiley faces, etc., from Japanese (Nishimura, 2003) and English (Werry, (1996); Shortis (2009)) to Spanish (Myslín & Gries, 2010), Arabic (Abolfazl & Omidvar, 2011), French (Anis, 2007) and Mandarin in Taiwan (Su, 2007), as well as less familiar languages as Romanian (Pacea, 2009) and Estonian (Kaalep & Muischnek, 2011)). In their deviations from standard texts, the text styles seen online ask us to rethink how text is processed and how such new styles are positioned within traditional genres. Similar to literacy, orthography has also come to be perceived as a *social practice*, with what is acceptable and approved of changing by a variety of social variables (for an overview of literacy as a social practice, see Street (2009)). Non-standard forms are in particular often associated with *sociocultural meaning* as writers align themselves through orthography (Androutsopoulos, 2000).

While manga may not immediately seem to fit into the above groups, being neither internet-based nor direct written communication between concerned parties, manga have long been mentioned as a source of unusual orthographic patterns (e.g., the use of non-standard symbols, Nishimura (2003); non-standard usages of geminate markers in Akitsuki (2009) and Coates (2010)). As Shortis (2009) argues for English, the trends observed in new forms of writing do not exist within a vacuum, but can be rather said to be replicating, recycling and reinventing forms already seen in other genres. Manga may be one such resource for Japanese: Nishimura (2003) points out that some of the characteristic symbols later seen in online chat forums, such as roman letters and non-orthographic symbols like ♪ and ☆, are commonly found in manga.

As a popular medium with its primary focus on spoken lines, manga may also be more closely aligned with texts approaching “written speech”. As Tannen (1980) argues, the line between spoken and written speech is not as finely drawn as has been assumed: While written and spoken speech are generally assumed to be two separate entities, in reality, some written texts may have more in common with spoken texts than with typical “written texts”, and some spoken texts may closer approach the characteristics of some written texts than typical “spoken texts”. Groensteen (2007) notes that in Western comics, it is not unusual to see non-standard spelling, which he argues is an attempt to create a sense of orality necessary for creating a sense of reality. Similar orthographic “deviations” may be expected in manga, and may be one way of assessing its positioning as a “written text”. Shibamoto & Schmidt (1996) note that there appear to be distinct genre differences and sociocultural associations with the different types of scripts used in Japanese, causing them to argue for a sociolinguistic approach to script usage and orthography. One may expect, then, that looking at the orthographic practices of manga would allow one to analyze how it is positioned *vis-à-vis* other texts using similar non-standard practices.

### 3.2.2 Orthography and Japanese

Before any actual discussion can be held on the nature of orthography, however, it is important to understand fully the complexities of the Japanese writing system. Today, Japanese is written primarily using three scripts: two syllabaries (*kana*)—hiragana, used for function and native Japanese words, and katakana, used for foreign loan words and onomatopoeia—and kanji, or Chinese characters. *Roomaji* (the roman alphabet) and Arabic numbers are also used to a limited extent. While hiragana and katakana are extremely transparent phonetically, mapping with few exceptions one-syllable-to-one-“letter”, kanji, on the other hand, map to several different

possible sounds. Due to the particular adaptations which were undertaken to use kanji for Japanese when they were imported from Chinese, most kanji are read with a number of different native-Japanese *kun* readings and Sino-Japanese *on* readings, a point I will touch upon again below. It should be noted that the complexities involved in using such multiple systems of writing have led to claims that Japanese is one of the most difficult languages in the world to read (Gottlieb, *Language and Society in Japan*, 2005). More details on the specifics of how the Japanese writing system developed and its characteristics may be found in Taylor & Taylor (1995).

In addition to the writing types mentioned above, Japanese also uses a variety of orthographic symbols, taken from both Japanese and western traditions. The most common orthographic symbols used are likely the Japanese comma and period, or the touten (、) and the kuten (。); other common punctuation marks include the exclamation and question mark (！, ?) and Japanese quotation marks (「, 」). An example of a punctuation mark in the Japanese tradition is the *odori-ji* (々), which shows that the same character will be repeated, such as in the name Nana (奈々). In general, the majority of text today is written in either hiragana or kanji, followed by katakana, punctuation marks, Arabic numerals and roomaji. However, as Shibamoto & Schmidt (1996) demonstrate, the usage of different script types is actually highly variable, and seems to differ by genres, which they interpret as being a reflection of the social practices involved in orthography.

Somewhat separate from the issues above but of particular relevance to orthography in manga is the use of furigana, or phonetic glosses, in text. Due to the peculiarities of kanji, while context generally makes the reading of compounds clear, there remains a significant degree of ambiguity when encountering kanji words. Many kanji words can be read in multiple ways. Semantic and stylistic context often makes the reading obvious; when the word “今日” (*now + day*) is used within text representing spoken speech, it is most likely read *kyou* (today), whereas when it is used in formal writing would most likely be read *konnichi* (these days). However, sometimes readings can still be ambiguous. For example, “明日” (*clear + day*), meaning tomorrow, is almost always read *ashita*, but may also be read *asu*. While *asu* is more formal, semantically it does not differ, meaning that it can generally if not colloquially always be read either way. In such cases, one may add furigana to the compound in the margins in order to make the compound unambiguous, even in texts which would otherwise not use furigana. Example 3 shows how one sentence would appear with and without furigana on the kanji.

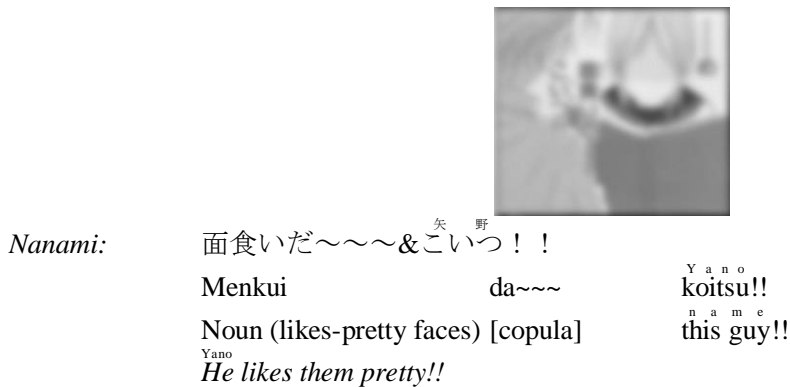
**Example 3: Text with and without furigana**

<i>Without furigana</i>	犬は嫌いじゃないけど、猫の方が好きだ。	=19 characters
<i>With furigana</i>	いぬ きらい じゃない けど、ねこ ほう す 猫の方が好きだ。	
<i>Character types</i>	KHKHHHHHHHPKHKHKHHP (K=Kanji, H=Hiragana, P=Punctuation)	
<i>Roomaji</i>	Inu wa kirai ja nai kedo, neko no hou ga suki da.	
<i>Gloss</i>	<i>I don't hate dogs, but I prefer cats.</i>	

Furigana have very early origins within the history of Japanese writing, having been born in the context of Buddhist texts with related examples found as early as the 9<sup>th</sup> century. Written in a form of Chinese, such texts were expected to be read in Japanese, using rules to change the word order and grammar. Furigana were originally used to demonstrate the word order as non-phonetic symbols, as well as to show the phonetic reading of given words (for an overview, see Wilkerson & Wilkerson (2000) and Imano (2009)). Today, furigana primarily demonstrate words' phonetic readings, generally appearing in slightly smaller font than the main text in the right margins in vertical text, or on the top in horizontal text. Furigana are not included in most texts, as it is assumed non-necessary for adults who have completed their years of compulsory education. However, some genres aimed towards younger readers always include them, including some types of manga.

More specifically, most manga for young readers, i.e., shoujo-manga and shounen-manga and *jidou-manga* (children's manga) include furigana for all kanji, which has been said to make them easier to read than other genres, particularly for non-native speakers (Ito, 2000). In addition, authors can chose to actively use furigana on a selective basis, even in genres that do not usually utilize furigana. In these cases, authors are sometimes motivated to use furigana for their clarifying functions, but it is also not uncommon for furigana to be used exceptionally, i.e., in ways not related to phonetic ambiguity. As Imano (2009) writes, furigana can be described as being a form of "expression" (*hyougen toshite no furigana*), or, as Wilkerson & Wilkerson (2000) similarly argue, a form of "poetics". An example from manga may include the following *Thought* from *Bokura ga Ita* (Figure 30).

**Figure 30:** Non-phonetic furigana from *Bokura ga Ita* (1:118)



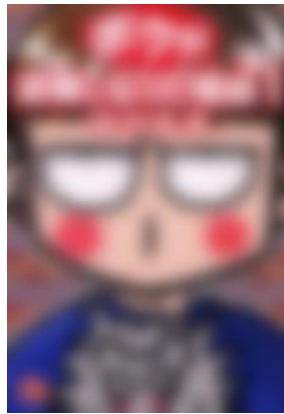
This *Thought* was made while the main female character Nanami is looking at a photo of the main male character Yano's dead ex-girlfriend. Here, the word *koitsu*, an informal third person pronoun meaning “this (person)”, which can be used for both sexes, is written in hiragana, meaning it is phonetically unambiguous. However, it has the kanji for the name Yano written in the margins, if anything thus making the phonetic reading unclear. It could be said to be clarifying deictic issues, however: Because Yano is not in the drawing, it is not clear whether it is him or his ex-girlfriend Nana that Nanami is referring to with *koitsu*. Thus, it adds no informative phonetic information, but can be said to be offering depth and clarity to the utterance in a different fashion.

While unusual, such exceptional usages of furigana are not unseen in manga, and further complicate their orthographic landscape. One can also expect that they might have an effect on the readability of manga: Although furigana are usually thought of as being elements which make reading *easier*, it may actually require more of readers to differentiate between these kinds of non-predictable usages and straightly phonetic furigana. This might immediately affect their usefulness in Japanese language education—furigana often being seen as a “godsend” (Schodt, 2006) for Japanese learners—but is also not without relevance to native Japanese speakers. It is interesting to note in regard to this that one of the professional opinions garnered from meetings in preparation to revise the jouyou-kanji in 2011 warned of young students' unfamiliarity with the proper readings of words like “理由” (*riyuu*, reason), which they often found to be read *wake* in many different texts (Japanese Ministry of Education, 2012). While this was not found in the series described here, examples can be found at least anecdotally in manga, a particularly conspicuous example being Yumi Tamura's *Boku ga Doroubou ni Natta Wake* (*The Reason I Became a Thief*), which writes *wake* (ワケ) in the title using “理由” throughout the series (Figure 31). Along with the possibility of analyzing these unusual furigana as an expressive aspect of

manga, one can see that unusual furigana in manga could be considered an influential aspect on how they are read and perceived.

With the above discussion in mind, I will offer two points as issues for examination: first, the types of writing used and their impact in how a manga style is developed, and the usage of furigana and their impact on how manga are read. To that end, I conducted two studies to observe and analyze the orthographic characteristics of manga. The first study focused on the actual types of text used (hiragana, kanji, etc.), and the second study focused on the use of the kinds of exception furigana described above. Below I review each study and their results before making some comparisons with other genres and assessing their possible impact on morpheme analysis, conducting a combined discussion for both parts.

**Figure 31:** An example of non-standard furigana in the title of *Boku ga Dorobou ni Natta Wake*



(Tamura, Yumi (1991), Tokyo, Japan: Shougakukan)

### 3.2.3 Study 1 – Writing types used: Methodology

To assess the overall orthographic characteristics of manga, I used regular expressions to divide the text into six groups: hiragana, katakana, kanji, roomaji, Arabic numerals and orthographic symbols. Four unusual types of kana styles were also targeted for extraction due to the reports by Akitsuki (2009), Coates (2010), Sadanobu (2008) and others as being common features in manga: (1) the small hiragana and katakana *tsu* (“っ” and “ッ”, respectively) geminate-marker (i.e., the double-k in the word *gakkou* “school” [がっこう, *ga-k-ko-u*]) in non-

geminate environments (e.g., at the ends of words, or proceeding vowels), (2) small hiragana and katakana vowels (“あ” vs. “ぁ”, “オ” vs. “ォ”), which are not normally used commonly, for long vowels, (3) the *bousen* symbol “ー”, normally used with katakana to create long vowels, in non-katakana environments (e.g., with kanji and hiragana), and (4) the *dakuon-ten* symbol “゛”, normally used with voiceless kana to express voiced consonants, with vowels to express issues unrelated to voicing.

### 3.2.4 Study 1 – Writing types used: Results

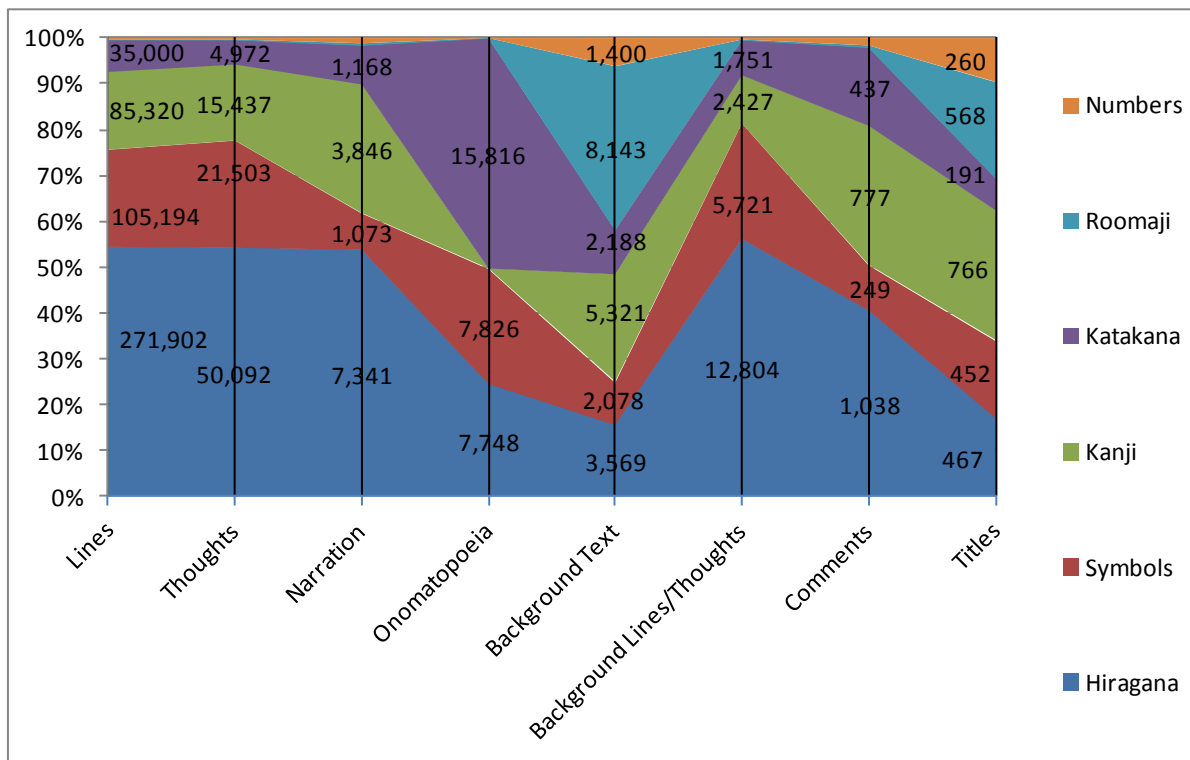
**Overall:** By far the most common text type in manga was hiragana, at 51.57% (Table 6). Interestingly, the second most common type of text was orthographic symbols at 20.93%. Kanji (16.55%) followed for third, and the remaining text types were all under 10%, with katakana at 8.94%; and 1.48% and 0.54%, respectively, for roomaji and Arabic numerals. Previously, Shibamoto & Schmidt (1996) looked at the scripts seen in manga, comparing them with several other types of writing, and the results appear to be overall comparable. Their results are also listed in Table 6, and it is interesting to note that the results appear rather different. In particular, they found that there was much more hiragana (73.05%). However, their study was somewhat less comprehensive than this one, in that they did not include symbols; they combined Arabic numerals with roomaji; and they only targeted text within speech bubbles or captions, e.g., text which would be called here *Lines*, *Thoughts* and *Narration*. They also did not take all of the text within their series as their target, but rather the first page and a mid-page only (Table 6 averages those two figures). Removing symbols from the total for *Lines*, hiragana would, however, still only account for 65.62% of all text, and kanji would go up to 21.06%. It is difficult to compare their data meaningfully, as the actual samples—including even the broad genres—and the sample sizes were not made available. The implication was that the sample series were for adults, which may make some difference, although it would be surprising for adult series to use *less* kanji. There is an approximately 10-year gap between these two studies, and it is also possible that manga has changed in that time, positing one possible place to look for advances in manga narratives.



**Table 6:** Distribution of text across genres

Genre		Hiragana	Symbols	Kanji	Katakana	Romaji	Numbers
Manga - Corpus	Shoujo	53.86	21.93	14.07	7.87	1.75	0.53
	Shounen	49.44	20.01	18.85	9.93	1.22	0.55
	Total	51.57	20.93	16.55	8.94	1.48	0.54
Manga - S.S. (1)		73.05	-	15.95	10.8	0.2	
Romance novels (1)		67.25	-	26.55	6.15	0.05	
Mystery novels (1)		66.95	-	28.95	4.1	0	
Science fiction novels (1)		65.75	-	26.75	7.25	0.25	
Business novels (1)		55.4	-	38.7	5.85	0.05	
Magazines (2)		51.4	6.6	20	20.4	0.7	0.9
Newspapers 1 (3)		35.32	10.70	46.15	6.07	0.41	1.36
Newspapers 2 (4)		34.06	15.58	41.46	6.34	0.46	2.10

(1)~(4) taken from (1): Shibamoto & Smith (1996); (2) Satake (1980); (3): Kyoudou-tsuushinsha (1971), *Moji-shutsudo chousa 100 manji shuukei kekka*, in Hayashi, Miyajima, Nomura, Egawa, Nakano, Sanada, and Satake (eds.) (1982); (4): Nozaki & Shimizu (2000)

**Graph 2:** Visualization of the distribution of text types

In comparison with other genres, manga appear to differ most dramatically in regard to symbols, with 20.93% much higher than what has been reported for both newspapers (10.70% (Kyoudou-tsuushinsha, 1971 in Hayashi, et al. (1982)) to 15.58% (Nozaki & Shimizu, 2000)) and magazines (6.6%, Satake (1980)) (Table 6). They also appear to have relatively few kanji, but this also appears to be partially related to the question of symbols. Note, however, that after removing symbols, the overall percentage of hiragana rises to 65.62%, and the percentage of kanji rises to 21.06%. Doing the same for Satake's results for magazines, hiragana and kanji would rise to 55.03% and 21.41%, respectively; for newspapers, Kyoudou-tsuushinsha's results would also rise to 39.54% and 51.68%, and Nozaki & Shimizu's results would rise to 40.35% and 49.11%. Thus, if one controls for symbols, manga would appear fairly similar to magazines in regard to their usage of kanji, but with more hiragana and fewer katakana; the distance between them with newspapers in regard to kanji would also be somewhat mitigated.

There were some important distinctions between text types and genres that should also be mentioned. In general, *Lines*, *Thoughts*, and *Background Lines/Thoughts* patterned similarly, which is perhaps not surprising given their overlapping textual/narrative roles (Table 7), averaging 54.16%~56.08% hiragana, 21.04%~25.06% symbols, 10.63%~17.07% kanji and 5.38%~7.67% katakana. *Narration* and *Comments* also patterned similarly, although somewhat differently from the other three types, with kanji taking a higher percentage (*Narration*: 28.16%, *Comments*: 30.40%) than symbols (*Narration*: 7.86%, *Comments*: 9.74%). This is presumably because *Narration* and *Comments* are both qualitatively different from the other three types; while they all make up sentences, *Narration* and *Comments* are not conversational, meaning they are less likely to use the symbols that help to make the text seem more like spoken speech, a point which will become more obvious before. On the other hand, *Background Text* included a large percentage of roomaji (35.87%), accounting for 80.16% of all roomaji. This is likely due to the text being primarily advertisements and slogans on clothing, meaning that the roomaji are largely used for design purposes, which seems consistent with the Japanese linguistic landscape (Kallen & Dhonnacha, 2010) and the nature of Japanese billboards (Someya, 2002). *Onomatopoeia* also use a very large percentage of katakana (50.30%), with very little kanji (0.05%). This is predictable, as Japanese onomatopoeia are primarily native-Japanese, as opposed to Sino-Japanese words which are the primarily word type using kanji, combined with the further tendency to write more sound-oriented words (such as foreign loan words) in katakana. With their distinct, boxy shapes, katakana may also stand out in the text, and as Kess & Miyamoto (1999, p. 30) note, they are often used as “visual italics” to “highlight exclamations”. This, too, might help explain why they are more common in onomatopoeia.

**Table 7:** Distribution of script types by category

Category	Hiragana		Symbols		Kanji		Katakana		Roomaji		Numbers		Total
<i>Lines</i>	271,902	54.38%	105,194	21.04%	85,320	17.07%	35,000	7.00%	1,047	0.21%	1,505	0.30%	499,968
<i>Thoughts</i>	50,092	54.16%	21,503	23.25%	15,437	16.69%	4,972	5.38%	219	0.24%	260	0.28%	92,483
<i>Narration</i>	7,341	53.75%	1,073	7.86%	3,846	28.16%	1,168	8.55%	74	0.54%	156	1.14%	13,658
<i>Onomatopoeia</i>	7,748	24.64%	7,826	24.89%	16	0.05%	15,816	50.30%	37	0.12%	0	0.00%	31,443
<i>Background Text</i>	3,569	15.72%	2,078	9.15%	5,321	23.44%	2,188	9.64%	8,143	35.87%	1,400	6.17%	22,699
<i>Background Lines/Thoughts</i>	12,804	56.08%	5,721	25.06%	2,427	10.63%	1,751	7.67%	55	0.24%	73	0.32%	22,831
<i>Comments</i>	1,038	40.61%	249	9.74%	777	30.40%	437	17.10%	14	0.55%	41	1.60%	2,556
<i>Titles</i>	467	17.27%	452	16.72%	766	28.33%	191	7.06%	568	21.01%	260	9.62%	2,704
Grand Total	354,961	51.57%	144,096	20.93%	113,910	16.55%	61,523	8.94%	10,157	1.48%	3,695	0.54%	688,342

In regard to genre differences (Table 8), shoujo-manga appeared to use hiragana somewhat more frequently than shounen-manga (53.86% vs. 49.44%), mostly accountable by differences in kanji (14.07 vs. 18.85%) and to a lesser degree katakana (7.87% vs. 9.93%). Interestingly, however, looking at the differences between text types in different genres and textual categories, it becomes clear that hiragana are used considerably more often in shoujo-manga than in shounen-manga (34.25% vs. 14.81%) for *Onomatopoeia* (Graph 3). For shoujo-manga, symbols, katakana and hiragana all take on similar percentages (28.61%, 36.85%, 34.25%, respectively), whereas for shounen-manga, the majority is katakana (64.06%). While Shibamoto & Schmidt (1996) found no significant differences between the scripts used by manga genres, these results for *Onomatopoeia* were statistically significant ( $t(8) = 4.09$ ,  $p = .0035^*$ ), pointing towards the importance of categories in genre, as was raised in Chapter 2.

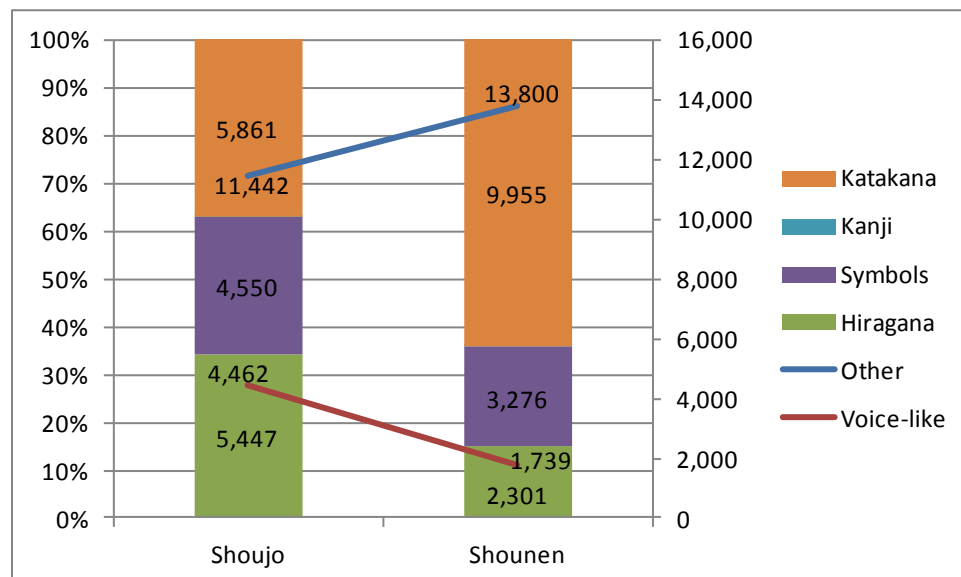
**Table 8:** Text types by manga-genre

Genre	Title	Hiragana		Symbols		Kanji		Katakana		Roomaji		Numbers		Total
Shoujo-manga	<i>Bokura</i>	27,713	56.79%	8,952	18.34%	7,517	15.40%	3,668	7.52%	570	1.17%	380	0.78%	48,800
	<i>Kimi</i>	39,259	50.03%	25,566	32.58%	9,301	11.85%	3,490	4.45%	661	0.84%	195	0.25%	78,472
	<i>Nana</i>	44,793	59.52%	10,725	14.25%	12,833	17.05%	5,173	6.87%	1,313	1.74%	422	0.56%	75,259
	<i>Nodame</i>	29,459	46.92%	13,745	21.89%	8,917	14.20%	8,404	13.38%	1,875	2.99%	389	0.62%	62,789
	<i>Rabukon</i>	37,313	56.38%	13,699	20.70%	8,084	12.22%	5,345	8.08%	1,377	2.08%	358	0.54%	66,176
	Total	178,537	53.86%	72,687	21.93%	46,652	14.07%	26,080	7.87%	5,796	1.75%	1,744	0.53%	331,496
Shounen-manga	<i>DeathNote</i>	43,867	48.48%	14,735	16.28%	21,243	23.48%	6,220	6.87%	3,131	3.46%	1,296	1.43%	90,492
	<i>Gintama</i>	32,414	54.48%	6,705	11.27%	11,782	19.80%	8,381	14.09%	173	0.29%	46	0.08%	59,501
	<i>Konan</i>	39,870	49.91%	18,624	23.31%	14,525	18.18%	6,206	7.77%	406	0.51%	249	0.31%	79,880
	<i>Naruto</i>	25,542	44.86%	14,037	24.65%	9,049	15.89%	8,034	14.11%	77	0.14%	202	0.35%	56,941
	<i>One Piece</i>	34,731	49.59%	17,308	24.71%	10,659	15.22%	6,602	9.43%	574	0.82%	158	0.23%	70,032
	Total	176,424	49.44%	71,409	20.01%	67,258	18.85%	35,443	9.93%	4,361	1.22%	1,951	0.55%	356,846
Grand Total		354,961	51.57%	144,096	20.93%	113,910	16.55%	61,523	8.94%	10,157	1.48%	3,695	0.54%	688,342

One potential reason for this difference in text types is the usage of more *voiced* onomatopoeia in shoujo-manga. Interestingly, 28.06% of *Onomatopoeia* in shoujo-manga directly describe voices, such as laughter, crying or screaming; this is somewhat lower than the percentage of hiragana, and interestingly enough, they

contain 53.29% hiragana. This corresponds to almost half of all of the hiragana seen (43.66%), suggesting that they use hiragana more frequently than the majority non-voice-like onomatopoeia. Indeed, symbols and hiragana make up 84.54% of all such voice-like onomatopoeia, showing that katakana is in the minority. On the other hand, such voice-like onomatopoeia account for only 11.19% of all *Onomatopoeia* in shounen-manga, and they contain 41.00% hiragana (30.99% of all hiragana used). Symbols and hiragana made up 64.18% of all voice-like onomatopoeia in shounen-manga. These results seem to suggest that voice-like onomatopoeia are more likely to be written in hiragana, but that katakana is still more central in shounen-manga than in shoujo-manga. Why voice-like onomatopoeia should favor hiragana is not clear, but one possibility might be the effects that different text types have. Katakana is generally favored for sound-oriented words like onomatopoeia, and it may be that that is being differentiated here. That shounen-manga should use more katakana may also be related to some of these points. As Shibamoto & Schmidt (1996) summarize, hiragana tends to be associated with young, female readers/writers, and stylistically offers a sense of softness of femininity. Katakana, on the other hand, is associated with youth and especially with male readers/writers, and they stylistically offer a sense of modernity and pop culture. As series written for and by men, one might say that the higher use of katakana in shounen-manga might correlate with these differences.

**Graph 3:** Distribution of text within onomatopoeia



**Kanji:** In total, 1,986 different kanji were seen 113,910 times. In comparison, Nozaki, Yokoyama, Isomoto, & Yoneda’s (1996) review of newspaper kanji garnered 4,476 different characters, more than twice as many. While the majority of actual uses of kanji were *kyouiku-kanji*—educational kanji, or those taught in grades 1 through 6 of elementary school—somewhat more than half of the unique characters seen (1,007) were not *kyouiku-kanji*. An additional 140 were on the *jouyou-kanji* (kanji for everyday use) list and 785 were on the *jinmeiyou-kanji* (kanji for use in names) list (Table 9). The *jouyou-kanji* consists of 2,136 characters, including the 1,006 on the *kyouiku-kanji* list. Most of the *jouyou-kanji* not included in the *kyouiku-kanji* are taught in junior high school, although there have been changes in the attitudes of the Ministry of Education towards kanji education, with new reports specifically stressing that changes in the nature of media also mean changes in the nature of kanji and how they should be taught (Bunka Shingi-kai, 2010). There were, however, 82 kanji that were off-list, meaning that they would not be taught formally at all, and *jinmeiyou-kanji*, too, are generally not prioritized within schooling. The coverage rate—that is, the percentage of all kanji seen that are *kyouiku-kanji*—came to 85.73%, which is actually somewhat lower than newspapers (89.84%, Nozaki & Shimizu (2000)). However, in general, the majority of the kanji seen will be familiar to individuals who have completed elementary school education. This issue will come back in Section 3.3 concerning readability.

**Table 9:** Use of kanji by type

Grade	Data type			
	Count		Tally	
1	80	4.03%	23,584	20.70%
2	159	8.01%	29,542	25.93%
3	198	9.97%	20,508	18.00%
4	198	9.97%	11,225	9.85%
5	177	8.91%	7,070	6.21%
6	167	8.41%	5,733	5.03%
<i>Jouyou</i>	785	39.53%	14,998	13.17%
<i>Jinmeiyou</i>	140	7.05%	943	0.83%
<i>Off-list</i>	82	4.13%	307	0.27%
Total	1,986		113,910	

Interestingly, many of the kanji which were off-list were actually ones which may appear familiar. They included “痺” (*shibi-reru*, *hi*; to become numb; 65 examples), “嘘” (*uso*, *kyo*, *ko*; a lie; 20 examples), and “狐” (*kitsune*, fox; 17 examples). These three are all off-list, but are actually fairly common. If one takes hits on

Google to be a sign of how common these kanji are seen in the language, then they actually appear to be more common than some kanji *on* the list. Whereas the above three culled 1,240,000, 138,000,000 and 128,000,000 hits, respectively, jouyou-kanji such as “犠” (*ikenie, gi*; sacrifice; 842,000 hits) and “壤” (*tsuchi, jou*; lot; 1,040,000 hits) resulted in fewer hits, suggesting that they are used less commonly. Even more important is that these off-list kanji were actually used in important concepts within the series. “痺”, for example, is used within *Death Note* in “麻痺” (*mahi*; paralysis), and more specifically, the compound “心臟麻痺” (*shinzou-mahi*; heart failure)—the way in which people whose name has been written in the *Death Note* die, a key plot point. “麻痺” in *Death Note* covered 41 out of 60 of the usages of “痺”. Likewise, of the 17 occurrences of “狐”, 12 appear in *Naruto*, a result of the main character being a fox who has been sealed in a human body—9 as “狐” alone, and 3 in the compound “妖狐” (*youko*), a supernatural fox. Off-list kanji tended to be more common in shounen-manga, which covered 83.6% of all such kanji (Table 10). As will be shown Section 3.3, this could be predicted to negatively affect their readability (that is, make them more difficult to read).

**Table 10:** Kanji usage by genre

Genre	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Grade 6		Jouyou		Jinmeiyou		Off-list		Total
<i>Shoujo</i>	10,934	23.44%	13,484	28.90%	7,636	16.37%	4,332	9.29%	2,091	4.48%	2,012	4.31%	5,662	12.14%	449	0.96%	52	0.11%	46,652
<i>Shounen</i>	12,650	18.81%	16,058	23.88%	12,872	19.14%	6,893	10.25%	4,979	7.40%	3,721	5.53%	9,336	13.88%	494	0.73%	255	0.38%	67,258
Total	23,584	20.70%	29,542	25.93%	20,508	18.00%	11,225	9.85%	7,070	6.21%	5,733	5.03%	14,998	13.17%	943	0.83%	307	0.27%	113,910

**Orthographic symbols:** As the second most common writing type seen, the high percentage of orthographic symbols alone differentiates manga from other genres. However, they are also typified by the types of symbols they do—and do *not*—use. A total of 73 different symbols were found; as can be seen from Table 11, most were seen less than 1% of the time (65). Note that Table 11 lists two types of information: The actual number or *instances* of symbols, including *grouped* symbols, and the number of characters in the corpus they contain (their *weight*). Grouped symbols (“ . . .”, “ — — —”, “ . . . . .”, “ . . . . .”, “ ~ ~ ~”) are symbols that seemed to be working together within the text, or appeared as one in the text, but which were written using a combination of symbols to demonstrate their relative weight or unclear relationship. For example, the 3-point lead ( . . . )—the single most common orthographic symbol, which is visually similar to English ellipsis (...) and which seems to be used to create pauses in the text, such as in Figure 33, Figure 35 and Figure 36—was treated in the corpus as three-individual *nakaguro* “ . ” instead of the one-character symbol “ … ” for three reasons. First, it was difficult to ascertain which form was being used in the originals. Second, not all combinations featured three points, and third, as a symbol which is composed of multiple entities, it has a longer perceptual weight and affects the length of sentences in ways that individual characters do not. Using the 3-point

lead symbol would thus oversimplify the data, but treating the 3-point as a single symbol clearly causes a gap between the numbers of characters seen, hence the notation of Table 11.

**Table 11:** Types of symbols seen

Symbol	Instances		Weight	
・ ・ ・	25,624	30.65%	76,872	53.10%
!	22,876	27.36%	22,876	15.80%
—	8,432	10.08%	8,432	5.82%
?	6,960	8.32%	6,960	4.81%
	6,260	7.49%	6,260	4.32%
――	3,495	4.18%	10,485	7.24%
、	2,930	3.50%	2,930	2.02%
〜	879	1.05%	2,637	1.82%
Symbols under 1% each (in order)				
・	♥	「	」	#
〜	”	“	・ ・ ・ ・	▼
・	。	)	(	・ ・ ・ ・ ・
”	:	=	♪	—
%	『	』	*	#
/	■	+	,	♪
,	☆	[	]	—
I	&	《	》	<
>	④	<	①	(
)	>	>	¥	○
[	]	□	△	→
+	②	③	}	【
】	©	☼	☺	♣
Total number = 73				

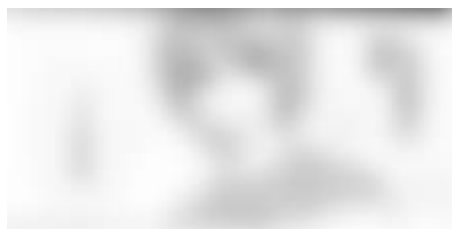
Other common symbols were the exclamation mark (“!”; 22,876) and the enigma (“?”; 6,960). The long-vowel marker bousen (“—”) was also common (8,432), suggesting that it is not used exclusively with katakana: To be used exclusively with katakana, it would have to account for one symbol out of every 4.09 character-strings seen. This is not likely given the length of Japanese words; after the morpheme analysis described in Section 3.4, the average word came to be just 2.18 characters. Additionally, symbols such as what I describe as the *chou-bousen*, or extra-long bousen (3,495), the *namisen* (tilde, 〜) (363), and the *chou-namisen*, or extra-long namisen (〜〜〜) (879), were all seen on a regular basis, and all seemed to be an extension of the bousen’s usages. That is, they often appear to express long vowels, such as in Figure 15, Figure 37 or Figure 38. They also occasionally, however, have roles similar to the 3-point lead: They are occasionally seen sentence-initially, wherein they appear to have no phonetic role except to create a sense of time passing (Figure 32).

**Figure 32:** An example of a chou-bousen being used sentence-initially *RabuKon* (2:40)



There is also a striking *lack* of certain common orthographic symbols. In particular, both the kuten and the touten are rarely seen. Of the 160 examples of the kuten, 94 are found in *Background Text*, i.e., part of *standard* text snippets interplayed into manga such as newspaper articles, and they generally did *not* appear within *Lines*. When they did, they tended to be used in more dramatic ways, such as in Figure 33. Here, Ootani's rejection is made more *final* through the use of the kuten, giving a strong ending to his utterance. On the other hand, while the touten appears to be used regularly on first glance, with 2,930 examples, they are in fact almost exclusively used in *Detective Conan* (96.14%). Akitsuki (2009) also notes this about *Conan*, suggesting that there might be publisher issues at work. While this is hard to prove given the current data, most of which come from the major publisher Shueisha, *Bokura*, also published by Shougakukan like *Conan*, does not follow this rule, suggesting that other issues such as individual authors or magazines may be more relevant.

**Figure 33:** An example of a strongly ended sentence using the kuten from *RabuKon* (2:129)



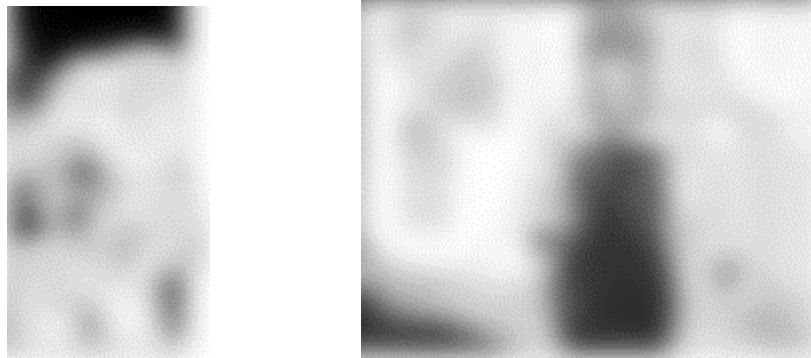
In regard to what they might be substituted with, it would appear that in many cases the kuten is not being replaced at all. 45.67% of all *Lines* finish without any symbolic marking, i.e., they end in kana, kanji,



roomaji or Arabic numerals. Of the remaining 54.53%, 21.81% end in an exclamation mark, 15.36% end in a nakaguro, suggesting that they are 3-point leads ( . . . ), and 13.14% end in enigmas. In regard to the touten, it appears that the space ( ) is often acting in its stead. Figure 34 and Figure 35 are examples of this: In Figure 34, the conjunction *dakara* is followed by a touten, and in Figure 35, it is followed by a space. One might interpret these issues as part of the creation of natural drama. While the exclamation mark has the effect of creating tension, the 3-point lead softens the text by creating a literal pause as one searches for the next string, much as the pauses and breaks common in natural conversation where disfluency is a salient feature (Maekawa, 2003). This also opens up the kuten as a marker of final, dramatic endings, giving it a saliency in those situations that would not be possible if it were always used. The avoidance of the touten, however, might be seen as a way of distancing manga from more formal texts, also contributing to its sense of orality.

**Figure 34 (Left):** The touten to describe pauses after a conjunction in *Detective Conan* (1:12)

**Figure 35 (Right):** The space being used to describe pauses after a conjunction in *Death Note* (3:127)



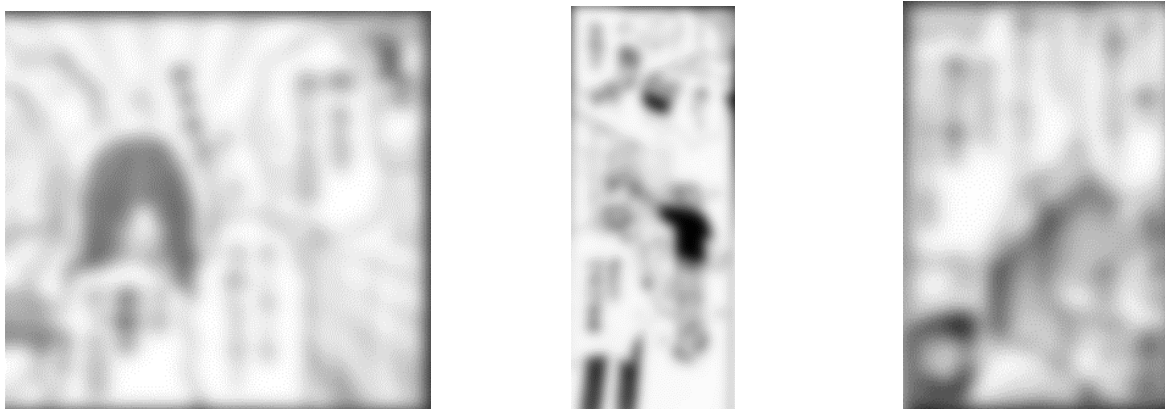
Finally, while some have stated that non-orthographic symbols were common in manga (e.g., Nishimura (2003)), such symbols and *emoji*—smiley faces or other non-phonetic graphic symbols not related to punctuation—were in fact relatively rare. Only three such symbols occurred with any regularity: the heart mark (♥), the double-musical note (♪), and the single-musical note (♫), which appeared 660, 64 and 17 times respectively. The other symbols seen were either punctuation related or had some relationship with phonetics in a broad way. Two commonly seen iconic symbol in manga, sweat marks or anger marks (Cohn, 2010), were rarely seen within speech bubbles, appearing only 207 and 38 times, respectively. Furthermore, they were generally *not* in line with the other text, but rather graphically incorporated. Examples of sweat marks graphically incorporated may be found in Figure 36 and Figure 37, the latter which also features a graphically incorporated

anger mark. Figure 38 shows a more ambiguous type of sweat mark that may be part of the text string. These were found only in *Detective Conan*, and only 14 times.

**Figure 36 (Left):** Example of sweat marks in *Kimi ni Todoke* (1:155)

**Figure 37 (Middle):** Example of sweat marks in *RabuKon* (1:104)

**Figure 38 (Right):** Example of sweat marks in *Detective Conan* (1:18)



**Exceptional kana uses:** While infrequent, all four points raised above (the geminate marker small *tsu* in non-geminate environments; small kana vowels used for long vowels; the bousen used outside of katakana; and the dakuon-ten used with vowels) were seen frequently enough to suggest that they are salient features in so far as they form regularly occurring markers differentiating the text from standard writing. First, a total of 2,998 instances of the small *tsu* being used word or sentence-finally were found. This suggests that approximately 1.15% of all words have this non-standard feature (using the results from the following Section 3.4 on morpheme analysis), making it an infrequent but regularly seen phenomenon.

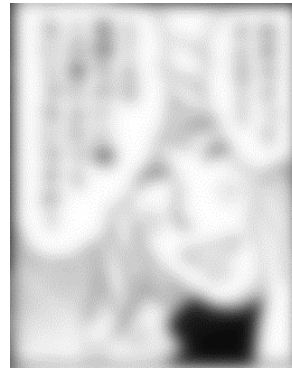
Second, 1,081 examples of small hiragana were found in addition to 1,517 examples of small katakana vowels not following katakana. The use of small kana vowels was more common in shounen-manga generally, with 1,829 examples found; interestingly, however, when shounen-manga used small kana vowels in unexpected environments, they tended to be katakana (79.22%), whereas in shoujo-manga there was a clear tendency to use hiragana (91.16%).

Third, out of the 8,432 bousen, 4,928 were found after characters which were not katakana, e.g., kanji or hiragana. At more than half of all of the examples found, this is clearly more than an incidental usage, but rather a distinct writing style. While many of these examples are of phonemic long vowels, such as writing *sou iu koto* (“That’s what it is”)—where both the long vowel in *sou* and the long vowel in the colloquial form of *iu* as *yu* are phonemic and register semantic differences—as “そーゆーこと” instead of “そういうこと” (Figure

39), many others describe vowels that are being elongated as one of the characteristics of spoken speech (Figure 40). Referring back to Table 11, remember that other non-traditional ways of describing long vowels such as the *namisen*, the *chou-bousen* and the *chou-namisen* were also all seen regularly, suggesting that the trend of writing long-vowels in non-standard forms is very common.

**Figure 39:** Example of a *bousen* being used to express a long vowel in *RabuKon* (1:119)

**Figure 40:** Example of a *bousen* being used for an irregularly elongated vowel in spoken speech in *Nodame Cantabile* (2:65)



Finally, 112 examples of the independent *dakuon-ten* were found, confirming that such forms are seen in manga, while relatively unusual. Normally, the *dakuon-ten* is incorporated into kana characters, such as in “*かゝ*” (*ka*) vs. “*がゝ*” (*ga*): The *dakuon-ten* is added to the upper right corner of the one-syllable character, changing it from the voiceless form *ka* to the voiced form *ga*. Independently-occurring *dakuon-ten* suggest that they are being used with vowels, which do not have a *dakuon-ten* forms. Being naturally voiced, adding the *dakuon-ten* would theoretically offer no additional information. Sadanobu (2008) describes these kinds of usages as *rikimi*, or straining, and argues that in Japanese speech, it is an empathetic way of speaking that can be obligatory. That is, he assigns to them a pragmatic function expressing speaker’s feelings of regret, or that the speaker is describing something they had experienced, similar to the use of creaky voice in English. While it is not clear whether these functions could be assigned to the examples found in manga such as Figure 41, it is important to note that they are thus features associated with spoken speech, and, while they may not have a specific *meaning* associated with them, they do represent particular qualities of sound and possible pragmatic functions.

**Figure 41:** Example of dakuon-ten being used with vowels in *GinTama* (1:90)



### 3.2.5 Study 2 – Use of furigana: Methodology

Of the 10 series in the corpus, all but one (*Nodame*) include furigana on all of their kanji. As such, the use of furigana could be said to be predictable, in that they are repetitive and do not signal alternates readings. As such, most examples are not of interest. As a result, in compiling the corpus furigana were not generally included. Only those which were *not* predictable in nature, i.e., ones which did not offer an established phonetic reading for the characters, were listed. Names, while unpredictable, were not included, given that it is difficult to say there is an acceptable phonetic reading in the first place (see Satou (2007) for more information on names in Japanese). In this second study, I compiled a list of all examples using non-established furigana, and classified their role in the text into one of five categories: *Emphasis*, *Censoring*, *Ateji* (non-established readings), *Pronunciation* and *Explanatory*. *Ateji* were further divided into deictic, *gairaigo* (loan words) and synonyms; *Pronunciation* into slang, foreigner talk, foreigner languages, katakana and letters/symbols, and *Explanatory* into Japanese translations and definitions. Details on each category are in Table 12; examples may be found in Figure 42 to Figure 53.

**Table 12:** Types of furigana

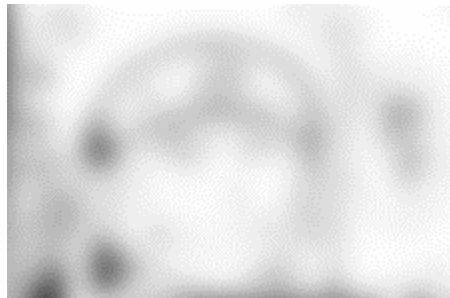
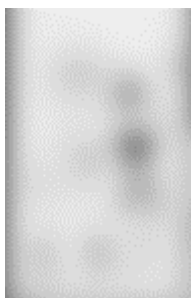
Major	Definition	Minor	Definition	Examples (Word - Furigana - Source)		
<i>Emphasis</i>	Symbols giving dramatic emphasis			アレ (are - that)	・・・	<i>Bokura</i>
<i>Ateji</i>	Readings which are not generally recognized for an individual kanji or compound	Deictic	Replacement of regular reading with words clarifying who/what/where the word is referring to	真選組 ( <i>Shinsengumi</i> [proper noun])	ここ ( <i>koko</i> - here)	<i>GinTama</i>
		Gairaigo	Non-normal foreign loan readings of kanji	銃 ( <i>juu</i> - gun)	ピストル ( <i>pistoru</i> - pistol)	<i>OnePiece</i>
		Synonyms	Replacement of regular reading with a word with a similar meaning	他人 ( <i>tanin</i> - stranger)	ひと ( <i>hito</i> - person)	<i>Bokura</i>
<i>Censorship</i>	Onomatopoeia or symbols marking censorship			――	びー―― ( <i>pii</i> - bleep)	<i>RabuKon</i>
<i>Pronunciation</i>	Regular readings with slight alterations	Letters/symbols	Readings assigned to roman letters or orthographic symbols	HR	ホームルーム ( <i>hoomuruum</i> - homeroom)	<i>Kimi ni Todoke</i>
		Slang	Colloquial pronunciations of words	強 ( <i>tsuyo</i> - strong)	つえ ( <i>tsue</i> - strong)	<i>Detective Conan</i>
		Foreign languages	Phonetic readings of foreign words	Nein	ナイン ( <i>nain</i> - nein)	<i>Nodame Cantabile</i>
		Katakana	Katakana where hiragana would normally be used	20歳 ( <i>hatachi</i> - 20 years old)	ハタチ ( <i>hatachi</i> - 20 years old)	<i>Nana</i>
		Foreigner talk	Changes in reading or writing type due to speaker's being foreign	今週 ( <i>konshuu</i> - this week)	コンシュウ ( <i>konshuu</i> - this week)	<i>GinTama</i>
<i>Explanatory</i>	Furigana elucidating the meaning of the words they are assigned to	Japanese translation	Japanese translations of foreign terms	カプチオーソ ( <i>kapuchirooso</i> - capriccioso)	気ままに気まぐれ ( <i>ki-mama ni ki-magure</i> - as one feels)	<i>Nodame Cantabile</i>
		Japanese definition	Definitions of difficult words	殉職した ( <i>junkyou-shita</i> - martyred)	(死んだ) ( <i>shinda</i> - died)	<i>Naruto</i>

**Figure 42 (Left):** Example *Emphasis* furigana in *Bokura ga Ita* (1:161)

**Figure 43 (Middle):** Example of *Censorship* furigana in *RabuKon* (1:40)

**Figure 44 (Right):** Example of *Explanatory* furigana (Japanese translation) in *Nodame Cantabile* (1:34)

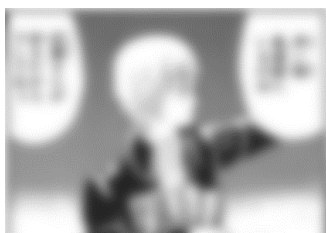
**Figure 45 (Right):** Example of *Explanatory* furigana (Japanese definition) in *Naruto* (2:16)



**Figure 46 (Left):** Example *Ateji* furigana (deictic) in *GinTama* (3:41)

**Figure 47 (Middle):** Example of *Ateji* furigana (gairaigo) in *One Piece* (1:146)

**Figure 48 (Right):** Example of *Ateji* (synonyms) in *Bokura ga Ita* (1:154)



**Figure 49 (Far left):** Example of *Pronunciation* furigana (letters/symbols) in *Kimi ni Todoke* (3:64)  
**Figure 50 (Mid-left):** Example of *Pronunciation* furigana (slang) in *Detective Conan* (1:108)  
**Figure 51 (Middle):** Example of *Pronunciation* furigana (foreign languages) in *Nodame Cantabile* (2:100)  
**Figure 52 (Mid-right):** Example of *Pronunciation* furigana (katakana) in *Nana* (3:150)  
**Figure 53 (Far right):** Example of *Pronunciation* furigana (foreigner talker) in *GinTama* (1:115)



### 3.2.6 Study 2 – Use of furigana: Results

**Table 13:** Distribution of furigana

Genre	Stress		Censoring		Explanatory		Ateji		Phonetic		Total		Characters	% characters
<i>Shoujo</i>	55	41.67%	1	0.76%	3	2.27%	44	33.33%	29	21.97%	132	100.00%	521	0.16%
<i>Shounen</i>	180	34.88%	1	0.19%	1	0.19%	197	38.18%	137	26.55%	516	100.00%	1,642	0.46%
Total	235	36.27%	2	0.31%	4	0.62%	241	37.19%	166	25.62%	648	100.00%	2,163	0.31%

% characters=Number of characters featuring unusual furigana

The distribution of all of non-predictable furigana was tallied by genre, with the results in Table 13. A total of 648 non-predictable forms were found, with *Ateji* (37.19%) forming the majority of all unpredictable furigana, followed by *Stress* (36.27%); a large part thus did not register reading changes at all. *Phonetic* types were also common (25.62%), with these three groups covering over 95% of all examples. *Explanatory* and *Censoring* furigana were exceptionally rare, having been seen only 4 and 2 times, respectively. In general, it appeared that shounen-manga used such irregular furigana more often, particularly when one considers that 41.67% of those seen in shoujo-manga were actually stress. This proved to be not quite statistically significant using a *t*-test between shoujo-manga and shounen-manga in regard to the number of characters with non-predictable furigana and the overall number of characters ( $t(8) = 1.93, p = .0894$ ). However, it does suggest that the data are trending, and additional data in the form of more series may show a clearer difference. Note that such non-standard furigana were featured on 2,163 characters throughout the text, or 0.31% of all characters. This is a very small

percentage overall, but with 5,289 pages in total, one would still see approximately one such example on every two pages, meaning they are a regularly occurring and likely salient feature.

Looking in more detail at those that are related to changes in readings, it is interesting to note that *Ateji* were made up of 82 deictic usages, 93 gairaigo, and 66 synonyms. In the case of deictic usages, the furigana could be thought of as playing an important clarifying role. That is to say, like the example given in Figure 30 or Figure 46, they often make it obvious about whom or what the speakers are referring to. Such deictic usages could primarily be seen as pragmatic efforts on the parts of authors, and are similar to *Emphasis* or *Explanatory* furigana in that regard as they play a meta-linguistic role in the text: While they may not change the meaning *per se*, they offer secondary information into how the text should be processed. On the other hand, gairaigo and synonyms fall more in line with the examples that Wilkerson & Wilkerson (2000) describe as *poetic*, in that they create polysemy in the text by their playing off the kanji with the non-matched readings. Readers are forced to consider *both* the word represented by the kanji themselves and the phonetic word they are being made to render, making for multiple levels of meaning within the text. *Phonetic* examples were primarily made up of slang (42) and letters/symbols (84), meaning they were, too, mostly clarifying in nature as they largely served to show the actual phonetic readings being sought. Foreign words (2) could also be seen as falling under this rubric. However, foreigner talk (10) and katakana (30) could be seen as falling under the meta-linguistic rubric, as they both demonstrate *regular* readings, but add something additional through their usage of script choices. In the case of foreigner talk, katakana is being used to demonstrate that it is not standard Japanese, and is consistent with the replacement of hiragana with katakana for function words throughout such speaker's utterances (Figure 53). The motivation behind using katakana in the other examples is less obvious, but it may be that they are playing with the affects that katakana have, mentioned previously in Section 3.2.4.

### **3.2.7 Discussion: Manga as spoken language, seen through orthography**

With the establishment of the orthography of manga now clearly established as unusual, one is now forced to consider, what is the effect of using such non-standard styles, and what is aimed to be achieved by such usages? A theme that has come up repeatedly, particularly in the question of what symbols are and are not used, has been that of giving manga an orality that would make it more associated with the spoken speech it emulates. This is clearly an important part of the issue, and some of the points raised, such as the high percentage of hiragana, could be related to it. That is, because manga are meant to be taken as *spoken* speech, they often emulate the

aspects of spoken language that are edited out of writing, such as non-standard forms, interjections and examples of misspoken speech. Whereas spoken speech can use intonation, volume and other factors to add paralinguistic meaning, written speech has orthography as its tools for adding similar nuances (Maekawa, 2008, pp. 27-29). Thus, just as Maekawa (2008, p. 29) suggests that keitai-meeru use emoji to add the paralinguistic information of spoken speech that is lost in writing, one can similarly argue that manga uses symbols to similarly bring the text closer to spoken speech.

Symbols are helpful in creating these effects, as they tell us not so much what *phonetic sounds* are being heard, but rather *how* they are being said: Do speakers pause here? Add a space, or a 3-point lead. Have they misspoken? Break the text up with a small *tsu*. As I mentioned previously, avoidance of certain symbols such as the touten and the kuten could also be interpreted as an attempt to move away from the more rigid impression of written language that they might supply. The relatively low usage of kanji could also be reflected in these points: Because kanji are not phonetically transparent, descriptions of phonetic changes are more suited to hiragana or katakana, which are highly transparent. When one wants to use kanji to represent irregular changes, authors must resort to manipulating the kanji themselves, such as with the use of non-standard furigana.

The effect of using non-standard orthography is also related to the greater question of developing a manga *style*. Tada (2001) describes the use of hiragana and dialectal and vernacular elements in keitai-meeru as a way of making them “keitai-meeru-esque” by giving them the keitai-meeru “mold” (*kata*), or in linguistics terms, a register. Similarly, the unique distribution of the different scripts, the abundance of unusual orthographic symbols and the usage of non-standard kana styles seen here could also be seen as a way of making them more *manga-like*. As I noted previously, the lack of kuten and touten also distance manga them from other styles of writing, while the usage of spaces creates further distance in the same direction. In this way, the usage of unusual orthography has a double effect, on the one hand making manga seem more like the dynamic conversational language it emulates, while at the same time distancing itself from more formal types of writing. The development of a manga style in this way also creates the effect of the text *as* manga—that is, it can be seen as contributing to how manga established its own unique style and genre of writing.

Note that Satake (1980) has also mentioned similar changes in the writing styles used by young people in letters and other personal writings, which he describes as the shin-genbun’icchi-tai, or the new style of unification of spoken and written language. One of the major changes in Japanese in the last 150 years has been the movement towards a written language that was more similar to that of spoken language. Before the Meiji period, written language was highly stylized, with government documents often in *kanbun* (Chinese writing) and even literary texts in a mixed, classical Japanese style. However, the creation of the modern state led towards the modification of Japanese to fit its new needs; additionally, pressure to emulate the Western style novels of



the realism school led to new styles of writing that were more similar to the language being spoken at the time. (See Twine (1991) for a thorough review of these issues.) However, as Nishimura (2011, p. 101) reviews, a large gap between spoken and written language is still considered typical of the Japanese language, hence one of the criticisms of *keitai-shousetsu* as being not of literary value. In comparison, Satake's *shin-genbun'icchi-tai* is characterized by the use of many unique characters and a more colloquial and personal style, which Satake argues is more expressive and emotional. These are also similar to what I have suggested for manga. Orthographic characteristics thus help relate manga to other styles of writing, and locate it within written Japanese language.

It is also interesting to note that even as these characteristics combine to create a general manga style, there were still individual differences between *shoujo-manga* and *shounen-manga*, with *shounen-manga* seemingly using more kanji and unusual furigana, both points that will affect their readability (Section 3.3) and the morpheme analysis results (Section 3.4). This suggests that even within manga, orthography can be further manipulated to create differences. It also points to the need for caution before making too many generalizations. Indeed, it should be noted that the trends seen here are not necessarily set in stone. In addition to differences between genres, individual series themselves are also not necessarily consistent in the orthographic styles they utilize. An interesting example is the way in which the raising of the diphthong /ai/ in masculine speech to [ei] are expressed, particularly with the common negative form *nai* (/nai/→[nei]), an issue I previously mentioned in Chapter 1 in regard to Example 2.

A total of 6 different forms were seen throughout all the series, excluding two additional forms of a further truncated form ([ne]): “ねえ”, “ねえ”, “ねー”, “ねエー”, and “無え”. As there is no standard way to write these forms, it is natural to see some variation, leading to this proliferation of forms. However, what is more interesting is while most series had a favored form that was used for more than half of all of examples, none of the series used any one form exclusively (Table 14). While the trend towards using a preferred form seems to suggest that there might be some internal trends towards standardization, this also points at the essential inconsistencies involved in non-“standard” texts. In this sense, while the orthographic characteristics of manga may have many large commonalities such as the use of the bousen, hiragana, spaces and the lack of *kuten* and *touten*, many of the smaller points are not so clear cut. Future research would be well served by looking at how individual series, genres and authors interact along with these general trends to find balance between their own individualities and consistency as manga as a genre.

**Table 14:** Forms and distribution of [ne:]

Genre	Series	Forms of [ne:]						Total
		ねー	ねえ	ねえ	ねエ	ねエー	無え	
Shoujo-manga	<i>Bokura</i>	28	1	26				55
	<i>Kimi</i>	48		2				50
	<i>Nana</i>	1	94	24				119
	<i>Nodame</i>	57	8	5				70
	<i>RabuKon</i>							0
Shounen-manga	<i>OnePiece</i>	3	4	4	210			221
	<i>DeathNote</i>	14						14
	<i>GinTama</i>	247	4		68			319
	<i>Konan</i>	34		2				36
	<i>Naruto</i>	26	3	1	14	13	1	58
Grand Total		458	114	64	292	13	1	942

## 3.3 Readability and the skills necessary for reading manga

### 3.3.1 Readability and manga

One of the major questions concerning manga remains how easy they are to read. As a whole, many assume that comics in general are easy to read, an attitude which has mixed consequences. Because of these assumptions, there exists a tendency to look down upon manga and comics as not requiring “real” reading (Ito, 2000), which has led to a reluctance amongst some to use them for more serious education purposes (detailed in Mori (2007)). At the same time, some argue that because they are easy to read, they are in fact a *good* resource for the classroom. Buono (2013) notes that for adult students of English as a second language, comics can be a positive resource thanks to their relatively simple vocabulary and visual aspects.

One thing that seems consistent is the tendency to think of manga as being easy to read, especially given the extra information supplied by their visual images (reviewed in Murakami & Bryce (2009)). Yet as Ingulsrud & Allen (2009) point out, the situation may not be that simple, as even seemingly *easy* series often contain very high-level vocabulary. It remains, however, largely speculation due to the lack of systematic research assessing the *readability*—or the ease of reading—of comics. This could be partially due to the lack of effective data sets, but it is also equally related to the fact that readability as a concept is actually largely new. Much of the research on readability has focused on the use of readability formulas, which Klare (1974) describes as a formula which uses linguistic variables in writing to give an index of the text’s probable difficulty. Readability formulas are meant to be predictive, in that “no actual participation by readers is needed (Klare, 1974, p. 64).” Such formulas have been developed for English since the 1920s, but theoretical work on what they really “mean” came only later (DuBay, 2004). For Japanese, formulas themselves have not been as well developed, with the result that there is far less work on readability in Japanese compared with English.

As Sato, Matsuyoshi & Kondoh (2008) note, while some of the lack of research on Japanese readability may be related to lack of interest, it is also been related to a lack of easy-to-use, readily available tools. While the corpus utilized here is a convenient solution to the earlier problem of a lack of appropriate data, recent developments in the field of readability in Japan have allowed for more objective ways of assessing the levels of texts, especially with new tools available. In recent years at least two such tools—one which Sato et al. designed and described, and another by the Readability Research Laboratory (2010) at the Nagaoka University of Technology—have become readily available. With these new tools in hand, assessing the readability of manga may prove key to understanding their nature as linguistic texts. In particular, through examining the readability

of manga, one would be better equipped both for assessing how they compare to other texts, especially in regard to their target audiences, but also how appropriate they might be for other purposes, such as use within the study of Japanese as a second language.

Within this section, I use the corpus data to assess the readability of manga using tools the developed by Sato, Matsuyoshi & Kondoh (2008) and the Readability Research Laboratory. In conducting this study, I focused on two major questions, the first concerning the reading level of the series and their respective audiences, and the second concerning differences in genre. In regard to the first question, while most of the series in the corpus were printed originally in traditionally shounen-manga or shoujo-manga magazines (*shounen-muke komikkushi*, *shoujo-muke komikkushi*), as defined by the Japan Magazine Publishers Association (JMPA) (2012), with the remaining two being published in women's comic magazines (*josei-muke komikkushi*), recent years have seen a leveling out of readership differences, with more readers reading out of their ostensible gender and age frames (Itou, 2005). Even the two series *not* falling under the shounen/shoujo-manga rubric as set by the JMPA are treated in most other regards as shoujo-manga, one having been awarded the Shougakukan Manga Award for shoujo-manga in 2002 (*Nana*), and the other the Kodansha Manga Award for shoujo-manga in 2004 (*Nodame Cantabile*). One might hypothesize thus that there will be less of a clear-cut relationship between series' assumed markets and their actual readability, i.e., they may be *more* difficult than assumed given the influx of older readers.

Second, shoujo-manga and shounen-manga differ greatly in terms of the types of plots they favor. All of the shoujo-manga series in the corpus feature everyday settings, generally in school (*Bokura*, *Nodame*, *KimiTodo*, *RabuKon*), or work (*Nana*). On the other hand, all of the shounen-manga series are extraordinary settings. *Naruto* features ninjas; *GinTama* samurais; and *One Piece* pirates. While *Death Note* and *Conan* feature everyday settings, they are greatly spiced up by involvement of the police/government along with the supernatural (*Death Note*) and science-fiction (*Conan*). Given the everyday settings of shoujo-manga, one might suppose that they tend to use less-technical language more similar to conversational speech, whereas shounen-manga would use more technical language to describe the unusual phenomena they depict. Due to the high association of conversational speech with native Japanese words and the use of kanji-based words for more complex and abstract concepts (Maynard, 1998), as well as to the association between kanji and readability in Japanese (Nishimura, 2011), this leads to the hypothesis that shoujo-manga will rate more readable than shounen-manga.

### 3.3.2 Methodology

I utilized two tools to assess the reading level of the corpus: first, Obi2: Japanese Readability Analyzer, developed by Sato, Matsuyoshi & Kondoh (2008), and second, the Japanese Readability Test (*Nihongo Riidabiritii Sokutei*) developed by the Readability Research Laboratory headed by Hideko Shibasaki at the Nagaoka University of Technology and partially described in Shibasaki & Tamaoka (2010). Obi2 is a free program, available online and for download through Satoshi Sato Laboratory's (2008) website, which utilizes a modeling tool described in Sato, Matsuyoshi & Kondoh (2008) to determine the readability of texts. Using a textbook corpus consisting of 1,478 sample passages from 127 textbooks from elementary school, junior high school, high school, and university, Sato et al. were able to develop a method of measuring text readability using character-unigram models, using six factors: the average number of characters per sentence, the ratio of *touten* to *kuten*, and the average number of *roomaji* and symbols, hiragana characters, kanji characters and katakana characters per run, respectively. Their method overcomes the problem of defining individual words by focusing on *runs*, that is, strings of character types, and sentences themselves. While the focus on *sentences* as defined by *touten* and *kuten* may seem to deter the use of Obi2 for non-standard text types such as manga which do not utilize such symbols, Obi2 was designed with priority given to incomplete sentences to accommodate its usage to a variety of texts. Indeed, with a high-level of performance over many text types ( $R > 0.9$ ), the measurement of readability of Sato et al. appears to be a reliable way of assessing the reading levels of texts. In terms of Obi2's production, it assesses texts as being of one of 13 levels, 1 to 12 each being one grade level, and 13 being university or *general*.

The Japanese Readability Test developed by Hideko Shibasaki with Shin'ichirou Hara and Jae-Ho Lee and available online at the Readability Research Laboratory's (2010) website uses similar methods to assess difficulty. They used a corpus consisting of all of the texts for elementary school Japanese language (*kokugo*) classes from Mitsumura Tosho Publishing, Tokyo Shoseki, and Kyoiku-Shuppan, and all of the junior high school Japanese language texts from Mitsumura Tosho Publishing, Tokyo Shoseki and Sanseido Bookstore. In terms of the factors considered, they included the average number of characters *per sentence*, the use of the different scripts, and the types of words seen (Sino-Japanese *kango*, native Japanese *wago*, etc.). They also used the number of predicates per sentence, arguing that more predicates suggest more complicated grammatical structures. The Japanese Readability Test categorizes texts into 9 levels grades correlating to elementary school 1 to 6 and junior high school 1 to 3. It also marks the kanji used into 11 grades, the first six correlating to the elementary school grades 1 to 6, and grades 7 to 11 to levels of the *Nihon Kanji Nouryoku Kentei* (Kanji-proficiency Examination; *KanKen* below). More specifically, 7 correlates to *KanKen L4*, or in-course junior high

school level; 8 to *KanKen L3*, or junior high school graduate level; 9 to *KanKen L2*, equivalent to a high school graduate's level; and 10 to the *KanKen PreL1* and 11 to the *KanKen L1*, both the college-graduate or *general* level, with the latter including approximately 3,000 additional characters (Nihon Kanji Nouryoku Kentei Kyoukai, 2012)

For the purposes of this study, I restricted the data to *Lines*. This was done as *Lines* form the majority of text; are the text type most likely to be in sentence-form; and are the text type given the most consideration during the reading process (Allen & Ingulsrud, 2007). This was also done in order to offer a comparison with the data presented here on the morpheme analysis of *Lines*. Readability in this sense is of course a limited function, in that it does not take the text in manga at a whole, including both the other types of text but also the images themselves. Aside from removing the half-width ampersand (&) denoting line breaks, the text was left as-is in order to best approximate the readability of the text in its original form. Each series was individually processed, leading to (1) an overall readability score on the Obi2 and the Japanese Readability Test scales, and, for the Japanese Readability Test only, breakdowns of (2) the characters used by writing type and (3) the kanji seen by educational level. Using the readability results, I also compared individual series by genre in order to ascertain whether there were any significant differences. For this, I used an unpaired t-test, conducted twice using both the data obtained from Obi2 and the data obtained from the Japanese Readability Test.

### 3.3.3 Results

***Readability and Grade Level:*** The readability results are presented below in Table 15. In general, the series averaged relatively low, or 6.05 on the Japanese Readability Test's scale and 5.9 on the Obi2 scale. According to Obi2's analysis, this would make manga *very easy*, and generally someplace between the fifth- and sixth-grade reading level. This does not seem to differ very much from the aimed audience of the series in the corpus, suggesting that they are not being marked up in an immediately obvious linguistic form for their more diversified audiences, not supporting the first of the above hypotheses.

***Readability and Genre:*** To test the differences of the readability scores for shoujo-manga and shounen-manga, I conducted two student's t-tests between their respective scores, one using those from the Japanese Readability Test and one using those from Obi2. Both the t-test using the Japanese Readability Test results and the t-test using the Obi2 results showed that the differences between shoujo-manga and shounen-manga were significant,

with shounen-manga rating more difficult: Table 16 and Table 17. This supports the second of the two hypotheses above, that is, that shounen-manga would be more difficult to read than shoujo-manga.

**Table 15:** Comparison of the readability test results

<i>Genre</i>	<i>Series</i>	<i>Japanese Readability Test Results</i>	<i>Obi2 Grades</i>	
		<i>Grade</i>	<i>Grade</i>	<i>Ease</i>
Shoujo	Bokura ga it	5.71	5	Very Easy
	Kimi ni todoke	4.05	5	
	Nana	5.29	5	
	Nodame Cantabile	6.52	6	
	RabuKon	4.98	4	
	<u>Average</u>	<b>5.31</b>	<b>5</b>	<b>Very Easy</b>
Shounen	Death Note	7.15	8	Somewhat easy
	Detective Conan	6.37	6	Very easy
	Gin Tamashii	6.88	6	
	Naruto	7.27	8	Somewhat easy
	One Piece	6.24	6	Very easy
	<u>Average</u>	<b>6.782</b>	<b>6.8</b>	<b>Very Easy</b>

**Table 16:** Comparative genre statistics using the Japanese Readability Test

**Table 17:** Comparative genre statistics using Obi2

<b>Obi2 t-test</b>			<b>Japanese Readability t-test</b>		
<i>Group</i>	<i>Shounen-manga</i>	<i>Shoujo-manga</i>	<i>Group</i>	<i>Shounen-manga</i>	<i>Shoujo-manga</i>
Mean	6.8	5	Mean	6.782	5.31
SD	1.1	0.71	SD	0.4601	0.9112
SEM	0.49	0.32	SEM	0.2058	0.4075
N	5	5	N	5	5
<i>t</i> = 3.0870	df = 8	SED = 0.583	<i>t</i> = 3.2246	df = 8	SED = 0.456
<i>p</i> = .0150*			<i>p</i> = .0122*		

### 3.3.4 Discussion

In general the manga in the corpus all ranked at a relatively low level of difficulty. With the series averaging between the fifth and sixth grade reading levels, this suggests that they are well within the reading levels of their target audiences, i.e., elementary and junior high school children. This goes against the hypothesis that shoujo-manga and shounen-manga may actually be reading at a higher level given their diversified audience. In fact, editors and magazines themselves have a large amount of control and say into the process, with editors having great say over stories themselves, in addition to the actual physical presentation of manga (Kinsella, 2000). Given the current target audiences of the magazines, it is reasonable to assume a certain level of conservative norms towards ease of reading for younger audiences. In the future, it might be of interest to research whether there have been any actual historical changes in the levels seen, through a comparison with older series. Meanwhile, however, it seems reasonable to argue that while current series are appealing to a wider audience, it is not necessarily through their use of more *difficult* or *advanced* language, at least in terms of their use of kanji, one of the major markers of readability.

To give these results more context, it is worth comparing the data with the results from other genres of texts. In particular, Nishimura (2011) used Obi2 to test the readability of 7 popular keitai-shousetsu and 88 conventional novels from the monitor version of the *Balanced Corpus of Contemporary Written Japanese* compiled by NINJAL. As Nishimura (2011, p. 89) describes, keitai-shousetsu have been viewed as being essentially *pulp*, meaning they are of less quality than conventional paper-printed novels, with common criticisms including that its language is “immature”, with only easy kanji and that they deviate from standard stylistic conventions through the use of non-standard manipulations like unnecessary blank space. As Coates (2010) notes, there are many similarities between manga and keitai-shousetsu, and these criticisms strike as more than a little familiar. Nishimura’s results showed that keitai-shousetsu averaged at the 7.71 grade level, and conventional novels at 8.2; thus, keitai-shousetsu only ranked somewhat lower than conventional novels. The results do place them at a higher reading level than manga, but narrations and quotations showed differing levels. That is, narrative texts—equivalent in function to *Narration*—read at a higher level in both keitai-shousetsu and conventional novels (7.71 and 8.7, respectively) than quotations—equivalent in function to *Lines*—which ranked at 5.86 and 6.68, respectively. Since only *Lines* were subject to the readability test here, it would seem more appropriate to compare manga to the data for quotations only. This would place them at very similar levels: Shoujo-manga read somewhat *lower* (5), but shounen-manga would read somewhat *higher* than keitai-shousetsu (6.4). While this would place them still somewhat lower than conventional novels, the gap would be far smaller.



These results, however, are largely predictable given the previous discussions on orthography and manga style as a kind of simulated spoken speech in Section 3.2, because both algorithms put great value on the ratio of kanji to kana and other symbols. Kanji education is strictly regulated for grade level within school (Gottlieb, 1995), meaning that there is a direct relation between school grades and kanji levels. Additionally, as Nishimura (2011) argues for *keitai-shousetsu*, texts such as manga which rely less upon kanji in favor of non-standard forms will accordingly be assessed as less difficult than texts more kanji-oriented. The pressures in manga to describe *colloquial* or *spoken* speech may also push it towards using non-standard forms of writing that are assessed as being easier, a point which is reflected not just in manga, but in the lower reading levels Nishimura (2011) reports for quotations.

Note, however, that the results also speak to the differences within manga, thus supporting the second hypothesis that shounen-manga would be more difficult. All of the shounen-manga read at least the grade 6 level on both the Japanese Readability Test's scale and Obi2's scale. On the other hand, only one shoujo-manga series, *Nodame*, one of two series in a more adult magazine, read at the grade 6 level on either scale, with the others all reading between levels 4 and 5. It is interesting to note that shoujo-manga have often been described as being characterized by their *sophisticated* linguistic structure (Takeuchi (2005), Ôtsuka (1994)), this is not maintained in so far as it is measured by difficulty. Paradoxically, while the action-oriented stories of shounen-manga may make one theorize that they should be easier, a relatively sophisticated vocabulary may be necessary in order to maintain and express their unique worlds: Creating worlds of ninjas, detectives, supernatural powers, alien samurai and pirates all require some understanding of the rules and organization of each locale, which, with the trend to fit more information into speech bubbles and out of narration (Ingulsrud & Allen, 2009), would all be detailed in *Lines*, a point which is likely also correlated to the higher percentage of *Lines* in shounen-manga and their average length (shoujo-manga: 13.52; shounen-manga: 15.53). In fact, the only one of the shoujo-manga series that similarly has a more *specialized* environment is *Nodame*, coincidentally the only one which ranked at the grade-6 level, and which, dealing with a music conservatory, requires a great deal of explanation to general readers. The everyday situations which characterize the shoujo-manga in this corpus may, thus, limit the environments in which more sophisticated vocabulary would be seen.

Of course, this discussion has simplified the issue of the overall manga reading experience. Reading manga successfully requires readers not just to read the linguistic content, but to be able to properly parse the icons used and move through the images. As both Nakazawa (2005) and Ingulsrud & Allen (2009) show, manga literacy skills must be acquired and developed, and cannot be taken for granted. The skills required to move through the pages also affect how language is processed. As mentioned in Chapter 2, Ôtsuka (1994) argued that certain kinds of out of speech-bubble, monologue-like text-types which had been found commonly in shoujo-

manga actually fell *out* of popularity due in part to people's perception that they were difficult to read: They were difficult to follow, with few clear leads between texts. Such text types still exist to some degree, in *Thoughts*, *Background Lines/Thoughts* and *Comments*. It is interesting to consider that this is more related to how the language is presented, than to the actual vocabulary seen. That is, the special relationship of the text within manga may, aside from the question of the impact of the visuals on understanding, affect how difficult people find they are to read. In addition, I have also simplified the situation here by limiting the data to *Lines* alone, and I have not touched upon the question of furigana. Since furigana takes away from the ambiguity of kanji for even the most unskilled of readers, their usage clearly affects how easily people will be able to read them.

This does not mean that these results are without value. While understanding the overall process of manga literacy will be essential to properly judging manga's readability, the readability scores described here could be useful for comparing them with other, more traditional texts when accompanied by the caveat that it is not a complete assessment. First, the results found here may also prove helpful in the Japanese language classroom, where manga has been increasingly utilized as a way of bringing in students' interests. As Chinami (2005) notes, one of the essential conditions for using manga in the classroom is understanding their characteristics, possibilities and limitations. With the lack of research on language in manga, this is something that is unfortunately less fulfilled than one might desire. Even given the above caution, knowledge of how easy manga may be would be extremely useful: A fifth-grade reading level may not seem very high for 18-year old native-speakers, but it would be a very different thing for 18-year old second-language learners in a beginner's Japanese class to read them. These data would allow for better planning, and for more effective usages of manga in the classroom. Secondly, the genre differences discovered here also give interesting perspective on the theme of re-defining shoujo-manga and shounen-manga. That shoujo-manga ranked easier to read in and of itself may be of only limited validity, but the differences that led to these assessments—such as the use of kanji and symbols—are points that may speak to more essential differences in how shoujo-manga and shounen-manga are constructed and presented. The results from the readability studies thus give perspective not just on the skills needed to read manga, but also on how manga are situated as a medium.

## 3.4 A Morpheme Analysis of Manga

### 3.4.1 Japanese morpheme analysis

This section introduces a *morpheme analysis* of the vocabulary seen in the corpus, with the goal of clarifying manga's lexical characteristics. While a specific definition of morpheme analysis will be given in due time, it is first important to consider how data from corpora is actually assessed, and how that data is obtained. With the development of corpus linguistics, a major question that has had to be broached is how to deal with the corpora themselves. In other words, now that the data has been compiled, how should one work with it? The key that has emerged is that of word-frequency, meaning the number of times an individual word appears in the corpus. By comparing the frequencies of multiple corpora, or the frequencies of different words within a given corpus, it is often possible to make striking observations about texts. Frequency lists have proven useful in finding the differences between corpora of various genres (Rayson & Garside, 2000), and they have also been helpful for practical purposes like compiling word lists for language learning (Coxhead, 2000). They have also been used in language learning to show differences between native and non-native speakers' use of vocabulary (Shirato & Stapleton, 2007).

What word frequency can tell us, however, is not just limited to questions of grammar or vocabulary. They can also give insight into larger questions about culture and society. Leech & Fallon (1992) compared the distribution of a large variety of words in the Brown Corpus (the major American English corpus) and the Lancaster-Oslo/Bergen Corpus (LOB; the major British English corpus) which could potentially offer non-linguistic contrasts only explicable by differences in subject-matters themselves, such as the use of “coffee” vs. “tea” rather than “movies” vs. “film”. Here, the former requires a difference in topic, whereas the latter refers to the same theme, even though they are linguistically different in form. Their results showed that there were not only strikingly clear differences in the kinds of vocabulary seen concerning things like sports, transportation and politics—all issues which speak of differences in culture and society—but also in regard to the abstract concepts and conjunctions that were seen. This led them to suggest that such abstract lexical items might, in addition to purely linguistic concerns, elucidate cultural differences between the United States and Britain.

Word frequency is clearly a powerfully informative statistic, and it is precisely data in this form—the number of personal pronouns and sentence final expressions—that will form the core of the examination on characters' speech patterns in Chapter 4. What is not as clear, however, is how word frequency should be calculated. Obtaining word frequency data requires corpora to be divisible into individual words which can then

be counted. This is not a great problem in English, where word boundaries are (generally) labeled by spaces: An ad-hoc definition could be *a chunk of text between two spaces*. However, as Furui, Maekawa & Ishihara (2000) have noted, the definition of words is *the* problem for Japanese corpus linguistics. Of course, ad-hoc definitions like the one I just formulated are problematic in English, too; John Sinclair, one of the founding-fathers of corpus linguistics, pushed strongly for corpora to *not* be overly-annotated, showing in (1991) that traditional parts-of-speech assignments are often misleading. But wordhood is particularly difficult in Japanese as there is no obvious ad-hoc definition: Because Japanese does not include spaces, or any definitive way of defining a word aside from possibly marking boundaries by script types, there is no standard way to divide a sentence. Searching for specific targets in corpora is one way of bypassing this problem, but while conducting individual queries for specific words can lead to fruitful results, they are inherently limited in scope. They only allow one to look for what one has been able to imagine, and they do not offer any information about the rest of the text. This severely limits their statistical power, as it is impossible to assess the sample size (e.g., how many other words are in the text).

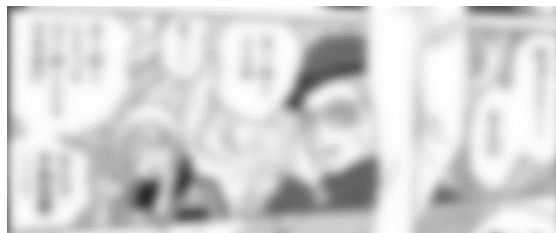
This requires a way to divide the entire corpus into countable units, so as to obtain a full, contextual sample. Japanese morpheme analyzers are one answer to the question. Here, the word *morpheme* refers to the smallest unit of language which has meaning. This might be an individual word, or it could be a *part* of a word. For example, the word *happy* has one morpheme—it cannot be further divided—whereas *happily* has two: It can be divided into its root, *happy*, and the bound morpheme *-ily*, used to form adverbs. Morpheme analyzers are software programs which use internal dictionaries and grammatical rules to probabilistically parse text into words, determining word boundaries based upon the surrounding environment of the text, including orthographic markings. Words are marked for word classes, usually including a central group of *nouns*, *verbs*, *adjectives*, *particles*, *adverbs*, *pronouns*, *proper nouns*, *interjections* and *symbols*. Several different such programs are available at the moment, with the major players for Japanese being Chasen, developed by Yuuji Matsumoto's Computational Linguistics Laboratory at the Graduate School of Information Science in the Nara Institute of Science and Technology; Juman, developed by Sadao Kurohashi and Daisuke Kawahara's Laboratory in the Department of Intelligence Science and Technology in the Graduate School of Informatics, Kyoto University; and MeCab, developed by Taku Kudou, a graduate of the Graduate School of Information Science in the Nara Institute of Science and Technology and currently a software engineer at Google. In terms of order, Juman was created first, which served as a partial base for Chasen; MeCab was later started with Chasen as its base, but was eventually re-done as an independent program (Kudou, 2012).

Chasen and MeCab, the two types used here, both are based upon a model of corpus learning, i.e., a small set of known words and rules is expanded upon with the introduction of additional data. While accuracy

rates for these different programs are not publicly available, using the UniDic dictionary data produced by the NINJAL (Nihongo Koopasu, 2004), both MeCab and Chasen showed a high-level of accuracy, with MeCab somewhat higher in terms of sentence endings (99.69% vs. 99.30%), parts of speech (98.89% vs. 97.92%) and word identification (98.59% vs. 97.22%). However, unusual orthographic styles appear to complicate issues. MeCab was less accurate for *Yahoo! Chiebukuro* (*Yahoo! Answers*), which uses more informal, conversational-like speech, and the *Corpus of Spontaneous Japanese*, a corpus of spoken Japanese compiled by NINJAL. MeCab's accuracy rates fell from  $F = .996$  for government white papers in regard to parts of speech to  $F = .984$  for both corpora, and from  $F = .994$  in regard to word identification to  $F = .981$  for *Yahoo! Chiebukuro* and  $F = .980$  for the *Corpus of Spontaneous Japanese*.

Given their unusual orthographic styles, manga can be predicted to have a comparatively low accuracy rate. This affects the data in two ways. First, the forms do not always match the versions registered in the program's internal dictionaries. For example, in *Nodame Cantabile*, the character Nodame often mixes katakana and hiragana in the forms she uses to end sentences where they would normally be written in hiragana alone; e.g., the copula “です” *desu* might be written as “でス”, or the formal sentence ending “ます” *masu* might be written as “まス” (Figure 54). Neither MeCab nor Chasen parse these properly, generally dividing such examples into two words (で - ス). Second, the lack of traditional orthographic symbols means that markers for word, sentence and clause boundaries are less clear. While this does not always affect word assignment *per se*, it is important when determining sentence boundaries and the forms of conjugation for predicates.

**Figure 54:** Example of non-standard writing of *desu-ka* (ですか→でスカ) in *Nodame Cantabile* (2:23)



As a result, manga may seem an unlikely target for morpheme analysis, and it is true that one is not without alternatives. One option would be n-gram analysis, which breaks the text into blocks of text repeatedly by numbers of units (characters, in the case of Japanese), and then probabilistically assesses which ones are likely functioning as words. N-gram analysis has the advantage of not requiring any dictionaries or other

annotated data in order to create its models; in fact, the text itself ends up being completely un-annotated, with no tags for parts of speech or conjugations at all, leading to a purely data-driven analysis. These are compelling points given the tenuous definition of words in Japanese. However, n-gram analysis is not an easy beast to handle. It leads to larger data output, as blocks of text are replicated over different combinations of units. For example, the simple 5-character phrase “学校に行く” (*gakkou ni iku*, (I) go to school) could potentially be broken into 14 different blocks (学, 校, に, 行, く, 学校, に行, 校に, 行く, 学校に, 校に行, 学校に行, 校に行く, 学校に行く). Data sets generated in this way can be enormous, even for small corpora. Furthermore, n-grams require more abstract statistical and probabilistic analyses in order to determine the word frequencies seen. This not only increases the burden that researchers feel, but also the burden of those who might be interested in the results. In addition, there are no widely available tools for doing Japanese n-gram analysis, requiring individuals to create their own.

While there are flaws with morpheme analyzer programs, they represent a comparatively accessible approach to dealing with the great task of text parsing. To overcome the issues that one might expect to see, I have taken an approach towards *standardizing* the text through the removal of non-essential orthographic markers which do not directly affect meaning, followed by thorough checks. Working with non-standard texts is an increasingly important part of morpheme analysis, as can be seen from some of the recent updates to Juman, which included adaptations for non-repetitive onomatopoeia, non-standard orthography using the bousen, and automatic recognition of long-vowels using the bousen and small kana (Kurohashi & Kawahara Laboratory, 2012). Conducting a morpheme analysis of manga could contribute to these larger issues by offering insight into what is necessary for dealing with non-standard texts, and what approaches are possible for researchers working with such data sets.

In terms of the data itself, the analysis here will focus on (1) the distribution of parts of speech, (2) the types of words seen in terms of their origin (*kango*, *wago*, *gairaigo*, etc.), (3) the number of sentences seen, and (4) the top-10 most frequent words from the most important word classes (nouns, verbs, adjectives, and adverbs). The first two points were selected for their ability to compare manga with other genres of text. In regard to (1), research has widely shown the distribution of the different parts of speech used in texts to be one aspect which varies between genres, such as with web diaries (blogs) and letters to newspapers (Kishimoto, 2008), as well as across languages (Nakano (1976) summarized in Hayashi et al. (1982)). This is also the case for (2), as the types of words used—e.g., whether they are of Japanese or Chinese origin—are also viable ways to differentiate genres. *Kango* words of Chinese origin, for example, tend to be more formal compared with *wago* words of Japanese origin (Gottlieb, 2008). Thus, *kango* are naturally associated with texts which are more formal, such as newspapers or academic writing. As was already noted in Section 3.3, the length and complexity of sentences is

another marker used for determining the readability of texts (3). Finally, the most frequent words seen (4) can, I will argue, tell us something about the nature of the texts and offer a point of comparison between genres and series.

Through this analysis, I will show that manga again show similarities with spoken language in regard to certain aspects of the distribution of parts of speech, such as their frequent use of interjections, but are less similar in terms of their use of vocabulary by origin, in particularly using fewer kango than spoken language does. This data seems to show that manga, while approaching spoken language in many ways, is indeed a *simulated* spoken language, and is thus hybrid in its textual styles. I will also point out several patterns in regards to the types of lexical items seen that appear to differ between shoujo-manga and shounen-manga. Particularly, shounen-manga appears to often use specialized vocabulary relevant to the themes they deal with, whereas shoujo-manga uses more everyday words. In this way, I will demonstrate that word frequency can offer interesting insight into how genre functions within manga itself as well.

### 3.4.2 Methodology

Data was limited to *Lines*, given (1) their central role in manga narratives and (2) that they are most likely to be in sentence-form. They were run through ChaMame, a morpheme analyzer interface. ChaMame was selected for its ability to use UniDic, a dictionary for morpheme analysis compiled by Yasuharu Den, Atsushi Yamamoto, Hideki Ogura, Hanae Koiso and Toshinobu Ogiso, and made available through NINJAL (<http://www.tokuteicorpus.jp/>). ChaMame was also created by the same group. Having been compiled as part of the Balanced Corpus of Contemporary Written Japanese, a project of unparalleled scope, the data is extremely robust, and is likely the most thorough of the currently available and easily accessible morpheme analysis dictionaries. ChaMame allows users to analyze the data using either Chasen or MeCab as support programs: In general, I used MeCab given its higher rate of accuracy, using Chasen only to compare misidentified data.

Results were exported into Excel, where they were relinked to the original data to include speakers' names, sexes and other information. UniDic automatically annotated each word for (1) sentence boundaries (*B* for sentence beginnings; *I* for words mid-sentence); (2) the main form seen in the text; (3) the main form's phonetic reading; (4) the phonetic-reading of the root; (5) the root form the word is associated with, which may be different for variant forms, verbs and adjectives; (6) the root's standard written form; (7) its part of speech; (8) for conjugating verbs and adjectives, its conjugating class; (9) the form of conjugation for verbs and

adjectives; (10) the katakana form of the word; and (11), the origin of the word (e.g., kango, wago, etc.). Example 4 shows how ChaMame treated one real line of text; each tag is in the order above.

**Example 4:** Text as analyzed by ChaMame

Source	Boundary	Main form	Phonetic	Root-phonetic	Root	Part of speech	Conjugating class	Form of conjugation	Form	Type
chamame	B	不幸	フコー	フコウ	不幸	名詞-普通名詞-形状詞可能			フコウ	Kango
chamame	I	を	オ	ヲ	を	助詞-格助詞			ヲ	Wago
chamame	I	与える	アタエル	アタエル	与える	動詞-一般	下一段-ア行	連体形-一般	アタエル	Wago
chamame	I	力	チカラ	チカラ	力	名詞-普通名詞-一般			チカラ	Wago
chamame	I	まで	マデ	マデ	まで	助詞-副助詞			マデ	Wago
chamame	I	は	ワ	ハ	は	助詞-係助詞			ハ	Wago
chamame	I	ない	ナイ	ナイ	無い	形容詞-非自立可能	形容詞	終止形-一般	ナイ	Wago
chamame	I	けど	ケド	ケレド	けれど	助詞-接続助詞			ケド	Wago
chamame	I	も	モ	モ	も	助詞-係助詞			モ	Wago
chamame	I	・・・			・・・	補助記号-一般				Symbol
chamame	I	!			!	補助記号-句点				Symbol

不幸/を/与える/力/まで/は/ない/けど/も/・・・/! Fukou/wo/ataeru/chikara/wa/nai/kedo/mo/・・・/! (I don't have the power to give people bad luck, though...)

Background Line/Thought from *Kimi ni Todoke* (1:15)

Using the results from the morpheme analysis, four different statistics were generated for analysis: (1) the distribution of parts of speech across the corpus; (2) the make-up of the corpus by lexical origins; (3) the number of sentences seen; and (4) the top-10 most frequent nouns, verbs, adjectives and adverbs. For (1) and (2), I tallied the data according to the part of speech and type. For (3), I calculated the number of sentences using the number of verbs, adjectives and auxiliaries in predicative form. Note that sentences here *include* strings such as quotations which are embedded but are in predicative form, but do not include dependent clauses using non-predicative forms. Predicative forms preceding other predicative forms—such as an adjective before the polite copula *desu*—were also removed to avoid double-counting. Finally, in regard to (4), I calculated the top-10 most frequent nouns, verbs, adjectives and adverbs for each individual series; when ties occurred, both forms were listed. Note that while Japanese has two major types of adjectival forms—*true* adjectives, which end in *-i* and conjugate, and *na*-nouns, which are actually nouns and do not conjugate independently—I chose to combine the two for the purposes of this more general overview.

Before the data could be statistically processed, however, ChaMame required thorough pre- and post-run checking in order to be used effectively, particularly through standardization. The text also required substantive manual checks, in addition to the setting of new parts of speech tags and word types particular to this data set. The following sections will describe these processes in detail prior to introducing the results.



**Pre-ChaMame standardization:** The unusual orthography that characterizes manga can make it more difficult to use morpheme analyzers as they (1) break up the data into unrelated strings (e.g., “いゝと思ふ” (*ii to omou*) vs. “いゝ・・・と思ふ” (*ii . . . to omou*): Both are one sentence, but the pause registered by the 3-point lead incorrectly leads Chamame to parse it as two separate expressions); (2) lead to incorrect data analysis in attempts to satisfy the role of unnecessary makers (“いゝのよ” (*ii no yo*-) analyzed as four parts: いゝ / の / よ / つ instead of recognizing “つ” as part of the final particle *yo*); and (3) cause unnecessarily large numbers of data lines, making it difficult to analyze the data through the noise generated. Using regular expressions, the text was standardized—i.e., made to conform to more standard orthographic styles—in regard to the following three points.

1. *The chou-bousen, namisen and chou-namisen:* Due to their similar functions (Section 3.2), *namisen* were replaced by the more standard *bousen*. *Chou-namisen* were replaced by the *chou-bousen* and then processed identically. Multiple strings of *bousen* and *chou-bousen* were reduced. *Chou-bousen* appearing at the beginning of sentences or phrases were replaced with the dash. *Chou-bousen* were deleted if they did not offer essential vowel elongation; those that did were replaced with the *bousen*. For example, the string “———かわい〜〜、ほしいな———” (*---kawai~~, hoshii na---*: “That’s cute, I want it”) would be reduced to “—かわい—、ほしいな”. The first *chou-bousen* is sentence-initial, the second non-essential elongation, and the *chou-namisen* essential elongation.

2. *Utterance-final small tsu:* Small *tsu* (っ) in sentence- or phrase-final environments were removed, as they could be easily determined to be non-geminate in nature. Such environments were defined as either being at the end of a *Line*, or proceeding orthographic symbols (e.g., ・ ・ ・ , 。 , ! , ? ).

3. *Multiple, repetitive symbol strings:* Multiple strings of 3-point leads, exclamation marks and enigmas were reduced (i.e., ・ ・ ・ ・ ・ ! ! ! reduces to ・ ・ ・ ! ). To maintain their status as compound symbols, 3-point leads made up of *nakaguro* were rewritten as a single character (...) for processing; strings of 2, 4, and 5 *nakaguro* were also replaced with characters not otherwise seen within in the text. Post-processing, they were returned to the appropriate number of *nakaguro*.

**Post-ChaMame alterations:** Manual checks were conducted in a three-part process using unique-word lists by root and parts of speech, and (1) checking for words which seemed unlikely (e.g., the *bousen* ー alone); (2)

assuring consistent analysis of parts of speech assignments by checking roots with multiple assignments; and (3), adjusting for between-series consistency of assignments. Several specific adjustments were made as follows.

1. *Notes column*: A column for *Notes* was created to add additional information about non-predictable furigana, dialectal forms, compound words, and phonetic changes, such as using [ne:] for /nai/ described in Section 3.2.

2. *New categories of word classes*: Several new word classes were defined given the special needs of the text.

Interjections—Misspoken: As simulated spoken speech, it is not uncommon to find what appears to be broken speech, words started and stopped, words only partially audible, or other examples of text whose meaning is either lost because the speakers stopped mid-way or were unable to complete the sentence. These were filed under the new category of Interjections—Misspoken, and characterized as *other* for word-type.

Foreign languages (English, German, Italian, Chinese): While ChaMame allows for foreign gairaigo loan words, it is not accommodating to interwoven, non-Japanese text. Examples of foreign language sentences were labeled as being from their original language, regardless of the scripts used.

Proper nouns—People names—Nicknames: To maintain the relationship between nicknames and who they are referring to, nicknames were given a special category, with the root form and the phonetic root form replaced by the full name.

3. *Recombination of compound morphemes*: Because UniDic seeks to break morphemes into the smallest unit possible, many compound words are separated into several lines of text. In addition to sound changes which can only be explained by treating them as compounds (e.g., realizing *koto-wa* ことは as *kottaa* こったあ: The noun, *koto*, and the particle, *wa*, have undergone phonetic changes and fused together), many pronouns, conjunctions and particles are often seen together, such as *nani-mo*, *sore-ni*, and *to-ka*. As leaving common compounds diminishes the informativeness of the data, such words were recombined into one, noting that they were compounds in the notes column.

4. *Dealing with non-predictable furigana*: Non-standard furigana present a unique issue for morpheme analysis: Should one favor their *phonetic* forms, or should one favor their *orthographic* forms? In general, I chose to

assume that the furigana were the *spoken* form, and replaced all but the main form with the furigana reading, as is done with proper nouns. Each example was tagged as such in the notes column.

5. *Treating symbols*: Combinations of enigmas and exclamation marks were rewritten as one combined symbol. This was done to reduce the overall number of text lines, and to accurately reflect the relationship between these two symbols. The enigma and the exclamation mark are usually written side-by-side, even in vertical-text, suggesting that they also functioned as a compound symbol; combining them onto one line allows one to show this relationship.

### 3.4.3 Results

**Overall:** 260,906 words were obtained from the morpheme analysis, 108,758 from shoujo-manga, and 152,148 from shounen-manga, a result of the difference in the number of *Lines* to begin with (219,513 characters and 280,455 characters, respectively) (Table 19). To give an idea of how much data the standardization process removed, out of the 499,968 characters originally found in *Lines*, 41,273, or 8.26%, were removed. The percentage of removed characters ranged from between 3.96% (*Detective Conan*) to 18.10% (*KimiTodo*).

Table 19 also divides the data into substantives, non-substantives, and symbols. Most research on the lexical make-up of texts has differentiated between substantives and non-substantives (i.e., freestanding words such as nouns and verbs compared with particles or suffixes, which must be used with other words), using only substantives due to the predictive nature of words such as particles: A noun will always potentially call upon a particle, so the number of particles largely correlates to the number of nouns. I also limited the data set to substantives, and removed particles, prefixes and auxiliary verbs following the precedents set in Kokuritsu Kokugo Kenkyuujo Houkoku 8 (1955), reported in Hayashi et al. (1982). Symbols were also removed. Suffixes and their roots were combined into compounds, following the class of the suffix (e.g., adverbial suffixes make an adverb). Non-substantives came to 39.50% of the data; this high percentage reflects the fact that they are often largely obligatory, and will come with most nouns as particles, and verbs as auxiliary endings (past tense conjugations, negations, etc.).

**Table 19:** Number of words by type

Genre	Titles	All		Substantives		Non-substantives		Symbols		Removed characters	
		Characters		Characters		Characters		Characters			
Shoujo-manga	Bokura	17,038	30,971	8,509	18,380	6,478	8,763	2,051	3,828	2,567	7.65%
	Kimi	20,379	37,187	9,107	20,022	7,611	10,265	3,661	6,900	8,221	18.10%
	Nana	27,884	49,692	13,340	29,951	11,782	15,603	2,762	4,138	2,967	5.63%
	Nodame	19,589	36,208	9,209	21,666	7,568	10,276	2,812	4,266	3,923	9.78%
	Rabukon	23,868	43,850	11,462	26,023	8,839	12,173	3,567	5,654	3,927	8.22%
	Total	108,758	197,908	51,627	116,042	42,278	57,080	14,853	24,786	21,605	9.84%
Shounen-manga	DeathNote	35,745	60,413	17,525	36,028	15,298	18,966	2,922	5,419	4,166	6.45%
	GinTama	28,870	50,282	13,903	30,353	12,302	16,134	2,665	3,795	2,320	4.41%
	Konan	36,309	61,518	15,384	32,046	13,790	17,038	7,135	12,434	2,537	3.96%
	Naruto	20,454	35,864	9,250	19,788	7,621	9,791	3,583	6,285	3,599	9.12%
	OnePiece	30,770	52,710	13,779	29,937	11,756	14,572	5,235	8,201	7,046	11.79%
	Total	152,148	260,787	69,841	148,152	60,767	76,501	21,540	36,134	19,668	7.01%
Total		260,906	458,695	121,468	264,194	103,045	133,581	36,393	60,920	41,273	8.26%

**Table 18:** Parts of speech for various mediums

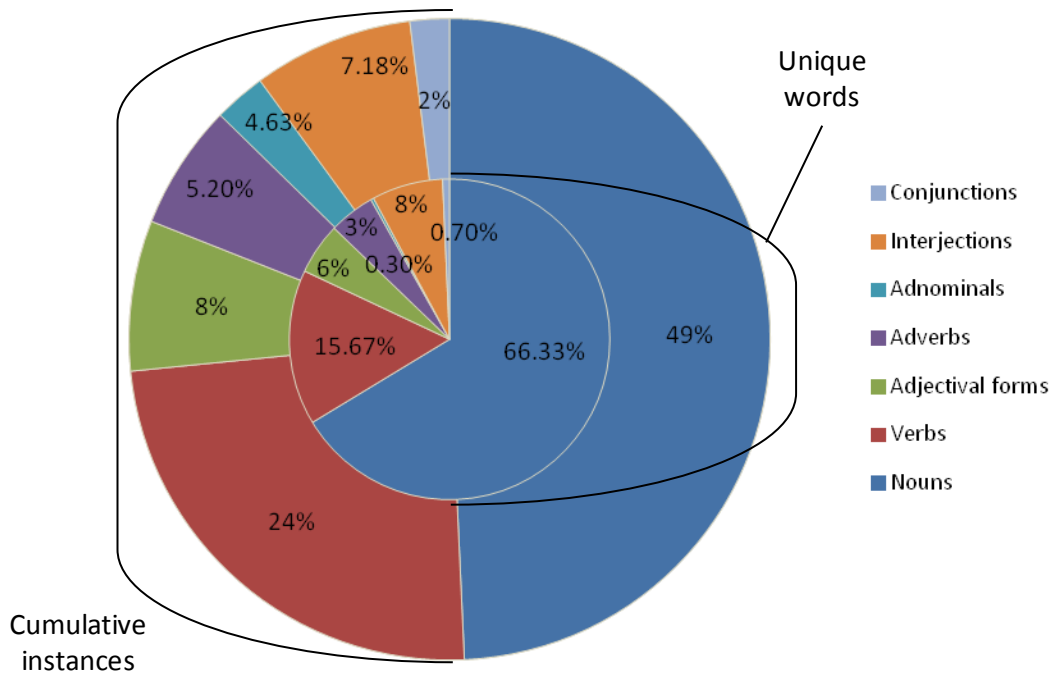
Genre	Details	Nouns	Verbs	Adjectival forms	Adverbs	Adnominals	Conjunctions	Interjections
Manga	<i>Shoujo</i>	47.01	22.99	<b>9.14</b>	7.15	2.08	2.32	9.30
	<i>Shounen</i>	50.93	25.00	6.43	5.71	<b>3.01</b>	1.73	7.19
	Total	49.27	24.15	7.58	6.32	2.61	1.98	8.08
Spoken language (1)		40.90	24.40	7.80	12.20	1.60	<b>3.80</b>	<b>9.40</b>
Blogs (1)	By women	41.90	<b>31.80</b>	7.90	<b>13.10</b>	1.70	2.20	1.40
	By men	46.50	31.10	8.30	7.70	2.00	2.40	2.00
	All	44.20	31.45	8.10	10.40	1.85	2.30	1.70
The Little Prince (2)		54.50	26.60	←18.90→				
Newspaper letters (3)		<b>55.50</b>	31.60	5.60	3.90	2.10	1.20	0.10
Magazines (4)		63.65	22.47	←12.3→			←1.76→	
Dictionary (Nihon Kokugo Daijiten) (5)		89.57	6.37	2.32	1.39	0.05	0.06	0.24

**Bold=Genre with the most in that type. Italics =Genre with the least in that type. (Excluding dictionary entries.)**

(1)~(5) taken from (1): Kokuritsu Kokugo Kenkyuujou Houkoku 8 (1955), *Danwago no jittai* (Tokyo: Shuei), in Hayashi, Miyajima, Nomura, Egawa, Nakano, Sanada, and Satake (eds.) (1982); (2): Kishimoto (2008); (3): Nakano, Hiroshi (1976), "Hoshi no oojima-sama" 6kakokugoban no goironteki kenkyuu (*Keiryuu Kokugogaku*, 79, 18-31), in Hayashi et al (1982); (4): Kokuritsu Kokugo Kenkyuujou Houkoku 25 (1964), *Gendai Zasshi 90shu no Yougo Youji Daisanbunsatsu Bunseki* (Tokyo: Shuei), in Hayashi et al (1982); (5): Matsui, Eichi (1979), "Nihon Kokugo Daijiten" shuuroku koumoku bunpu-hyou (*Kokugotenbou*, special edition), in Hayashi et al (1982)

**Parts of speech:** Table 18 shows the percentages of the major parts of speech seen in manga compared with other major genres. The most striking difference between manga and other genres was its relatively high-percentage of interjections (8.08%). It has this quality in common with spoken language (9.4%), which actually used interjections somewhat more frequently. Nonetheless, both of these two genres had 6 points over any other genre. Manga were also similar to spoken language in their relatively low usage of verbs, both placing somewhat over 24%, with only magazines lower at 22.47%; the Japanese translation of the *Little Prince*, a widely popular classic children's novelette with drawings, also placed near at 26.60%. However, it did not have very much else in common with manga, although it is similarly for a young audience and is a mixed-visual medium.

**Graph 4:** Distribution of parts of speech - unique words vs. number of occurrences



Shoujo-manga and shounen-manga also showed an interesting similarity with blogs written by female and male authors. Kishimoto's (2008) study showed that blogs by women tended to have a lower percentage of nouns and more adverbs, with small differences between the other parts of speech. Likewise, shoujo-manga also feature *fewer* nouns than shounen-manga (47.01% vs. 50.93%), and *more* adverbs (7.15% vs. 5.71%). While other differences—the higher number of adjectival forms, for example—are not explicable by this model, it may be that a similar phenomenon is at work: The shoujo-manga series were all written by women, with the shounen-manga largely written by men (the possible exception being *Death Note*, whose writer has remained anonymous). This is question of authorial influence will come up again in Chapter 4 on language and characterization.

Finally, while the data in Table 18 show the number of overall occurrences, Graph 4 compares the number of unique new words to the number of overall occurrences of each part of speech. These are interesting to compare as discrepancies between these two percentages give insight into whether or not there are many words with low frequencies, and how the data actually distributes amongst individuals: A high number of unique words compared with a fewer occurrences suggests that many of those words must not be occurring often, whereas a low number of unique words and more occurrences suggests that the same words are being used frequently. Interjections, conjunctions, adnominals, adverbs and adjectival forms all had similar percentages for

both calculations, but nouns and verbs were interesting in regard to this point. Whereas the number of unique nouns came to 66.33% of the data, they were only 49% of all occurrences; in contrast, verbs made up 15.67% of all unique words, and 24% of all occurrences. This suggests that the group of nouns being selected from is relatively wide, but that many are used infrequently. It also suggests that the number of verbs being used is relatively few but that at least certain members are used very often. These two points will be picked up again below looking at the most frequent words by parts of speech.

***Types of words seen:*** Table 20 shows the types of words seen in percentages, as compared with other genre. Manga show the lowest usage of kango (16.95%), along with one of the highest percentages of wago (70.55%). They seemed to place closest to children's books, which had the second lowest percentage of kango (18.70%) and the highest percentage of wago (78.00%). While there did not appear to be much in common between manga and children's books in regard to parts of speech, it does seem to imply that there are some similarities. They appeared to be furthest from letters to newspapers, a comparatively controlled and literary form of public writing, which had the highest percentage of kango (46.40%) and the lowest percentage of wago (48.20%). While manga showed some similarities with spoken language in regard to the parts of speech used, spoken speech actually appeared to use kango much more frequently (23.60%), although the percentage of wago was very similar (71.80%). As can be predicted from the orthography studies, shounen-manga tended to use somewhat more kango (18.69% vs. 14.61%), with shoujo-manga slightly higher in all of the other four categories. Manga also had the highest percentage of mixed words (i.e., words using two or more morphemes of different origins) than other genres, but this is likely related to how suffixes were dealt with. It was not made clear within the summaries of the other genres which were used as resources how suffixes were treated, and it is possible that they were treated as non-substantives.

**Table 20:** Types of word seen by origin for various mediums

Genre	Details	Kango	Wago	Gairago	Mixed	Other
<i>Manga</i>	<i>Shoujo</i>	14.61	71.28	3.65	<b>4.14</b>	<b>6.33</b>
	<i>Shounen</i>	18.69	70.01	2.26	3.29	5.75
	<i>Total</i>	16.95	70.55	2.85	3.65	5.99
<i>Children's books (1)</i>		18.70	<b>78.00</b>	2.10	1.20	-
<i>Blogs (2)</i>	<i>By women</i>	19.80	70.20	5.30	1.00	3.80
	<i>By men</i>	26.50	62.40	<b>7.30</b>	0.70	3.10
	<i>All</i>	23.15	66.30	6.30	0.85	3.45
<i>Spoken language (3)</i>		23.60	71.80	3.20	1.40	-
<i>Magazines (4)</i>		41.30	53.90	2.90	1.90	-
<i>Newspaper letters (2)</i>		<b>46.40</b>	<i>48.20</i>	2.80	<i>0.30</i>	<i>2.30</i>
<i>Dictionary (Kadokawa Kokugo Jiten) (5)</i>		37.10	52.90	7.80	2.20	-
<b>Bold</b> =Genre with the most in that type. <i>Italics</i> =Genre with the least in that type. (Excluding dictionary entries.)						
<i>Other</i> =Foreign languages, misspoken interjections, proper nouns						
(1)~(5) taken from (1): Nomura & Yanagise (1979) , <i>Jidou yomimono no goi kousou (Keiryuu Kokugogaku , 12 (2))</i> , in Hayashi, Miyajima, Nomura, Egawa, Nakano, Sanada, and Satake (eds.) (1982); (2): Kishimoto (2008); (3): Nomoto et al (1980), <i>Nihonjin no chishiki-kaisou ni okeru hanasikotoba no jittai (Monushou Kagaku Kenkyuuhi Tokutei Kenkyuu "Gengo" Kenkyuu Houkokusho )</i> , in in Hayashi et al (1982); (4): Kokuritsu Kokugo Kenkyuujo Houkoku 25 (1964), <i>Gendai Zasshi 90shu no Yougo Youji Daisanbunsatsu Bunseki</i> (Tokyo: Shuuei), in Hayashi et al (1982); (5): Sanetoukeishuu (1973), <i>Kindai Nicchuu Koushoushiwa</i> (Tokyo: Shunshuusha), in Hayashi et al (1982)						

**Number of sentences:** In total, 34.59% of all verbs, adjectives and auxiliaries were in the predicative form (Table 21). While one might be tempted to treat one speech bubble as one sentence, their being one block of text, the number of predicative forms per entry only averaged 0.67, meaning that approximately 33% of the time, any given entry will *not* include a predicative form. This means that a large percentage of entries are not sentences, but rather fragments or non-complete replies. This makes for an interesting comparison with Nishimura's (2011) results for keitai-shousetsu and conventional novels. Within keitai-shousetsu, 24.4% of quotations—similar to *Lines*—featured complete endings, compared to 36.0% in conventional novels. However, for narration—similar to *Narration*—these percentages went up to 67.1% and 87.0%, respectively. As I argued in Section 3.3, it seems appropriate to compare *Lines* to Nishimura's quotations, which would place manga somewhere between keitai-shousetsu and conventional novels in regard to fragments. However, in both keitai-shousetsu and conventional novels, narration forms the bulk of text (62.8% and 60.4%, respectively). Thus, while manga compare more closely to quotations within conventional novels in terms of the number of sentences vs. fragments, one will actually see fragments much more commonly in manga because *Lines* are so much more common. Allen & Ingulsrud (2007, p. 8) suggest that there is a trend to take the necessary information in manga out of the captions of *Narration* and into put it *Lines* in order to force readers to read them; if this were the case, then this might

mean that *Lines* in manga are doing double-duty, in that they may be at times acting more similarly to Nishimura's narration than quotations. This may help account for differences with other mediums.

**Table 21:** Number of sentence-forming predicates seen by genre and series

Genre	Series	Sentence-forming		Non-sentencing forming		Sen./Entry	Char./Sen.
Shoujo-manga	<i>Bokura ga ita</i>	1,630	35.81%	2,922	64.19%	0.47	20.58
	<i>Kimi ni todoke</i>	1,702	33.90%	3,319	66.10%	0.60	26.68
	<i>Nana</i>	2,671	35.02%	4,956	64.98%	0.81	19.72
	<i>Nodame Cantabile</i>	1,859	38.39%	2,984	61.61%	0.57	21.59
	<i>Rabukon</i>	2,386	37.98%	3,897	62.02%	0.70	20.02
	<i>Total</i>	10,248	36.18%	18,078	63.82%	0.63	21.42
Shounen-manga	<i>DeathNote</i>	2,926	30.52%	6,662	69.48%	0.95	22.07
	<i>Detective Conan</i>	2,854	34.11%	5,513	65.89%	0.82	18.43
	<i>GinTama</i>	2,342	30.16%	5,422	69.84%	0.56	27.35
	<i>Naruto</i>	1,606	34.99%	2,984	65.01%	0.53	24.57
	<i>One Piece</i>	2,986	38.55%	4,759	61.45%	0.69	20.01
	<i>Total</i>	12,714	33.41%	25,340	66.59%	0.70	22.06
Grand Total		22,962	34.59%	43,418	65.41%	0.67	21.77

Sen./Entry=Number of sentence-forming predicates over the number of entries seen; Char./Sen.=Number of characters seen over the number of sentence-forming predicates

Since the number of predicates and sentence lengths were used as one marker of readability, one might also anticipate genre differences here, there were no significant differences between shoujo-manga and shounen-manga in regard to the percentage of verbal/adjectival forms which were sentence-forming ( $t(8) = 1.44, p = .186$ ), nor for number of predicative forms per entry ( $t(8) = 0.81, p = .4366$ ) or number of characters per sentence-forming predicates ( $t(8) = 0.38, p = .7158$ ). This suggests that the major difference in readability was that of kanji usage.

**Most frequent nouns:** Table 22 lists the top-10 most frequent nouns for each series. The top-10 nouns in shoujo-manga take up a higher percentage of all nouns seen compared with shounen-manga (22.74% vs. 17.94%), which was found to be statistically significant ( $t(8) = 2.864, p = .02103^*$ ). This may suggest that shounen-manga cover a larger diversity of topics, and one may notice some interesting differences between the two genres just by looking at the top-10 nouns. While many of the top-10 nouns in shoujo-manga were personal pronouns (*atashi*, *omae*; 3~4 per each series, average 3.2), proper nouns (e.g., the names of characters; 2~4 for each series, average 3.4), and nouns with a grammatical function (*koto*), all of the shounen-manga series with the exception of



*GinTama* had at least one noun related to their themes: *sousa* (investigation, *Death Note*), *kaizoku* (pirate; *One Piece*), *tantei* (detective, *Detective Conan*), and *ninja* (ninja, *Naruto*) for example. In comparison, only one shoujo-manga series featured a word related to its theme in the top-10 list: *Nodame Cantabile*, with *okesutora* (orchestra) and *shiki* (lead/conduct), in line with *Nodame*'s musical theme. As I discussed in Section 3.2 on unusual kanji, *Nodame* appears to be somewhat exceptional compared with the other shoujo-manga series in the corpus in regard to its strong orientation towards a particular theme, in this case, music and the development of the main characters into professional musicians. This is not dissimilar to the tendency in shounen-manga to be focused on a theme and the main character's maturation in that field such as Luffy and his dream of becoming a pirate in *One Piece*, or Naruto and his maturing into a strong ninja on the road to be the head ninja Hokage in *Naruto*. *GinTama*, the only shounen-manga with no theme-nouns in the top-10 list, is also exceptional in its lack of such a theme. This may seem surprising at first given that it superficially is similar to the other shounen-manga: It also has an unusual setting (an alternate world Edo—with aliens), but this theme is actually rather less important to the series. As a comedy, it is less centered on character development within a certain system, and this may have affected how often those kinds of thematic words appear.

One may also observe another important difference between the top-10 nouns seen in shoujo-manga and shounen-manga, which I would describe as the importance of interpersonal relationships in shoujo-manga compared with shounen-manga. As noted above, many of the top-10 nouns in shoujo-manga were proper names and personal pronouns; combined, these two types came to 15.43% of the top-10 nouns in shoujo-manga compared to 8.20% in shounen-manga. Individually, the difference between the percentage taken up by the proper names was found to be significant ( $t(8) = 4.25, p = .0028^*$ ), but this was not the case for personal pronouns ( $t(8) = 0.81, p = 0.4423$ ). The significantly high percentage of proper names might suggest that the characters are very concerned about each other: While there are no words related to themes, the usage of names suggests that one of the topics of conversation is likely *themselves*. Because it is not uncommon to avoid using second person pronouns in Japanese in favor of other forms of reference, such as titles and proper names (Whitman, 1999), it is possible that many of these usages are actually being used as second-person pronouns. While the usage of personal pronouns was not significantly different between the two genres, if one considers that some of the names are likely being used in this way, the actual amount of interaction between characters may actually be higher in shoujo-manga. This point might also speak of more continuous and direct, personal interaction between characters in shoujo-manga.

**Most frequent verbs:** The most frequent verbs were generally everyday basic verbs such as *suru* (to do—number one for all but one series), *iu* (to say), and *naru* (to become) (Table 23). The same verbs were held in common

throughout all of the series, with only 16 different forms seen; this is consistent with the earlier data (Graph 4) suggesting that a small number of verbs would make up a large number of the examples seen. A large percentage of the verbs were potential auxiliaries: 10 out of the 16 different most frequent verbs were verbs that could possibly be non-independent. These include the *iru*, to be, which is used with the conjunctive form of the verb to form the durative tense (e.g., *tabe-teiru*, eat+-*iru*, '[I'm] eating'); as well as *iku*, to go, used with the conjunctive form to express change over time (e.g., *kawatte-iku*, change+-*iku*, '[things] will change'). As was mentioned above, most of the verbs were not in the predicative form, which came to just 18.33% of all verbs. Combined with the high frequency of potential auxiliaries, this suggests that there are a large number of fairly long sentences, with complex verb strings and dependent clauses.

Compared with nouns and adjectives, there are fewer clear theme-oriented differences between the genres or individual series. With the low number of verbs seen overall, the majority of the top ten verbs are shared in common with all the series. Within this list, however, there is one interesting point to mention: *Death Note*, *Detective Conan* and *Naruto* all feature the word *korosu*—to kill—within the top-10 verbs. While no other obviously theme-related verbs were to be found, the dramatic nature of the word alone stands out. Both *Death Note* and *Detective Conan* have murder, in one way or another, as themes; to a lesser extent, one might say the same for *Naruto*, with its fight-after-fight formula. Expanding the list may show more differences in themes of this nature. While not directly related to themes, *per se*, another interesting distinction is also the high use of *chigau* (to be different) in *RabuKon*. This is likely because the characters in *RabuKon* are in the Osaka area, and all actively use *kansai-ben*, or the dialect spoken in the Kansai area which includes Osaka, Kobe and Kyoto. In the Kansai dialect, the word *chigau* is often weakened to the form *chau*, and used to mean something along the lines of 'that's not it'. Thus, *chigau* has a special role in the Osaka dialect, and we can easily predict that many of these examples are actually *chau*—which they are (84 examples; 81.55%). This shows how the importance of the Osaka setting and the usage of the Osaka dialect in *RabuKon* come through linguistically.

**Most frequent adjectival forms:** Before looking at the frequency list in Table 24, it is important to note that a large number of those showing up amongst the top-10 adjectives were adjectival forms which have unique grammatical roles in Japanese. The most obvious example would be *nai*, which was the most common adjective in eight of the series—and the second most common form in the remaining two. *Nai* is used to express the negative form: To conjugate adjectives and to express the negative of the verb *aru*, one must use *nai* (e.g., *yoku-nai*, 'not good': the adjective *yoi* in its conjunctive form *yoku*, followed by *nai*; or *ikanai*, 'will not go': the verb *iku* in its *mizenkei* irrealis form *ika*, followed by *nai*). Likewise, the adjective *yoi* (good), which is often used in

idioms and compound adjectives (*kakko-ii*, cool) was the most common adjectival form in the remaining two series, and the second most for all other series. These two alone accounted for 35.85% of all adjectival forms.

Other common adjectives with grammatical forms were *you* (as/like; 8 series); *mitai* (as/like; 9 series); and *sou* (like; 6 series). While they do not tell very much about the themes of the different series, they might be informative about the *grammatical complexity* being seen. *You*, *mitai* and *sou* all attach to other verbs, nouns or adjectives to report information: It was *like* this, it *appeared* that, etc. This suggests that sentence structures being seen are not very simple—e.g., the basic subject-object-verb form—but rather have more embedded structures, a point which would have influence on their readability.

At least some of the remaining adjectival forms may be suggestive of the series' themes. For example, *RabuKon* features *chiisai* (small; 24), which is relevant to the main story line of a very tall girl, Risa, who falls for the shortest boy in her class, Ootani. While placing just out of the top-10, *dekai* (enormous; 22) was also common. In *Naruto* and *One Piece*, which both have fight scenes as one of their central aspects, the word *tsuyoi* (strong; 13, 24 respectively) is also comparatively common. With the exception of *KimiTodo*, all of the shoujo-manga featured *suki* (like/love) in the top-1- list, as opposed to shounen-manga, where it only placed in the top-10 in two series (*GinTama*, *One Piece*). Shoujo-manga are often said to be concerned with romance (Fujimoto (2008), Shamoan (2004), etc.), and the usage of *suki* could be a continuation of the theme of interpersonal relationships in shoujo-manga brought up by the most frequently seen nouns.

**Most frequent adverbs:** A total of 27 different words formed the group of most frequent adverbs, suggesting a fairly high repeat rate between series: Each must have ranked in the top-10 for an average of 2.7 series (Table 25). The most common form was, for all but two series, *sou*, meaning *that (way)*. In *Nodame*, it was the third most common, following *nan-de* (why) and *mou* (already). It was also the third most common adverb in *GinTama*, following *mou* and *dou* (how). *Sou* has a wide variety of usages, but is often used along the lines of *yes* in meaning (e.g., *Kureta no?—Sou na-no*: 'Did [they] give [you] [that]?' 'That's right (=yes)'), which might partially account for its high frequency.

One interesting point in regard to genre is the use of *nan-ka*, which ranked high in 3 of the shoujo-manga series (*KimiTodo*, *Nana* and *RabuKon*), but only one of the shounen-manga series (*GinTama*). *Nan-ka* is a hedge or filler that has been given attention recently for its increasingly wide usage amongst young people, and has been said to be similar in usage to the characteristics of young girls' speech patterns (Iio, 2006). It is a form mostly seen in spoken speech, which may be another hint at the subtle differences between shoujo-manga and shounen-manga. As with nouns, there were also some examples of words which might be related to series' themes, too. *Death Note* frequently used four adverbs that did not rank in any other series: *tsumari* (that is to

say), *naze* (why), *moshi* (if), and *mazu* (first). One might associate these words with logical thinking: *tsumari* in its role in restating sentences, *naze* in its ability to not just ask for reasons, but to state them (*naze ka to iu to* “to say the reason why”), *moshi* in setting up relations between events, and *mazu* in creating an order for actions. This may be too much speculation, but one of the narrative characteristics of *Death Note* might be said to be how the characters—in particular, Light and L—examine and analyze their situations, making predictions about each other’s actions as they try to determine their mutual identities. Such analytical discussions may have resulted in its particular adverb pool. While not as stark, *Detective Conan* also features two adverbs not seen in any others: *osoraku* (probably) and *doushite* (why). Like *Death Note*, *Detective Conan* is fairly analytical, with its focus on solving mysteries. *Osoraku* is a natural part of the speculative nature of the story: looking at the clues to draw out predictions of what had happened. *Doushite*, too, is related, in its usage both as demanding that ever needed motive (*why* would they do that?), and, too, from those who listen to the speculation and ask to know *why*.

Aside from such speculation, there is also a final and compelling question to ask about adverbs concerning the high-frequency of forms that straddle the line between adverbs and interjections. While I have chosen to not look at interjections at this time due to their great variability and almost infinite number of possibilities (Norrick, 2009), many of the most frequent adverbs seen in the corpus can, in certain circumstances, be interpreted not as adverbs, but rather as interjections. For example, *mou* can be used both as an adverb to mean ‘already’ (*mou tabe-ta* –already eat-past: “[I] have eaten already”), but it can also be used as an interjection showing one’s displeasure or that one wants someone to stop something: *Mou, anata-ttara!* (already/you+conditional) could be translated as *Oh you!*, with similar nuances. Attempts were made in the manual corpus checks to differentiate between these forms; in particular, adverbs and interjections that are phonetically identical were encoded using different characters so that they would not be compiled into one. For example, *mou* the adverb is listed as “も う”, whereas *mou* the interjection is listed as “もー”.

However, the distinctions are not always clear. While *mou ya-da* (already/dislike/copula) seems like an interjection—*Uck I don’t like it!*—but it could also be interpreted as an adverb (*I already don’t like it*). Which one is accurate depends on the situation. Adverbs are indeed a less stable class than might be assumed, and they often double as discourse markers as well (Fraser, 1996), which could also be potentially filed under interjections. While the distinction between adverbs and interjections is problematic for the corpus, they might be another way that manga are being encoded as spoken language: If one considers these ambiguous forms to possibly be a part of interjections, the rate of interjections would actually be higher.

Table 22: Most frequently seen nouns

Table 23: Most frequently seen verbs

Genre	Top 10 Nouns																				Total %			
Series																					%			
Shounen-manga	Boku ga Ito	ヤノ 160 (name)	Yano (name)	あたし 106	atashi I	何 105	nani what	俺 96	ore I	事 90	koto thing	御前 71	omae you	タカハシ 66	Takahashi (name)	人 61	hito people	ヤクモト 53	Yamamoto (name)	ナナミ 47	Nana (name)	855 22.28%		
	Kimi ni Todoke	サダコ 192 (name)	Sadako (name)	カゼハヤ 135	kazehaya (name)	何 111	nani what	事 93	koto thing	あたし 91	atashi I	私 87	watashi I	俺 83	ore I	ヤノ 78	Yano (name)	クロサマ 77	Kurosama (name)	あんな 76	anna you	1,023 25.23%		
	Nana	ナナ 243 (name)	Nana (name)	あたし 220	atashi I	シヨウジ 145	Shouji (name)	何 143	nani (name)	事 143	koto thing	俺 137	ore first	先輩 125	senpai (name)	御前 125	omae I	あんな 105	anna you	93	you	88 that 73	Jonko (name)	21.52%
	Nodame Contable	チヅキ 168 (name)	Chiki (name)	何 119	nani what	ノタメ 113	Nodame (name)	俺 111	ore I	先輩 108	senpai (name)	私 107	watashi I	オオサトウ 85	okusatoro orchestra	74	thing	御前 61	omae you	58	conduct	20.83%		
	Robukon	何 222	nani what	オオサトウ 207	Ootani (name)	俺 137	ore I	あたし 135	atashi I	事 134	koto thing	リサ 119	Lisa (name)	御前 88	omae police	其れ 81	sore that	コイヌミ 77	Koizumi (name)	男 65	otoko (name)	1,265 24.42%		
	DeathNote	キヲ 352 (name)	Kira (name)	事 303	koto thing	エル 194	L (name)	私 158	watashi I	捜査 130	sousa search	僕 120	boku I	警察 119	keisatsu police	人間 117	ningen human	其れ 111	sore that	者 98	mono person	1,702 18.19%		
	GinTama	何 200	nani what	俺 181	ore I	御前 165	omae you	事 142	koto thing	物 105	mono thing	てまえ 98	tenae you	奴 83	nu him	此れ 77	this	馬鹿 66	baka idiot	私 65	watashi I	1,182 16.78%		
	Detective Conan	事 200	koto what	私 174	watashi (thing)	犯人 157	hanjin I	貴方 111	anata culprit	其れ 109	sore you	何 99	nani that	人 85	hito what	此れ 81	this	先生 79	sensei I	探偵 79	toneri detective	1,054 13.81%		
	Naruto	御前 176	omae you	俺 156	ore I	何 120	nani what	事 113	koto thing	ナルト 110	Naruto (name)	其れ 74	sore that	忍者 67	ninja person	術 80	jutsu this	先生 54	sensei technique	サスケ 51	Sasuke (name)	985 20.67%		
	OnePiece	俺 361	ore I	何 216	nani( I	海賊 191	kaizoku pirate	事 153	koto thing	御前 148	omae you	馬鹿 83	baka idiot	ゾロ 81	Zoro (name)	私 80	watashi I	てまえ 76	tenae you	彼奴 71	afsu him	21.55%		
Shounen-manga																								

Genre	Top 10 Verbs																				Total %			
Series																					%			
Shounen-manga	Boku ga	為る	suru	言う	iu	来る	kuru	行く	iku	遅る	yaru	思う	omou	見る	miru	有る	aru	成る	naru	居る	iru	808		
	116	to do (A)	116	to say	74	to come (A)	66	to go (A)	65	to do (A)	60	to think	57	to look (A)	57	to have (A)	52	to become (A)	47	to be (A)	42.15%			
	Kimi ni	為る	suru	言う	iu	遅る	yaru	来る	kuru	有る	aru	思う	omou	分かる	wakaru	成る	naru	居る	iru	聞く	kiku	927		
	Todoke	215	to do (A)	155	to say	88	to do (A)	77	to come (A)	73	to have (A)	69	to think	69	to know	68	to become (A)	58	to be (A)	53	to hear	45.22%		
	Nano	為る	suru	言う	iu	行く	iku	成る	naru	来る	kuru	思う	omou	遅る	yaru	有る	aru	居る	iru	見る	miru	1,372		
	432	to do (A)	175	to say	135	to go (A)	113	to become (A)	112	to come (A)	93	to think	87	to do (A)	87	to have (A)	72	to be (A)	66	to look (A)	42.11%			
	Nodame	為る	suru	言う	iu	有る	aru	遅る	yaru	行く	iku	見る	miru	成る	naru	居る	iru	分かる	wakaru	869	to know	44.18%		
	Contable	282	to do (A)	100	to say	88	to have (A)	83	to do (A)	65	to go (A)	62	to come (A)	58	to look (A)	55	to become (A)	42	to be (A)	34	to know	44.18%		
	Robukon	為る	suru	言う	iu	成る	naru	遅る	chigau	行く	iku	遅る	yaru	来る	kuru	呉れる	kureru	有る	aru	思う	omou	1,231		
	338	to do (A)	227	to say	103	to become (A)	103	to differ	98	to do (A)	89	to do (A)	86	to come (A)	66	to give (A)	61	to have (A)	60	to think	45.97%			
Shounen-manga	DeathNote	為る	suru	居る	iru	言う	iu	有る	aru	成る	naru	見る	miru	殺す	korosu	分かる	wakaru	思う	omou	出来る	dekiru	2,178		
	634	to do (A)	410	to be (A)	280	to say	176	to have (A)	157	to become (A)	121	to look (A)	115	to kill	105	to know	92	to think	88	to be able (A)	45.58%			
	GinTama	為る	suru	言う	iu	来る	kuru	遅る	yaru	居る	iru	行く	iku	有る	aru	見る	miru	呉れる	kureru	1,991	to give (A)	54.70%		
	397	to do (A)	768	to say	151	to come (A)	126	to do (A)	125	to be (A)	122	to become (A)	79	to go (A)	75	to have (A)	75	to look (A)	70	to go (A)	63	to know	43.18%	
	Detective	為る	suru	居る	iru	言う	iu	来る	kuru	有る	aru	成る	naru	見る	miru	殺す	korosu	行く	iku	分かる	wakaru	1,492		
	Conan	398	to do (A)	294	to be (A)	174	to say	129	to come (A)	111	to become (A)	98	to have (A)	82	to look (A)	73	to kill	70	to go (A)	63	to know	43.18%		
	Naruto	為る	suru	成る	naru	言う	iu	遅る	iru	遅る	yaru	有る	aru	来る	kuru	殺す	korosu	行く	iku	呉れる	kureru	845		
	273	to do (A)	95	to become (A)	90	to say	86	to be (A)	86	to do (A)	55	to have (A)	55	to come (A)	36	to kill	36	to go (A)	33	to give (A)	33	to look (A)	39.21%	
	OnePiece	為る	suru	言う	iu	遅る	yaru	来る	kuru	成る	naru	行く	iku	有る	aru	居る	iru	見る	miru	呉れる	kureru	1,310	to give (A)	38.14%
	314	to do (A)	177	to say	153	to do (A)	121	to come (A)	117	to become (A)	106	to go (A)	104	to have (A)	98	to be (A)	65	to look (A)	55	to give (A)	55	to give (A)	38.14%	

**Table 24:** Most frequently seen adjectives  
**Table 25:** Most frequently seen adverbs

Genre	Series	Top 10 Adjectival Forms																									Total	%	
Shounjo	Boku no ga	無い	ない	良い	よい	好き	suki	悪い	わるい	凄い	sugoi	嫌	yao	みたい	mitai	可愛い	kanwai	まじ	maji									521	
	Ito	173	(negation)	139	good	40	like/love	36	bad	35	great	28	dislike	22	(like)	18	cute	15	serious	15								62.85%	
	Kimi ni	無い	ない	良い	よい	凄い	sugoi	怖い	kowai	みたい	mitai	悪い	warui	you	嫌	yao	まじ	maji	serious								494		
	Todoke	149	(negation)	147	good	36	great	33	scary	26	(like)	23	bad	23	that	20	dislike	19	serious	18								59.02%	
	Nana	良い	よい	無い	ない	嫌	yao	嫌	you	そう	sou	大丈夫	daijoubu	凄い	sugoi	好き	suki	悪い	warui									766	
	Nodame	263	good	238	(negation)	43	dislike	39	(like)	39	that	32	alright	32	great	31	like/love	26	bad	23								439	
	Contable	127	(negation)	126	good	54	great	37	amoi	悪い	warui	みたい	mitai	嫌	you	大丈夫	daijoubu	そう	sou									59.89%	
	Robukon	良い	よい	無い	ない	好き	suki	みたい	mitai	悪い	warui	可愛い	kanwai	嫌	yao	小さい	chisai	面白い	omoshiroi									649	
		202	good	151	(negation)	66	like/love	45	(like)	41	bad	34	cute	32	dislike	29	small	25	interesting	24								60.20%	
	Shounen	DeathNote	無い	ない	良い	よい	様	you	可能	kanou	早い	hayai	悪い	warui	凄い	sugoi	面白い	omoshiroi	大丈夫	daijoubu									586
		252	(negation)	98	good	87	(like)	27	possible	27	early	22	bad	21	great	18	interesting	17	alright	17								57.91%	
Gintama		無い	ない	良い	よい	様	you	悪い	warui	早い	hayai	好き	suki	みたい	mitai	大丈夫	daijoubu	そう	sou									659	
		327	(negation)	141	good	40	(like)	32	bad	21	early	21	like/love	21	(like)	20	alright	20	that	16								57.40%	
Detective		無い	ない	良い	よい	様	you	早い	hayai	悪い	warui	そんな	sonna	確	aka	みたい	mitai	凄い	sugoi									431	
Conan		189	(negation)	55	good	41	(like)	32	early	23	bad	21	such	20	certain	18	(like)	17	great	15								56.19%	
Naruto		無い	ない	良い	よい	様	you	みたい	mitai	早い	hayai	凄い	sugoi	強い	tsuyoi	大丈夫	daijoubu	悪い	warui	そんな	sonna							306	
		106	(negation)	68	good	30	(like)	19	(like)	18	early	16	great	15	strong	12	alright	12	bad	10								51.43%	
OnePiece		無い	ない	良い	よい	強い	tsuyoi	悪い	warui	様	you	そう	sou	早い	hayai	はばばら	hababara	大丈夫	daijoubu	凄い	sugoi	みたい	好き	suki				540	
		215	(negation)	135	good	27	strong	26	bad	22	(like)	21	(like)	18	early	16	separate	15	terrible	15	great	15	(like)	15	like/love				55.84%

Genre	Series	Top 10 Adverbs																				Total	%
		そう	sou	何て	nande	どう	dou	一寸	chotto	もう	mou	全然	zenzen	未だ	mada	矢張り	yohari	こう	まあ	まあ	moo		
Shounen	Boku no ga	そう	sou	何て	nande	どう	dou	一寸	chotto	もう	mou	全然	zenzen	未だ	mada	矢張り	yohari	こう	まあ	まあ	moo		
	Ita	79	thatway	53	why	37	how	35	a little	26	already(!)	17	not at all	17	still	16	after all	15	this way	14	well (!)		
	Kimi ni	そう	sou	一寸	chotto	どう	dou	もう	mou	なんか	nanka	矢張り	yohari	何て	nande	良く	yoku	未だ	mada	ちゃんと	chanto		
	Todoke	121	thatway	60	a little	51	how	36	already (!)	30	like	28	after all	22	why	16	well, often	15	still	12	for sure		
	Nana	そう	sou	どう	dou	何て	nande	もう	mou	矢張り	yohari	一寸	chotto	未だ	mada	又	mata	なんか	nanka	ちゃんと	chanto		
		165	thatway	66	how	57	why	45	already (!)	42	after all	34	a little	25	still	22	again	20	like	18	for sure		
	Nodame	何て	nande	もう	mou	そう	sou	一寸	chotto	34	a little	ちゃんと	chanto	もうと	moto	more	矢張り	yohari	良く	yoku	未だ		
	Contable	69	why	63	already (!)	61	that way	51	how	一寸	a little	26	for sure	23	more	17	after all	16	well, often	15	still		
	Radukon	そう	何て	nande	どう	dou	もう	mou	一寸	chotto	又	mata	目茶	mecha	なんか	nanka	矢張り	yohari	未だ	mada	still		
		130	thatway	66	why	54	how	49	already (!)	49	a little	41	again	31	very	30	like	26	after all	24	still		
Shounen	DeathNote	そう	sou	もう	mou	どう	dou	詰まり	tsumori	何故	naze	若し	moshi	こう	kou	先ず	mazu	未だ	mada	矢張り	yohari		
		106	thatway	68	already (!)	37	how	24	that is to say	23	why	22	if	21	this way	21	first	20	still	20	after all		
	Gintama	もう	mou	どう	dou	そう	sou	一寸	chotto	何て	nande	又	mata	なんか	nanka	未だ	mada	全く	matteku	良く	yoku		
		99	already (!)	69	how	48	that way	48	a little	31	why	30	again	22	like	17	still	17	quickly (!)	15	well, often		
	Detective	そう	sou	どう	dou	もう	mou	一寸	chotto	未だ	mada	まさか	osomaki	まあ	doushie	恐らく	moshika	どうて	moo	急度	kito		
	Canon	141	thatway	63	how	55	already (!)	41	a little	35	still	27	probably	25	why	25	surely (not)	21	well (!)	21	for sure		
	Naruto	そう	sou	どう	dou	もう	mou	一寸	chotto	何て	nande	未だ	mada	一番	ichiban	良く	yoku	又	mata	もうと	motto		
		51	thatway	49	how	43	already (!)	24	a little	17	why	17	still	16	number one	15	well, often	14	again	12	more		
	OnePiece	そう	sou	もう	mou	どう	dou	未だ	mada	まあ	moo	何て	nande	良く	yoku	又	mata	まさか	moshika	矢張り	yohari		
		97	thatway	51	already (!)	48	how	34	still	27	well (!)	27	why	26	well, often	22	again	18	surely (not)	15	after all		

### 3.4.4 Discussion

The morpheme analysis developed here has offered two interesting points to consider: first, how manga are similar (and dissimilar) from other genres in regard to its usage of words, and second, what thematic differences between shoujo-manga and shounen-manga the words seen hint at. Examination of the distribution of parts of speech and the types of words seen have shown that manga appear to be similar to spoken language in their high use of interjections, and in their high use of wago. While manga did not fully overlap with any one genre, as a written text, they seemed closest to diaries in regard to parts of speech, and children's books in regard to word origins. On the one hand, some of these results help to answer the question of what makes a conversation between characters in a series feel like a realistic conversation: In particular, the high percentage of interjection was an obvious similarity with spoken language, making it seem an especially relevant point. On the other hand, their close alignment to children's books reminds one of the readability results, reminding us that they were ranked comparatively low for difficulty, which was largely associated with the non-use of kanji, e.g., the high percentage of wago that makes them similar to children's books.

While the top-10 data is limited by nature, they also made hints of some interesting patterns, particularly between the genres. Whereas shounen-manga consistently had words that appear to be related to their general themes show up on the top-10 lists, shoujo-manga rarely did. Indeed, for the four major types of words—nouns, verbs, adjectives and adverbs—one could find some, if not many, examples of words which clearly bespoke of the major themes of the shounen-manga series, whereas this was clearly less evident in shoujo-manga. As with the results from the orthography and readability studies, this supports the importance of themes to shounen-manga. It is interesting to consider that this is relevant here even though the series are actually very diverse. None of the shounen-manga stories here follow exactly the same pattern, yet—with perhaps the exception of the comedy *GinTama*—they all show similar usages of theme-related lexical items. One way to interpret this is that shounen-manga are more plot-oriented, with the themes coming back lexically in communicating how the plot is moving forward. Shoujo-manga, on the other hand, could be said to be more character-oriented, with most of the vocabulary made up of everyday familiar words, with a special emphasis on personal pronouns and proper nouns due to the interaction between characters.

While the discussion of the methodology involved required lengthy discussions of how ChaMame was used in order to deal with this special kind of text, the details described here also may prove helpful to others hoping to conduct similar analyses. One of the key issues with morpheme analysis software, especially for

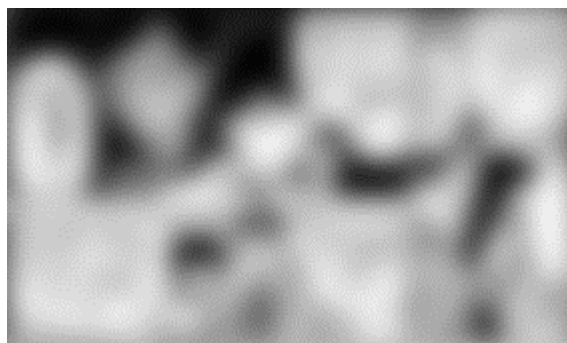
Japanese, continues to be how to adapt them to texts which fall out of their predictive frameworks. The procedures taken here, while not actively changing the internal dictionaries or settings of the programs, offer ways of thinking about the needs of non-standard texts that may make them more accessible to other researchers. In the next chapter, I will give a more in-depth analysis of how these data can be used by looking at how characters actually speak. The analysis will continue with the theme of how shoujo-manga and shounen-manga are characterized by showing distinct differences between the ways that characters are constructed.

There were, however, some problems. A major issue that was *not* dealt with here was that of text split between different speech bubbles. As in the discussion of sentences, the breaking up of the text into small fragments appears to be a characteristic of manga, likely partially brought on by the spatial restrictions of putting text into speech bubbles—which should be noted as another commonality with similarly spatially-restricted keitai-shousetsu (Nishimura, 2011). However, this does lead to problematic interpretations. While ChaMame analyzes what form the verb is in—meaning that, in theory, one could assess the number of sentences (including dependent clauses) by the number of verbs in predicative form (*shuushikei*), as I did to some degree above—it is not always accurate given that without the usage of touten and kuten, it can be unclear where a sentence ends or begins. More importantly, because utterances are often broken up between speech bubbles, the form of the verb is often ambiguous. In particular, when a relative clause is split between its head noun and its dependent clause over two speech bubbles, the verb is assessed as being in the predicative form because no other clues directly follow it. This is obviously incorrect, but requires hand-checking of the data.

While such examples are relatively benign, the real problems occur when the text calls for interpretations. For example, when a character says the first part of an embedded sentence (e.g., the *dependent sentence*), but then abbreviates the noun that is its header, should the verb be assessed as being attributive or predicative? Program-wise, because there is no noun directly following it, the verb is assessed as being in the predicative form. Likewise, if the noun followed in the next speech bubble, it would be possible for hand-checkers to interpret as the attributive form, but when it is abbreviated, this predictable pattern falls apart. Instead, one must arrive at the interpretation that the noun is abbreviated, which may not be universal. A good example of these phenomena may be observed in Figure 55: Ran's fragmented utterance (1) is continued in (2), making the verb in (1) an attributive form rather than a predicative form. This is easily corrected, but one may wonder whether Detective Moori really intended to end his sentence as he did in (3), or whether the 3-point lead points to his intending to follow it with *Masami*, as it did with Ran. Examples like this are a weakness of morpheme analyzers, and are not immediately correctable without thorough manual checks. Further research may require considering how to accommodate these kinds of fragments.



**Figure 55:** Lines split between multiple speech bubbles in *Detective Conan* (2:70)



- (1) Ran      ほら、この前、お父さんを捜して&ほしいって依頼に来た・・・  
 Hora, kono mae, otousan wo      sagasite & hoshii tte irai ni ki-ta...  
 Hey this before father (d-ob.) search (conj.) wants (quote) request (i-ob.) come+past  
 Hey, [the one who] came to request that you look for her father...
- (2) Ran      広田雅美さん&よ!!  
 Hirota Masami-san yo!!  
 Hirota Masami-san (emph.)  
 Masami Hirota!
- (3) 毛利      ああ・・・&山形からわざわざ&来た・・・  
 Aa... & Yamagata kara wazawaza & ki-ta...  
 Aa... Yamagata from purposefully come+past  
 Ahh... [the one who] came out from Yamagata...

### 3.5 Conclusions

In this chapter, I looked at three related aspects of language in manga: their orthographic characteristics; their readability; and their lexical characteristics. It has become obvious that manga have a unique orthographic style that is typified by their overuse of symbols, a large percentage of hiragana, and somewhat lower percentages of kanji compared with other types of writing. I have argued that this is a result of the kind of simulated spoken speech which typifies manga, similar to the modified writing styles Groensteen (2007) has described for western comics. As a partial result of the unique orthographic styles, which utilize less kanji than other media, manga ended up being rated as very easy; however, compared with Nishimura's (2011) data on conventional novels, it seems to only be a few grade levels below, suggesting that it is not as dramatically *easy* as might be assumed. On the other hand, the lexical analysis also showed some interesting trends and patterns, including more similarities with spoken speech in its use of a higher frequency of interjections and *wago*. The most frequently seen nouns also further supported trends I noted in the orthography analysis on the importance of themes in shounen-manga compared with shoujo-manga, which appeared to use more words related to interpersonal relations.

While the studies I reviewed here offered some interesting points to consider in and of themselves, they are just the beginnings of the ways the corpus can be utilized. Particularly, as the morpheme analysis is now available, any number of different properties could be investigated, from particular lexical items to grammatical constructions using concordances. In the next chapter, I specifically discuss several studies that utilize that data, and in that way it is situated as one way to show the potential of these data. More specifically, I will use the results of the morpheme analysis to investigate the use of gendered speech in manga, looking at how realistic it is and comparing differences between genres. Like this chapter moved the analysis one level deeper, the next chapter will thus move the level down one to that of characters and speech patterns.

## 4. Language Patterns and Characterization

### 4.1 Chapter themes and overview

In comparison to the previous chapter, where I looked at the overall lexical trends, in this chapter I take the analysis one layer deeper by examining how language is used at the most basic of levels: How characters express themselves and their identities through language, or, to put it differently, how language is used to create and develop characters in manga. In doing this, I will introduce three studies, two of which use data obtained from the morpheme analysis described in Chapter 3 on personal pronouns and sentence final particles. One keyword that will come up repeatedly is *yakuwari-go*, or role-language: that is, language that is aligned to characters' images, generally in a stereotypical way. *Yakuwari-go* has become a popular topic in Japanese linguistics since first being introduced by Kinsui (2003), and manga have often been described as a likely location of such stereotypical speech patterns. Yet while *yakuwari-go* offer a convenient and easy way to create certain types of characters, there is also pressure in manga to make characters' speech patterns realistic so that readers might relate to them. With comparatively few quantitative studies on the relationship between language and characterization in manga, there are many questions still unanswered about characters' language use in manga. This chapter seeks to rectify this, and with gendered language patterns being both a location of *yakuwari-go* and also of linguistic change in recent years, they are an ideal place to look at how language is manipulated to create characters.

To that point, after an overview of Japanese gendered language, both in general and in manga, I will introduce three studies: the first on the distribution of *Lines* between male and female characters; the second on the use of personal pronouns in *Lines*; and the third on sentence final expressions in *Lines*. What will become obvious is that there are significant differences between *shoujo-manga* and *shounen-manga* in regard to their usage of gendered speech, with female and male characters both speaking more stereotypically feminine and masculine, respectively, in *shounen-manga*. Speech in *shoujo-manga* will come to be seen as relatively naturalistic, with the results not too different from the real-life speech patterns of young women and men. In addition to the issues of authors' genders and plots, I will argue that these differences are related to a gap in the role of male and female characters in the two genres. In particular, where *shoujo-manga* shows a relatively balanced distribution of *Lines* between male and female characters, *shounen-manga* are deeply skewed towards male characters, with most of the female characters not given major speaking roles. With deep differences

between characters' tendencies to use *yakuwari-go* by their centrality, this means that with most female characters in shounen-manga being essentially extras, they are likely to use stereotyped speech patterns.

#### 4.1.1 Japanese gendered language

Today, gendered language is widely considered to be one of the major characteristics of Japanese (Kindaichi (1989), Tsujimura (2007)). Structurally, gendered language in Japanese is non-obligatory, meaning that speakers are not required to use *gender-appropriate* speech patterns in order to form grammatical sentences. Rather, the use of gendered languages speaks of one's social position, one's relationship to one's listener(s), and one's own personal feelings. Many aspects of Japanese are subject to such non-obligatory gendered differences, and they are found in a variety of phonological, morphological and lexical environments. Shibamoto (1987) lists some of the more common examples as intonation; politeness; personal pronouns; and emphatic sentence final particles (SFPs). Personal pronouns and SFPs are particularly interesting in Japanese, in part because they are clearly marked forms that are easily extracted for analysis, but also because these are both elements which appear to be undergoing change.

As Okamoto (1995) reviews, the use of feminine language in Japanese has particular social values associated with it, and the characteristics ascribed to it—such as women's speech being polite, soft-spoken, and nonassertive—are partially interpretable as a reflection of women's lower social status. Perhaps surprisingly, however, the recent shift in women's roles in Japanese society has been accompanied with a marked decrease in the differences in speech patterns between male and female speakers. These changes can be seen in many different locations; one particularly prominent example is that of sentence final particles (SFPs). SFPs are—at risk of sounding redundant—special particles which occur at the end of sentences. SFPs are *non-referential* elements, meaning that they do not change the semantics of the sentences themselves (Inoue M. , 2006, p. 53); rather, they are “an affective and evidential marker with which speakers signal their social attitudes or stances . . . (Inoue M. , 2006, p. 76).” Without a specific linguistic meaning, they thus might be considered to fall under the category of *pragmatic markers*, or “. . . the linguistically encoded clues which signal the speaker's potential communicative intentions (Fraser, 1996, p. 168).” Generally, they give emphasis or show the relationship between speakers, and are commonly categorized into feminine, masculine and neutral forms that are thought to be primarily used by the respective genders. SFPs may be used together, and the categorization of a given SFP depends on the other lexical items it is being used with. *Yo*, for example, can be neutral, feminine or masculine

depending on the context: When it appears after a verb in the polite *masu* form, it is considered neutral, but when it is after a noun, it is considered feminine, and after a verb in the plain form, it is considered masculine. Some examples of feminine forms are *wa* with a rising intonation; *no* followed by the particles *yo/yo-ne/ne*; and *ne* following nouns. Some masculine forms are *ze* and *zo*; and *yo* after the command form of the verb. Neutral forms include *yo-ne* after verbs and adjectives.

Numerous studies on changes in the use of SFPs have recently come out, however, and it has been widely reported that the female-specific forms are essentially out of use amongst young women, with clear differences in the percentages used by older and younger women (Philips (2001), Ozaki (1999), Kobayashi (1993), Okamoto (1995), etc.). Even more interesting, however, is the fact that while the feminine forms appear to be on their way to extinction in colloquial speech, some young women are choosing to use some masculine SFPs (e.g., Philips (2001), Okamoto (2005)). These changes seem to suggest that these *traditional*—with caution as regards the meaning of traditional, given the strong arguments to consider gendered speech differences as the creation of the Meiji period and the modern state (Nakamura M. , 'Onna-kotoba' wa Tsukurareru, 2007)—distinctions are becoming less important, causing some such as Kobayashi (1993) to suggest that Japanese women's speech is undergoing a process of neutralization (*chuuseika*).

In comparison with SFPs, personal pronouns have long seemed to be holdouts in gendered speech, showing relatively little change or diversification. Unlike standard English, which is restricted to one form for each pronoun, Japanese features many forms of first and second person pronouns, and their choice depends not only on gender, but also on formality, the speaker's relationship with the listener, and their own preferences. For first pronouns, women tend to use *atashi* or *watashi*, whereas men tend to use *ore* or *boku*. For second person pronouns, women tend to avoid them altogether or use *anata*, and men tend to use *omae* or *kimi*. While there are no exclusively feminine forms, *kimi*, *omae*, and *temee* are generally considered masculine forms (Shibamoto J. , 2005). Because Japanese is a pro-drop language, meaning that it allows subjects (and, to a lesser degree, objects) to be dropped when context is sufficiently rich to make the dropped words clear, personal pronouns are generally used less frequently in Japanese than they are in languages such as English. Examining the first chapter of the *Little Prince* in several different languages, Nakano (1976), summarized in Hayashi, et al. (1982), found that pronouns took up just 3.1% of substantives, compared to 20.9% in English and 20.7% in French. There is a tendency to avoid using personal pronouns at all, with the use of second person pronouns often considered to be rude: While there are second person pronouns for inferiors or equals, there is no such form for superiors, making them thus inappropriate for such usages (Whitman, 1999). When they cannot be avoided, it is common to use names or titles instead of second person pronouns. It is also possible to refer to oneself by one's name, although this is a habit largely associated with young women (Kajino & Podesva, 2007).

As Abe (2004) describes, personal pronouns are generally thought to be the strongest index of gender in Japanese, and as such, are the slowest to change. In reflection of that, as both Ozaki (2002) and Hishikari (2007) report, there are no first person pronouns used in common with both male and female young adults in Tokyo of junior-high-school to college-age. However, the situation may not be that clear-cut. While young girls do not appear to be using them *en masse*, there have been several reports of young women using the masculine first person pronouns *boku* and *ore* at some point in their lives. In a survey of a 136 female students, Endo (2001) found that 23 students have used *boku* and *ore* at some point; 72 students reported observing other girls use them, too. Miyazaki (2004) also observed their being used amongst some junior high schools girls. While they may be a distinct minority, it appears that some girls are now choosing to use masculine pronouns. People also appear to be conscious of changes in the usage of personal pronouns, with girls using *boku* and *ore* labeled *boku-kko* (*boku-girls*) and *ore-onna* (*ore-ladies*) (Shinomoto, 2008). Changes in girls' personal pronoun patterns also appear to be a point of popular concern, with many discussions on forums such as Yahoo Chiebukuro—literally, 'Yahoo Bag of Wisdom,' otherwise known in the English-speaking world as Yahoo Questions—concerning girls' usage of *boku* and *ore*, which I will describe further below. With changes in their usage patterns seemingly looming in the horizon, and with deep and wide impacts on gendered language, personal pronouns are an element to follow in these coming years.

#### 4.1.2 Gendered language and manga

Manga are in many ways a natural place to look for language change given that they are one of the most important linguistic resources in Japan today. As I argued in Chapter 1, the most popular type of manga, story-manga, are characterized by strong narrative structures (Takeuchi I. , 2006), making them heavily reliant upon linguistic elements to express their narratives. Even with sales numbers going down from their previous highs (Nakano, 2009), there is no doubt that they form one of the most widely read bodies of text in Japan today. It seems but natural to think that they should have an impact on the Japanese language; indeed, this appears to be a common line of thought, with manga often popularly noted as one location of linguistic change amongst young Japanese. As noted previously, according to a 2008 survey conducted by the Agency for Cultural Affairs (2010), approximately 45% of people thought that manga were influential in young people's speech, second only to TV, family and friends. This maintained the results from the Agency's (2001) survey, where 44.6% of respondents gave manga as a linguistically influential medium.

Popular attention has been paid particularly to the role that manga may play in the changes to gendered speech patterns observed above. For example, one often sees discussions of manga and their influence on young women's choice of the masculine first person pronouns *ore* and *boku*. In her surveys of female college students' and their and their friends' use of masculine personal pronouns, Endo (2001) found that some students attributed such usages to the influence of manga. Likewise, Nakamura ('Sei' to Nihongo: Kotoba ga Tsukuru Onna to Otoko, 2007) describes the usage of *ore* amongst female elementary school children, who associated it with the popular *anime* (animated cartoon) *Crayon Shin-chan*, whose rambunctious, mischievous and very crude (male) protagonist, Shin-chan, uses *ore*. In his analysis of *boku-kko*, Nishida (2011) also claims that *boku-kko* are common in manga, and calls upon several examples of them in his analysis.

Such explanations are also commonly found in popular discourse. Looking at *Yahoo Chiebukuro*, one finds many posts on the role of manga in girls' use of *boku* and *ore*. Searching with the key words *onna no ko* (girl), *ichininshou* (first person pronoun), *boku* or *ore* and manga, controlling for orthographic differences through repeat searches, led to 89 different questions, of which 29 were related to questions about specific manga and/or instances of female characters using *boku/ore*, and 43 were questions concerning actual usage of *boku/ore* amongst women, with manga coming up in the discussions. (The remaining 17 were only incidentally related to these topics.) The earliest question was from October, 2004, and my search was conducted in September, 2012. Amongst the 43 relevant questions about personal pronoun usage, the discussions would often point to manga as a reason for why girls would use *boku* and *ore*. To answer a question as to why men use *ore* and women use *watashi*, one user wrote "I know girls who use *boku* because of manga's influence (Anonymous, 2006)." Out of four responses to another user's post asking if any girls around them used *boku* or *ore*, two attested to their usage by girls around them who read manga (Taruruda, 2007). Another user, self-describing herself a female high school senior, asked for opinions on girls' usages of *ore* and *boku*, and herself attested such patterns to manga; additionally, two out of nine responses commented that manga were influential (Misamisa091v, 2009).

However, it is not entirely clear how influential manga may actually be. While the posts on *Yahoo Chiebukuro* show people's associations between manga and girls' usage of masculine personal pronouns, they do not specifically say what about manga might lead girls to do so. Do female characters in manga frequently use male forms? Are female readers copying the speech patterns of male protagonists? Or are there other issues at work here? These kinds of questions have been very difficult to assess objectively given the lack hitherto of empirical research on language in manga. While offering interesting and suggestive results, works such as Endo's (2001) or Nakamura's (2007) have focused on conscious surveys and interviews amongst young people on their personal pronoun usage and why they chose their personal pronouns, or individuals' own anecdotal usages. What is known from such surveys is only that manga are thought to be (potentially negatively) influential,

but we do not know anything about how language is actually being used in manga. While this tells us much about how manga are *perceived* in Japanese society, in evaluating such claims it is crucial that one have a clearer grasp of how personal pronouns—and more widely, other aspects of gendered speech—are actually used. Even Nishida's (2011) study, while citing real examples of female characters using *boku* and *ore* in manga, is limited in that he fails to discuss what *kind* of manga he is using, and instead treats all manga as one type, which is clearly not the case.

In considering what kind of a linguistic resource manga might be, it is important to reflect upon its nature as fiction. As a popular linguistic resource, manga must maintain a certain level of realism in order for readers to be engaged by them, and as Ueno (2006, p. 17) describes, manga may “influence readers in framing their gender identities as well as their linguistic behavior.” As noted earlier, if the speech it uses is too unrealistic, readers may not be able to fully relate to the characters. Linguistic realism being important, one might anticipate great pressure on authors to make their speakers appear naturalistic. Yet at the same time, as works of fiction, manga are constantly trying to find balance between the needs for realism and the utilization of linguistic conveniences such as stereotypical speech patterns.

As I suggested earlier, of relevance here is the concept of *yakuwari-go*, which has become popular in recent years. Defined by Kinsui (2003, p. 205) as “speech patterns which are aligned to the images of their speakers,” *yakuwari-go* could be thought of a specific kind of register, or a linguistic variety “associated with particular situational contexts or purposes (Biber, 1995, p. 1).” Registers have often received attention for their role in creating social identities, such as how the use of medical terminology by doctors helps define them socially as professionals, and *yakuwari-go* are similar in that they help authors create easily recognizable identities for characters. While registers are parts of natural speech, however, *yakuwari-go* are as if they had been taken one step further, so that while they help identify characters, they no longer are necessarily naturalistic. Kinsui (2007, p. 98) has argued that they are particularly prevalent in manga, to which he says they are “essential”. This has resulted in a renewed interest in language patterns in manga, accompanied by several new reports on *yakuwari-go*. Takahashi (2009) showed how different (female) characters are made to speak depending on their character type using the shoujo-manga series *Life* as a case study. Jung (2005) looked at the ways that *yakuwari-go* are used in Japanese manga and their Korean translations, finding that where the original Japanese texts tended to use more *yakuwari-go* related to gender, their Korean translations tended to use more *yakuwari-go* related to speakers' ages. Kurosaki (2011) focused on the use of the non-standard masculine first-person pronoun *washi*, used commonly in manga and anime amongst professorial characters, looking for its origins as *yakuwari-go* in dialects and Edo-period texts.



The tension between realism and stereotype makes one wonder how gendered speech is really used. While the studies described above all pointed to the commonness of *yakuwari-go*, others have suggested that speech patterns in manga may be comparatively realistic. Ueno (2006) examined the use of SFPs in *shoujo-manga* and *rediusu-manga* (ladies'-manga; manga for female readers one generation older than *shoujo-manga* readers), and found that the use of strongly feminine SFPs goes up amongst with characters' ages in both genres, similar to the gradated patterns that are expected. Young girls in *shoujo-manga* particularly used SFPs that are comparatively realistic; but they overused masculine forms in ladies'-manga. Older characters also overused strongly feminine forms in *shoujo-manga*. Ueno attributes this gap partially to differences in how older and younger women perceive each other, but it is also similar to the results that Okamoto (1995) and Philips (2001) found for real life SFP usage: That is, younger women are using feminine forms less and masculine forms more. In examining the use of SFPs in three series published approximately 15 years apart starting in the 1970s, Aizawa (2003) found that feminine SFPs went down over time, from the oldest to newest series, suggesting a drop in feminine forms similar to that which was found in real life. Similarly, Chinami (2001) found that while female characters still tended to use feminine SFPs, they appeared to be on the decline. Akitsuki (2006) also notes that recent *shoujo-manga* such as *Nana* represented a departure from traditional characterization with their use of masculine-speech heroines, which he suggested might form a new type of *yakuwari-go*.

The use of gendered speech and *yakuwari-go* in manga thus appear to be extremely complex, and there remain some major gaps in our understanding. While there have been some quantitative studies on the use of language in manga such as Ueno (2006), they are exceptions rather than the rule. Most research on *yakuwari-go* has been qualitative, making it particularly difficult to evaluate whether claims that *yakuwari-go* are common in manga are well founded. Even Ueno's study is limited in that it targeted only series for female readers, and looked exclusively at female speech; thus there is very little knowledge of what male speech is like aside from what Kinsui (2003), (2007) originally described. One can expect, however, that there are differences between how characters are depicted between the genres. As Ueno's (2006) study showed, female characters differed in usage patterns even within genres aimed for women, and it is not hard to imagine that this would be the case between *shoujo-manga* and *shounen-manga* as well; there is no reason to think that such differences would not also exist for male characters.

How characters utilize gendered speech in manga is clearly central to how characters are constructed, and the corpus used here presents a unique opportunity to look at how *all* of the characters—male and female alike—use gendered speech in representative series from *shoujo-manga* and *shounen-manga*. To consider how some of the issues described play out, I used the data from the morpheme analysis described in Chapter 3 to conduct three surveys on (1) the use of *Lines* by different gendered characters; (2) the use and distribution of

personal pronouns; and (3) the use and distribution of SFPs. Personal pronouns and SFPs were targeted in (2) and (3) because they are both considered key elements of yakuwari-go (Kinsui, 2003), and because of their obvious importance in gendered language. However, it is also important to understand how much text is actually being used by male and female characters—the goal of (1)—as this would be expected to have an impact on the types of text characters will use. As I will describe in more detail below, the role of characters in the text, and particularly their centrality, has impacts on these issues (Kinsui, 2003). Thus, examining the distribution of *Lines* themselves is a necessary first step as it offers crucial background information on how characters are being manipulated. While it will become obvious that, particularly in shounen-manga, yakuwari-go speech patterns are not unusual, it is important to consider what type of characters are more prone to using them, and whether this effects how commonly such forms are seen.

## 4.2 Study 1: Distribution of Lines by sex and number of characters

### 4.2.1 Methodology

This study looked at two important and related factors: first, the distribution of *Lines* between the sexes, and second, how many characters can be observed. As regards the first question, I noted for all *Lines* whether their speakers were *male*, *female*, *both/unclear*, *onomatopoeia* or *animal*, and calculated the number of entries and characters for each group, using the raw data from the corpus. Using the data procured from the morpheme analysis described in Chapter 3, I also calculated the number of words spoken by each group, except for onomatopoeia, as they were not included as targets for the morpheme analysis. The results for the amount of text spoken by each sex in the two genres are listed in Table 26 below.

**Table 26:** Number of *Lines* by the sex of speakers and genre

Genre	Sex	Entries		Characters		Words	
Shoujo-manga	Male	7,076	43.57%	91,481	41.67%	46,844	43.07%
	Female	8,881	54.69%	125,660	57.24%	60,944	56.04%
	Both/Unclear	217	1.34%	2,173	0.99%	968	0.89%
	Onomatopoeia	64	0.39%	194	0.09%	-	-
	Animal	1	0.01%	5	0.00%	2	0.00%
	Total	16,239		219,513		108,758	
Shounen-manga	Male	14,284	79.13%	232,026	82.73%	126,869	83.39%
	Female	2,996	16.60%	43,978	15.68%	23,945	15.74%
	Both/Unclear	191	1.06%	2,202	0.79%	1,138	0.75%
	Onomatopoeia	489	2.71%	1,697	0.61%	-	-
	Animal	92	0.51%	552	0.20%	196	0.13%
	Total	18,052		280,455		152,148	
Grand total	Male	21,360	62.29%	323,507	64.71%	173,713	66.58%
	Female	11,877	34.64%	169,638	33.93%	84,889	32.54%
	Both/Unclear	408	1.19%	4,375	0.88%	2,106	0.81%
	Onomatopoeia	553	1.61%	1,891	0.38%	-	-
	Animal	93	0.27%	557	0.11%	198	0.08%
	Total	34,291		499,968		260,906	

As regards the second question, I took the names and sexes assigned in the original corpus, and calculated the number of *Lines* and text characters for each individual character, as well as the number of words

spoken by each character from the morpheme analysis data. In order to test the role of peripheral characters in the two genres, I also divided characters into those who were identifiable, and those who were extras. For the purposes of this study, I defined *identifiable characters* as individuals with 10 or more *Lines* and a total of over 150 text characters. Characters with fewer *Lines* than that were determined to be *extras*: With an average of 6.48 *Lines* per page, such characters are unlikely to be part of multipage dialogues. The number of characters seen in each series is listed in Table 27.

**Table 27:** Number of characters seen in each series

Genre	Series	Sex				Total
		Male		Female		
Shoujo-manga	<i>Bokura</i>	26	47.27%	29	52.73%	55
	<i>Kimi</i>	47	46.53%	54	53.47%	101
	<i>Nana</i>	22	57.89%	16	42.11%	38
	<i>Nodame</i>	59	51.30%	56	48.70%	115
	<i>Rabukon</i>	27	50.00%	27	50.00%	54
	Total	181	49.86%	182	50.14%	363
Shounen-manga	<i>DeathNote</i>	119	86.23%	19	13.77%	138
	<i>GinTama</i>	104	88.14%	14	11.86%	118
	<i>Konan</i>	61	66.30%	31	33.70%	92
	<i>Naruto</i>	42	91.30%	4	8.70%	46
	<i>One Piece</i>	200	95.69%	9	4.31%	209
	Total	526	87.23%	77	12.77%	603
Grand Total		707	72.89%	262	27.01%	970

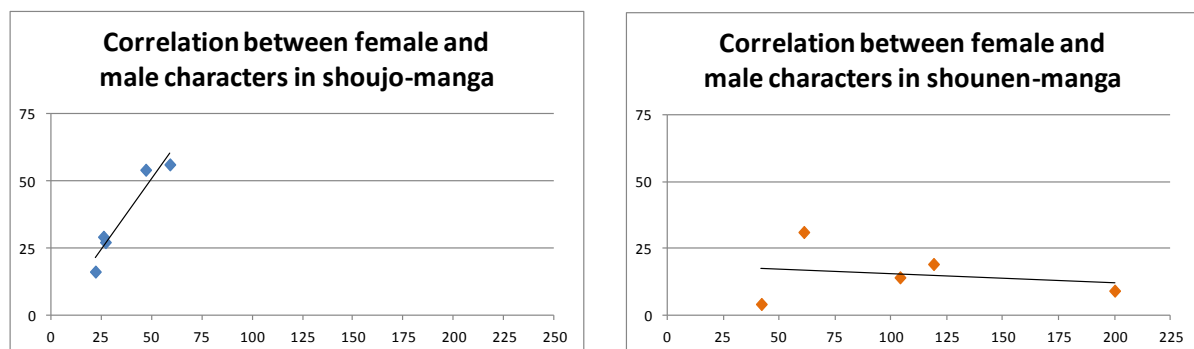
## 4.2.2 Results

The data were generally skewed towards male characters, who made up 62.29% of entries as compared with female characters, at 34.64% (Table 26). This percentage was slightly higher for characters (64.71%), suggesting that male characters' *Lines* were slightly longer than female characters'. Entries whose speakers were unclear, or were a mixed group, as well as onomatopoeia and *Lines* spoken by animals were only incidental (0.27~1.61%). However, this imbalance between the sexes is mostly a reflection of the distribution found in shounen-manga, where female characters' *Lines* only accounted for an average of 16.60% of entries. Male characters accounted

for 79.13%. This is somewhat predictable: One might hypothesize that the main characters are aligned with the gender of their target audiences, leading to more *Lines* for that sex because the main characters will likely have a larger percentage of the overall text. To that point, shoujo-manga also does show a similar gap, but it is remarkably less so: *Lines* by male characters came to 43.57% of entries, and female characters 54.69%. Some of the data also suggest that more than the alignment of the characters with their gendered readers, the sex of the main characters determines the distribution of the text: One shoujo-manga series, *Nodame Cantabile*, actually features more male *Lines* (61.19%), and its main character could be argued to be a male character, Chiaki, who individually counts for 24.25% of all characters (Table 28). This skew towards male characters in *Nodame*, however, is not as dramatic as that of shounen-manga.

It would appear that the skew described above is actually more of a result of a difference in the numbers of characters, and the number of male and female characters in particular. Shounen-manga tended to have a large cast of characters, averaging 121 per series (range: 46~209), compared to 73 in shoujo-manga (range: 38~115) (Table 27). While this was not statistically significant ( $t(8) = 1.56$ ,  $p = .1576$ ), there was a strong positive correlation in shoujo-manga between female and male characters ( $r = .958$ ,  $N = 10$ ,  $p = .009^*$ ); Graph 5), meaning that as the number of characters increases, they tend to increase for both sexes. While there was a moderate negative correlation between female and male characters in shounen-manga, this was not significant, likely due to the small number of cases ( $r = -0.206$ ,  $N = 5$ ,  $p = .7389$ ). Given the limited number of cases, this is not unexpected, but as Graph 6 suggests, *none* of the shounen-manga series showed a balanced ratio of female to male characters as is seen in shoujo-manga. These points make it clear that (1) shounen-manga feature more male characters than shoujo-manga, (2) shoujo-manga feature more female characters than shounen-manga, and (3), shoujo-manga have a more balanced ratio between male-to-female characters.

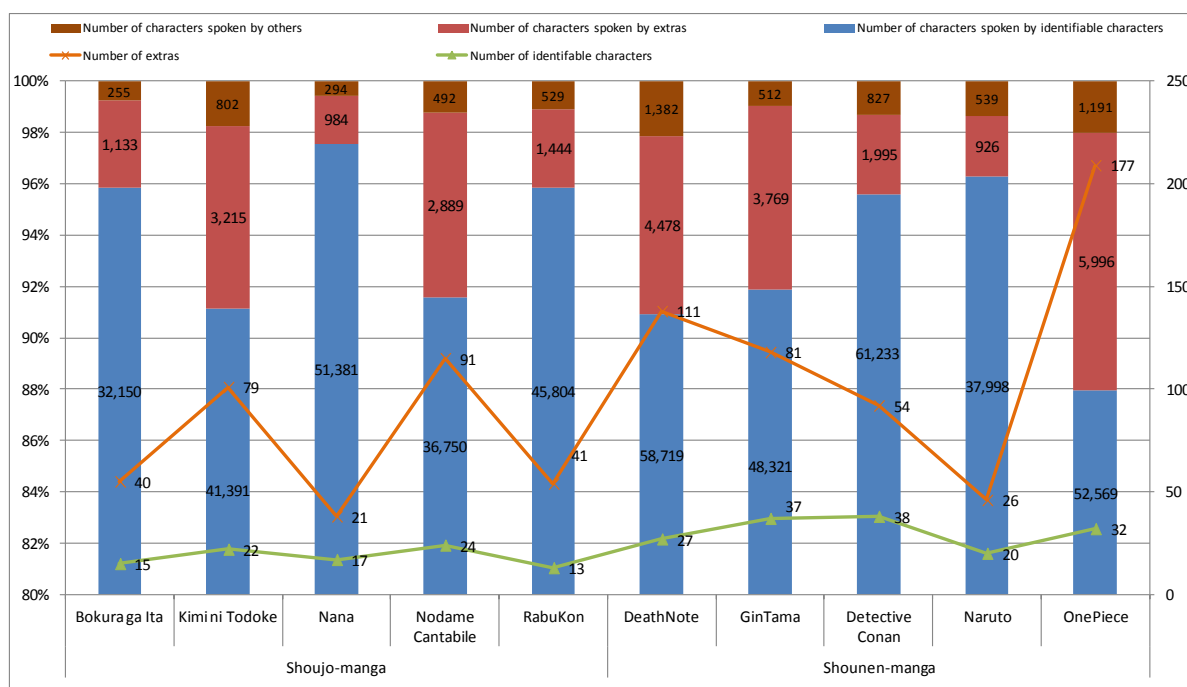
**Graph 5 (Left):** Scatter plot of the correlation between female and male characters in shoujo-manga  
**Graph 6 (Right):** Scatter plot of the correlation between female and male characters in shounen-manga



The differences between shoujo-manga and shounen-manga become even starker when one looks at the number of identifiable and extra characters (Graph 9). First, at 30.8 characters on-average, shounen-manga tended to feature a much larger cast of identifiable characters. Only *Naruto*, at 20 characters, was exceptional; however, it may be consistent in its exceptionality, having also been so in its use of *Background Lines/Text* (Chapter 2). In comparison, shoujo-manga averaged just 18.2 identifiable characters; this difference was significant ( $t(8) = 3.20, p = .0126^*$ ), suggesting that shoujo-manga focus on a smaller group of core characters. While the overall percentage of characters taken up by identifiable characters did not differ significantly between shoujo-manga and shounen-manga, this also suggests that the average amount of text spoken by identifiable characters is longer in shoujo-manga. They did indeed average longer than shounen-manga (2,279.96 vs. 1,680.78 text-characters), and while this difference was not clearly statistically significant, it was trending ( $t(8) = 1.74, p = .1195$ ).

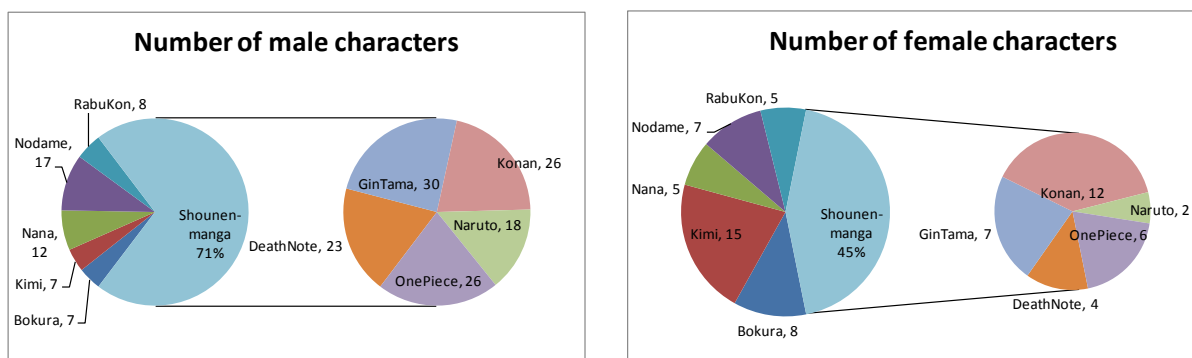
Second, there were important differences in the gender of characters (Graph 7, Graph 8). The low ratio of female to male characters is even more obvious when looking at identifiable characters. In shounen-manga, there were many identifiable male characters, with the minimum 18 (*Naruto*) and the maximum 30 (*GinTama*); yet identifiable female characters were limited, ranging from as low as 2 (*Naruto*) to as high as 12 (*Detective Conan*). Even at their highest, the number of female characters is several points below the lowest number of male characters. Shoujo-manga were more balanced; in some series they even skewed towards males, with a minimum of 7 (*Kimi ni Todoke*) to a maximum of 17 (*Nodame*). Female characters ranged from 5 (*Nana*, *RabuKon*) to 15 (*Kimi ni Todoke*). This suggests that female characters may play a less central role in the narratives of shounen-manga.

**Graph 9:** Distribution of *Lines* by character types



**Graph 7 (Left):** Number of identifiable male characters by series

**Graph 8 (Right):** Number of identifiable female characters by series



Additionally, there are also differences in how much characters speak: Specifically, the number of *Lines* spoken by the main (=the character with the most *Lines*) male and female characters differed within the genres. All the shoujo-manga series had at least one male and female character with more than 10% of all *Lines*. The main female characters accounted for up to 37.62% (Risa Koizumi in *RabuKon*); the lowest as 22.60% (Nodame) in *Nodame Cantabile*, but this is likely because the main character could be argued to be male (Chiaki: 24.25%). The lowest percentage for a main male character was 12.77% (Shouji) in *Nana*, which has a relatively high

number of male characters, and the maximum 29.81% (Yano in *Bokura ga Ita*). On the other hand, all shounen-manga but *One Piece* had two or three male characters covering 10% of *Lines*, but only two have female characters over 10%, topping off at just 10.16% (Nami in *One Piece*).

**Table 28:** Percentages of text (characters) spoken by the main male and female characters for each series

Genre	Title	Male			Female		
		Name	Total	% Text	Name	Total	% Text
Shoujo-manga	<i>Bokura</i>	Yano	9,999	29.81%	Nanami	11,999	35.78%
	<i>Kimi ni Todoke</i>	Kazehaya	6,381	14.05%	Sawako	12,827	28.25%
	<i>Nana</i>	Shouji	6,722	12.77%	Nana	17,059	32.40%
	<i>Nodame Cantabile</i>	Chiaki	9,733	24.25%	Nodame	9,069	22.60%
	<i>Rabukon</i>	Ootani	12,188	25.51%	Risa	17,976	37.62%
Shounen-manga	<i>DeathNote</i>	Light	18,607	28.81%	Naomi	2,518	3.90%
	<i>GinTama</i>	Sakata	12,424	23.62%	Kagura	3,370	6.41%
	<i>Detective Conan</i>	Conan	13,846	21.62%	Ran	6,455	10.08%
	<i>Naruto</i>	Kakashi	9,139	23.16%	Sakura	3,508	8.89%
	<i>One Piece</i>	Luffy	9,742	16.30%	Nami	6,071	10.16%

### 4.2.3 Impact of the distribution of characters on speech patterns

While a more detailed analysis of these issues will be saved for the combined discussion in Section 4.5, the data presented here lead to several predictions for gendered speech in manga. As has become obvious, shoujo-manga and shounen-manga differ primarily in the number of characters seen and number of characters for both genders. With female characters at a minimum—and relegated to primarily secondary or extra roles—one can expect that shounen-manga will feature different linguistic patterns from female characters in shoujo-manga, where there are many more central female characters. This is because of the relationship between *yakuwari-go* and peripherality. As Kinsui (2003) has argued, characters who are less central to the narrative are more likely to use *yakuwari-go* than central characters. That is because there is a performance cost involved in *yakuwari-go*. Having a main character utilize *yakuwari-go* would require great effort on the part of the author to be consistent in their speech patterns, and to maintain that overtime; failure to follow the stereotypes that one is calling upon properly would have negative effects by breaking the spell, so to speak. It is less likely that such gaps would be noticed for peripheral characters, being out of the spot-light and speaking less overall. For these reasons, one can



hypothesize that shounen-manga, with its larger cast of secondary characters, is more susceptible to yakuwari-go. Furthermore, because female characters are primarily extras, they will be particularly susceptible to yakuwari-go. On the other hand, because shoujo-manga feature fewer characters who could be potential targets for yakuwari-go, and because male and female characters both have strong presences, one can hypothesize that shoujo-manga will appear more similar to natural speech.

## 4.3 Study 2: Distribution of personal pronouns in Lines

### 4.3.1 Methodology

This study used the data procured from the morpheme analysis described in Chapter 3 from the word class categorized as pronouns (*daimeishi*). Pronouns include a variety of related types, including demonstrative pronouns such as *sore* (that) and indefinite pronouns such as *nani* (what); however, the actual targets were limited to first and second person pronouns. Third person pronouns were exempt as they do not feature a variety of forms, are not as clearly identifiable in terms of gender. All of the targets were identified and labeled masculine, feminine or neutral, largely based upon the groupings in Shibamoto (1987) and Shibamoto (2005), with some modifications; Table 29 lists forms of first and second person pronouns and their gender assignments. All orthographic variations were included. Some phonetic variations thought not to be separate forms were grouped together (e.g., *omae* vs. *omee*). While names can be used as substitute forms for first and second person pronouns, they were targeted here due to the difficulty of distinguishing between uses. Note also that I did not differentiate between plural forms, as aside from *uchi* and *wareware*, they are usually modified from the singular forms through the addition of suffixes (e.g., *watashi* can be used for the plural by adding *tachi*: *watashi-tachi*). Finally, in order to make accurate judgments as to the role of gender in their distribution, only forms used by male and female characters were utilized; *Lines* whose speakers were unclear or were spoken by mixed groups, as well as those spoken by animals, were not included. Occasionally, characters quote other individuals directly, but as the language patterns seen within quotations may be used for very particular functions (see Maynard (2002) for a detailed analysis of the non-quotative use of quotations), personal pronouns found in such contexts were removed as targets using quotation marks and the quotative markers *tte* and *to*.

**Table 29:** Types of personal pronouns

Status	Singular			Plural	
	Feminine	Masculine	Neutral	Neutral	Masculine
1st Person Pronouns					
In Use	atashi	boku	watakushi	uchi	
	uchi	ore	watashi		
	(speaker's name)	jibun			
Dialectical		washi			
Out of use/limited		ware			wareware
		yo			
2nd Person Pronouns					
In Use		omae	anata		
		temee	anta		
		kimī			
Dialectical		ware			
Out of use/limited		kisama			

**Table 30:** Number and types of personal pronouns seen by genre, series and gender (details)

Genre	Series	Sex	1st Person Pronouns										2nd Person Pronouns														
			Feminine					Neutral					Masculine					Neutral					Masculine				
			atashi	uchi	watashi	watashi	boku	ore	ware	washi	yo	uchi	wareware	anata	anta	nanji	onushi	otaku	kimi	kisama							
Shoujo-manga	Bokura ga	Male				1		95				2			1	26		1									
	Ita	Female	106			1						6															
	Kimi ni	Male						79				1															
	Toodoke	Female	88	2		85						2			76												
	Nano	Male				3	15	125				2	1		1				13								
		Female	222		1							5			3	92											
	Nodame	Male			1	51	21	110				6			5	12			25								
	Contable	Female				56						1			2	3											
		Male	5			13	6	133				2							1								
	Robukon	Female	129		1	10						1			1	62											
Shounen-manga	Total	Male	5	0	1	68	42	542	0	0	0	13	1	5	13	1	0	1	39	0							
		Female	545	2	2	152	0	0	0	0	0	15	0	7	259	0	0	1	0	0							
						124	120	59				8	26	41	3			1	7								
	Death	Male																									
	Note	Female	4			34						1		12	2												
		Male				5	44	188	6	7	11	13	5	4	38		1	2	16	23							
	Gintama	Female	1			60		8				4		9	14												
	Detective	Male				60	33	79	1	19		1	3	78	14				34	10							
	Conan	Female	5			97						1		31	17		2										
	Naruto	Male				7	34	155		18		1	4	6	9				32	1							
Variations seen		Female				20									8												
	One Piece	Male				18	32	361	1	34			7	15	7				27	22							
		Female	6			62						1		2	29				1								
	Total	Male	0	0	0	214	263	842	8	78	11	23	45	144	71	0	1	3	116	56							
	Female	16	0	0	273	0	8	0	0	0	6	1	54	70	0	0	2	1	0								
Variations seen			atashi, atasha	uchi-ra			boku, bokwa	ore, ora		washi, wosha																	

## 4.3.2 Results

A total of 11 different forms of first person pronouns and person pronouns, each, were found (Table 30). This includes 2 plural first pronouns (*uchi* and *wareware*); however, in the discussion below the statistics given will only be for the singular forms unless otherwise noted. There were 3,072 instances of first person pronouns (male: 2,074, female: 998), and 2,025 instances of second person pronouns (male: 1,563, female: 462). There were no significant differences between the number of pronouns used by male and female characters by genre; there was also no significant difference between the number of first person pronouns used by male and female characters. Male characters, however, used a higher number of second person pronouns per total words than female characters (0.90% vs. 0.54%), a difference which was found to be significant using the t-test ( $t(8) = 2.88, p = .0099^*$ ). This seems consistent with reports that female speakers generally try to avoid using second person pronouns more than males (Peng (1973), reported in Shibamoto (2005)).

**Table 31:** Types of personal pronouns used by female characters

Genre	Series	1st Person Pronouns										2nd Person Pronouns								
		Feminine			Masculine			Neutral		Total		Masculine			Neutral		Total			
		#	%1P	%W	#	%1P	%W	#	%1P	%W	#	%W	#	%2P	%W	#	%2P	%W	#	%W
Shojo-manga	Bokura	106	99.07%	1.09%	-	-	-	1	0.93%	0.01%	107	1.10%	2	6.67%	0.02%	28	93.33%	0.29%	30	0.31%
	Kimi ni Todoke	90	51.43%	0.00%	-	-	-	85	48.57%	0.59%	175	1.22%	1	1.30%	0.01%	76	98.70%	0.53%	77	0.54%
	Nana	222	99.55%	0.01%	-	-	-	1	0.45%	0.01%	223	1.28%	23	19.49%	0.13%	95	80.51%	0.55%	118	0.68%
	Nodame Cantabile	-	-	-	-	-	-	56	100%	0.81%	56	0.81%	1	16.67%	0.01%	5	83.33%	0.07%	6	0.09%
	Rabukon	129	92.14%	0.01%	-	-	-	11	7.86%	0.09%	140	1.11%	-	-	-	63	100%	0.50%	63	0.50%
	Total	547	78.03%	0.00%	-	-	-	154	21.97%	0.25%	701	1.15%	27	9.18%	0.04%	267	90.82%	0.44%	294	0.48%
Shounen-manga	DeathNote	4	10.53%	0.00%	-	-	-	34	89.47%	1.10%	38	1.23%	1	6.67%	0.03%	14	93.33%	0.45%	15	0.49%
	Gintama	1	1.45%	0.00%	8	11.59%	0.00%	60	86.96%	1.18%	69	1.36%	29	55.77%	0.57%	23	44.23%	0.45%	52	1.03%
	Detective Conan	5	4.90%	0.00%	-	-	-	97	95.10%	1.05%	102	1.10%	2	3.85%	0.02%	50	96.15%	0.54%	52	0.56%
	Naruto	-	-	-	-	-	-	20	100%	1.00%	20	1.00%	1	11.11%	0.05%	8	88.89%	0.40%	9	0.45%
	OnePiece	6	8.82%	0.00%	-	-	-	62	91.18%	1.37%	68	1.50%	9	22.50%	0.20%	31	77.50%	0.68%	40	0.88%
		Total	16	5.39%	0.00%	8	2.69%	0.00%	273	91.92%	1.14%	297	1.24%	42	25.00%	0.18%	126	75.00%	0.53%	168
Bold are the largest group. %1P=Percent of first person pronouns; %2P=Percent of second person pronouns; %W=Percent of all words																				

Bold are the largest group. %1P=Percent of first person pronouns; %2P=Percent of second person pronouns; %W=Percent of all words

Looking at the distribution of first person pronouns, I found that the central male first person pronoun was *ore* (1,384 instances, in 6.84% of male *Lines*), whereas the central female first person pronoun was *atashi* (545, in 4.59% of female *Lines*). However, female characters in male-oriented series used *watashi* more commonly, with few examples of *atashi* (273 [7.14% of female *Lines*] vs. 16 instances [0.53%]). When female characters in shounen-manga do use *atashi*, they tend to be older, sometimes rather rough and *strong* characters (Figure 56 compared with Figure 57). Male characters show more variety in first person pronouns than females overall: Female characters only used 3 forms, *atashi*, *uchi*, *watashi*, and *ore*, whereas male characters used all of the first person pronouns with the exceptions of *uchi* and *watakushi*. However, not all of them are forms commonly used in real-life: *washi* (78), *yo* (11) and *ware* (8) are all either dialectical or no longer in current use.

There were also no examples of the feminine singular form *uchi* and the masculine *jibun*, which Hishikari (2007) reports as recently being in common use. (*Uchi* used as a feminine plural with the ending *ra* [*uchi-ra*] or as the neutral plural possessive [*uchi-no*] were seen infrequently.)

**Figure 56 (Above):** Usage of *atashi* by a female character in *One Piece* (1:69)

**Figure 57 (Below):** Usage of *watashi* by a female character in *One Piece* (1:207)



*While the female character using atashi is older and unattractive, the female character using watashi is younger and attractive. This type of differential usage was not uncommon in the shounen-manga .*

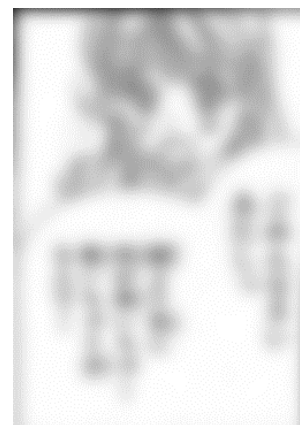
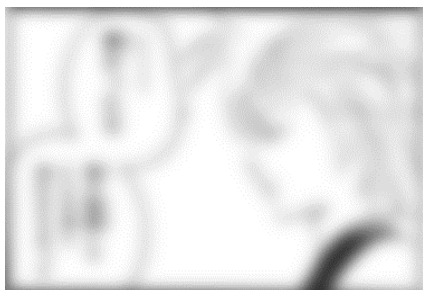
There were also differences between series. *Death Note* uses *ore* comparatively infrequently (59) in favor of *boku* and *watashi* (*boku*: 120; *watashi*: 124). *Death Note* features many public scenes because of its peculiar plot line, which may have influenced characters towards the more formal *boku* and *watashi*. As Kobayashi's (1999) study has shown, men tend to avoid use of *ore* in the workplace when dealing with outsiders or superiors, and this is likely a similar change. Female characters in *Nodame Cantabile* also used first person pronouns infrequently. They may be using alternatives forms like their first names; when they do use first person pronouns, however, they favored *watashi* (51), and did not use the female form *atashi* even once. *Nodame*'s main female characters are all college students, and this may be a reflection of that: Comparing high school students and college students, Hishikari (2007) found that the usage of *watashi* went up as women increased in age, so that while *atashi* was still the most common form used amongst friends (47% vs. 44%) and parents (50% vs. 17%), it was fell behind *watashi* with under-classmates (55%) and teachers (93%~96%). Male characters in *Kimi ni Todoke* also have a very low percentage of first person pronouns, possibly because it features few male *Lines*, giving perhaps insufficient environments for their appearance.

**Table 32:** Types of personal pronouns used by male characters

Genre	Series	1st Person Pronons										2nd Person Pronouns									
		Feminine			Masculine			Neutral			Total		Masculine			Neutral			Total		
		#	%1P	%W	#	%1P	%W	#	%1P	%W	#	%W	#	%2P	%W	#	%2P	%W	#	%W	
Shounen-manga	Bokura	-	-	-	95	98.96%	1.31%	1	1.04%	0.01%	96	1.32%	70	97.22%	0.97%	2	2.78%	0.03%	72	0.99%	
	Kimi ni Todoke	-	-	-	79	100%	1.39%	-	-	-	79	1.39%	41	100%	0.72%	-	-	-	41	0.72%	
	Nana	-	-	-	140	97.90%	1.35%	3	2.10%	0.03%	143	1.37%	119	99.17%	1.14%	1	0.83%	0.01%	120	1.15%	
	Nodame Cantabile	-	-	-	131	71.58%	1.05%	52	28.42%	0.42%	183	1.46%	86	83.50%	0.69%	17	16.50%	0.14%	103	0.82%	
	RabunKon	5	3.18%	0.05%	139	88.54%	1.26%	13	8.28%	0.12%	157	1.43%	89	100%	0.81%	-	-	-	89	0.81%	
	Total	5	0.76%	0.01%	584	88.75%	1.25%	69	10.49%	0.15%	658	1.40%	405	95.29%	0.86%	20	4.71%	0.04%	425	0.91%	
Shounen-manga	DeathNote	-	-	-	179	59.08%	0.56%	124	40.92%	0.39%	303	0.95%	58	56.31%	0.18%	45	43.69%	0.14%	103	0.32%	
	Gintama	-	-	-	256	98.08%	1.08%	5	1.92%	0.02%	261	1.10%	280	86.15%	1.18%	45	13.85%	0.19%	325	1.37%	
	Detective Conan	-	-	-	132	68.75%	0.49%	60	31.25%	0.22%	192	0.72%	97	51.32%	0.36%	92	48.68%	0.34%	189	0.71%	
	Naruto	-	-	-	207	96.73%	1.13%	7	3.27%	0.04%	214	1.16%	217	93.53%	1.18%	15	6.47%	0.08%	232	1.26%	
	OnePiece	-	-	-	428	95.96%	1.64%	18	4.04%	0.07%	446	1.71%	267	92.39%	1.02%	22	7.61%	0.08%	289	1.11%	
	Total	-	-	-	1,202	84.89%	0.95%	214	15.11%	0.17%	1,416	1.12%	919	80.76%	0.72%	219	19.24%	0.17%	1,138	0.90%	
Bold are the largest group. %1P=Percent of first person pronouns; %2P=Percent of second person pronouns; %W=Percent of all words																					

**Bold** are the largest group. %1P=Percent of first person pronouns; %2P=Percent of second person pronouns; %W=Percent of all words

As opposed to the way that manga are commonly perceived, there were very few examples of female speakers using masculine first person pronouns. Indeed, the only example seen was the semi-recurrent character Tatsumi from *GinTama*, who appears as a repeat character in an early story arc (Figure 58). (While she reappears in a later chapter, she does not have a constant role.) No other female characters, from either genre, used any masculine first person pronouns to refer to themselves; however, there are cases where they used masculine forms in quotative environments, where they are borrowing someone else's speech patterns. (These examples were not included as targets.) Interestingly, there was one example of a male character who used *atashi*: Seiko, a female-identifying transsexual in *RabuKon* (Figure 59). In both of these cases, the characters *appear* similar to the opposite gender: Tatsumi and Seiko not only talk like members of the opposite gender, they also dress and (stereotypically) behave like them, too.

**Figure 58 (Left):** Usage of *ore* by a female character in *GinTama* (3:160)**Figure 59 (Right):** Usage of *atashi* by a male character in *RabuKon* (3:87)

With regard to the distribution of second person pronouns, it appeared that the main male second person pronoun was *omae* (919 [4.3% of male *Lines*]). The main female second person pronoun was *anta* (329 [0.70% of female *Lines*]), with *anata* second (70 [0.5%]). As with first person pronouns, male characters tended to use a wider variety of second person pronouns, although this difference was somewhat less dramatic. There were also some differences between series. Male characters in *Detective Conan* tended to favor *anata* over *omae* (78 vs. 47); and *anata* is also much more common in *Death Note* (41 instances, compared with *omae*'s 48). This may also have to do with its frequent public scenes, as *omae* is extremely familiar. There were also some examples of female characters using masculine second person pronouns. Some characters used *omae*, although they tended to be older women—or, in the case of *Death Note*, be non-human. There were also some examples of female characters using the highly masculine *temee*, but these were almost all attributable to the punk character Nana in the series of the same title. However, there appears to be an asymmetrical usage here within gendered first and second person pronouns: While Nana refers to others commonly in the familiar, masculine and rude *temee*, she also uses the feminine *atashi* to refer to herself. Compared with first person pronouns, the choice of second person pronouns is less related to individuals' identity than to their relationship with their listeners, and these kinds of asymmetrical usages may not be surprising as using masculine first person pronouns has less personal significance for their speakers.



## 4.4 Study 3: Distribution of sentence final expressions in Lines

### 4.4.1 Methodology

This study looked at the use of SFPs and other sentence final expressions (SFEs) by male and female characters. Sentence final particles were targeted using the category of *Particles-Final Particles* (*joshi-shuu*/*joshi*) from the morpheme analysis. As with personal pronouns, forms in clearly quotative environments were exempted. When necessary, orthographic symbols and clues, such as periods, exclamation and question marks, white space and other items common at the end of sentences in manga were used to define sentences. I also controlled for orthographic differences, not distinguishing between the syllabaries (katakana, hiragana) used to write SFPs, nor for whether the vowels were being lengthened (*ne* vs. *nee*, *sa* vs. *saa*).

All SFPs were tagged for gender, but before introducing the SFPs examined I must note that there is some disagreement on which items qualify as SFPs and how they should be categorized, gender-wise. As Inoue (2006, p. 54) describes, their determination to a large degree is “political”, and the situation is not as objective as one might think. This must be kept in mind when considering the data described here, and the observations I make are naturally relative to how the SFPs have been categorized. That being said, I based the groupings found here primarily upon McGloin (2005), Okamoto (1995) and Ueno (2006), with some modifications (Table 33). I also included Kansai dialect SFPs as *RabuKon* is set in Osaka; for these, I followed Sturtz Sreetharan (2004).

It should be noted that some of the elements included, in particular, *da*, *n+da*, *desho(u)* and *daro(u)*, may not be strictly considered SFPs (although, as Maynard (2008) notes, *da* has pragmatic functions that allow one to believe it is more than just a copula). This is primarily because they are associated with the larger question of sentence endings and gender. Giving an accurate assessment of how many SFPs are used also requires knowing where they are *not* used (but possible), such as the endings of verbs or adjectives. There are also environments where the *lack* of non-SFP endings determines whether a sentence is neutral, masculine or feminine. For example, the informal copula *da* is generally treated as moderately masculine, and the bare noun treated as strongly feminine; thus, saying *hon da yo*—“this is a book”, with the copula *da* and the SFP *yo*—is moderately masculine, whereas *hon yo* is strongly feminine. As a result of these factors, I also chose to include such non-SFPs sentence final null-forms in the survey and, I have followed Izumitani (1996) in treating them as SFPs.

**Table 33:** Types of sentence final expressions by gender

Strongly Feminine (SF)	Moderately Feminine (MF)	Neutral (N)	Moderately Masculine (MM)	Strongly Masculine (SM)
A/V/+no+yo/ne/yo-ne	A/V-no	(nen)de K	(da/ka/wa)i	(yan)ke K
kashira	deshou	A/V plain form	(da/ka/wa/yo)nou	/ai/-[ei]
N/masu-no±yo/ne/yo-ne	mon(o)(ne)(yo)	A/V/masu(nen/yan/yo)na K	ya	ka-yo
N-yo(ne)	N-ne	A/V/masu/other+ne	na	V-com.(yo)(na)
wa ↑(±ne/yo/yo-ne)	yon	aru(ka/yo) F	A/V/da/other+yo	V-na(yo)(neg. com.)
		datte	be	wa ↓ na
		jan	da(rou/tta)(yo/ne/yo-ne)	ze
		ka-na/nya:	ga-ne	zo
		kke(na/ne)	ka	
		masu form	ka-ne	
		masu-ka/yo	sa	
		nen K		
		noni		
		Q-no		
		raa		
		tteba		
		V-te (req.)		
		wa ↓ na K		
		ya K		
		yan K		

N=Noun, A=Adjective, V=Verb, K=Kansai form, com.=command form, neg.=negative form, req.=request, ↓=falling intonation, ↑=rising intonation

There are also limitations when differentiating between SFPs of similar form as text-data. It is nearly impossible to tell differences in intonation, which is an issue with the SFPs *no* and *wa*. *No* generally forms neutral questions when its intonation is rising, but forms either strongly or mildly feminine statements when used with a falling intonation and follows adjectives and verbs or nouns, respectively. Here, any *no* proceeding a question mark was determined neutral, and the rest sorted manually. *Wa* is also found in two forms of rising and falling intonation. With rising intonation, *wa* is considered a feminine SFP, and may be seen in combination with *ne*, *yo* and *yo-ne*. With falling intonation, however, *wa* is considered masculine, and does not take *ne*, *yo* and *yo-ne*, although it may be observed with *na*. As *wa* used with the rising intonation is generally considered to be a strongly-feminine form, I assumed that male characters were using the strongly-masculine falling intonation, unless in combination with the other SFPs, or when used with specific characters known to use female forms exceptionally, such as *Nodame*'s Masumi, an *okama* (effeminate, likely gay or transgender male) character who actively uses other parts of feminine speech like *kashira* and *no-yo*. This is an admittedly problematic solution, and future studies may require more subtle differentiation using other orthographic clues or comparisons with their animated versions.

SFPs are also sometimes used in non-sentence final environments. SFPs or related forms can often be used following dependent clauses and individual lexical items. For example, in the sentence *watashi wa ne, hon ga suki na-no* (=“I ‘ne’, like books”), *ne* does not occur at the end of the sentence, but rather after the subject (*watashi*, “I”) and the subject-particle (*wa*), and draws attention to the subject. Morita (2007) notes that there are some similarities between these mid-sentence particles—described as *kantoujoshi*—and SFPs, but also that their roles have yet to be fully understood. For the purposes of this study, *all* SFPs were included, meaning that

kantoujoshi were also treated as SFPs. This was chosen partially due to the difficulty of defining absolutely what a sentence *is*, as was discussed in Chapter 3, but also in acknowledgement of the uncertainty of the differences between SFPs and kantoujoshi and the need to establish those facts before they are treated as distinct entities.

Note also that many of the SFPs examined here can also be seen in combination with each other. Well-established combinations of SFPs with similar levels were grouped together, such as the feminine form of *wa* with *yo-ne*, *yo*, or *ne*; in order to avoid over-counting, for other types whose combination could not be considered an established variation, the string was categorized by the strongest form found within. One exception was made in the use the masculine reinterpretation of the diphthong /ai/ as [e:], described previously in Chapter 3. Because it is a strongly masculine form which is not a true SFP and is primarily related to phonetic change, it was always treated separately from any SFPs following it.

**Table 34:** Types of sentence final expressions seen (SJ=Shoujo-manga, SN=Shounen-manga)

Sex	Genre	Strongly feminine				Mildly feminine				Mildly masculine				Strongly masculine																	
		A/V/N/ masu-no yo/ ne/ yo-ne	kashira	N/ masu- no	N- yo( ne)	wa ↓ ( ne/ yo/ yo-ne)	A/V- no	deshou	mon( o)( ne)( yo)	N- ne	yon	( da/ ka/ wa) !	( da/ ka/ wa/ yo) nou	ya	na	A/V/other yo	be	da( rou/ tta)( yo/ ne/ yo-ne)	ga- ne	ka	ka- ne	sa	( yan) ke( na) k	/ai/ - [ei]	ka- yo	V- Com.(yo)( na)	V- na( yo) (Neg. Com.)	wa ↑( na)	ze	zo	
Male	SI	○	○	○	23	○	○	○	53	○	○	○	○	○	○	○	○	840	○	○	○	○	○	○	166	○	○	○	○	○	○
		SN			16	○	○	61	○	○			○	○	○	○		2,461	○	○	○	○	○	○	○	708	○	○	○	○	○
Female	SI	○	○	○	○	121	144	○	○	○	○	○	○	○	○	○	○	766	○	○	○	○	○	○	○	○	○	○	○	○	○
		SN	○	○	○	146	57	○	○	○	○	○	○	○	○	○	225	○	○	○	○	○	○	43	○	○	○	○	○	○	○
Total	Total	146	21	49	226	283	301	227	190	122	1	77	20	57	1,143	1,410	6	4,276	8	1,376	68	351	23	886	68	965	288	87	227	453	

Sex	Genre	Neutral																			
	Jan	( nen) de k	A/V plain form	A/V/ masu( nen/ yan/ yo) na k	A/V/ masu/ other ( datte/ yo) ne	aru( ka/ yo) F	datte	ka- na/ nya:	kke( na/ ne)	masu form	masu- ka/ yo	nen k	noni( ne)	Q- no	raa	tteba	V- te (Req.)	wa ↓ na k	ya k	yan k	
Male	SI	○	○	1,485	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	
	SN	○	○	3,086		○		○	○	○	○	○	○	○	○	○	○			○	
Female	SI	○	○	1,865	○	○	○	○	○	○	○	○	○	○	○	○	○		○	○	
	SN	○	○	459		○	○	○	○	○	○	○	○	○	○	○	○			○	
Total	Total	242	131	6,895	264	981	44	99	199	53	1,691	629	214	152	530	2	39	133	1	319	52

#### 4.4.2 Results

**Forms found:** As shown in Table 34, a large variety of SFPs of all gendered-types were seen. In general, male and female characters used a combination of different gendered SFPs, and not necessarily restricted to their gender: Both sexes used all levels of gendered SFPs, from the most strongly feminine (SF) forms to the most strongly masculine (SM) forms, as well as neutral (N), mildly feminine (MF) and mildly masculine forms (MM). In fact, almost all SFPs were seen used by both sexes to some degree, although not in all series. The major exception was the Kansai forms, which were seen almost exclusively in *RabuKon*. Below, I will discuss the specific distribution of SFPs by the sex of characters within genres. Note that as I looked at all of the SFPs by gender and genre exclusively, there may in fact be cases where similar types of SFPs were being used to create very different characters even within the same gender and genre, such as the use of feminine forms to create childish or gentle female characters, but also snobbish ones (e.g., the older sister in Figure 64).

**Female characters:** Female characters in both genres used primarily neutral and moderately masculine forms. These two types covered an average of upwards of 87% of all SFPs in shoujo-manga, with the remaining made up of, in order, moderately feminine forms, strongly feminine forms and strongly masculine forms (Table 35; Figure 60~Figure 64). In shounen-manga, however, this percentage was lower (69.15%), with female characters using a lower percentage of neutral forms, suggesting that their language is more marked for gender than in shoujo-manga. The high-usage of neutral forms is not surprising for either genre, given that many more neutral forms were targeted than the other types and that the neutral forms also included forms such as the unmarked plain forms and the *masu* form verbs and adjectives as well as the copula *desu*. These could be said to be default null forms, in the sense that if an SFP is not used, these would be the standard way to end a sentence. As SFPs are non-obligatory, a high percentage of unmarked defaults is to be expected.

**Table 35:** Sentence final expressions used by female characters

Genre	Series	SF		MF		N		MM		SM		Total	% words
Shoujo-manga	<i>Bokura</i>	15	1.41%	92	8.65%	643	<b>60.43%</b>	292	<u>27.44%</u>	22	2.07%	1,064	10.99%
	<i>Kimi ni Todoke</i>	51	3.65%	101	7.22%	751	<b>53.72%</b>	452	<u>32.33%</u>	43	3.08%	1,398	9.76%
	<i>Nana</i>	58	3.04%	108	5.67%	993	<b>52.13%</b>	680	<u>35.70%</u>	66	3.46%	1,905	10.96%
	<i>Nodame Cantabile</i>	35	4.93%	36	5.07%	511	<b>71.97%</b>	122	<u>17.18%</u>	6	0.85%	710	10.29%
	<i>RabuKon</i>	105	7.37%	58	4.07%	1102	<b>77.39%</b>	122	<u>8.57%</u>	37	2.60%	1,424	11.26%
	Total	264	4.06%	395	6.08%	4,000	<b>61.53%</b>	1,668	<u>25.66%</u>	174	2.68%	6,501	10.67%
Shounen-manga	<i>DeathNote</i>	8	3.25%	13	5.28%	195	<b>79.27%</b>	29	<u>11.79%</u>	1	0.41%	246	7.97%
	<i>GinTama</i>	35	6.24%	29	5.17%	234	<b>41.71%</b>	186	<u>33.16%</u>	77	13.73%	561	11.08%
	<i>Detective Conan</i>	186	26.61%	65	9.30%	312	<b>44.64%</b>	130	<u>18.60%</u>	6	0.86%	699	7.54%
	<i>Naruto</i>	43	24.71%	14	8.05%	85	<b>48.85%</b>	30	<u>17.24%</u>	2	1.15%	174	8.70%
	<i>OnePiece</i>	119	26.44%	45	10.00%	186	<b>41.33%</b>	86	<u>19.11%</u>	14	3.11%	450	9.94%
	Total	391	18.36%	166	7.79%	1,012	<b>47.51%</b>	461	<u>21.64%</u>	100	4.69%	2,130	8.90%

**Bold** is the largest group; underscore the second largest group; and *italics* the third.

As regards the high percentage of moderate masculine forms, it must be remembered that they include the copula *da* and its derivatives (*darou*, *da-ne*) and *yo* following adjectives and verbs in the plain form (e.g., *suru yo*, *kawaii yo*); this was similar to Okamoto's (2005) reports for natural speech. *Da* forms made up 11.46% of usages on average, or 11.77% in shoujo-manga and 11.11% in shounen-manga. *Yo* after adjectives, verbs and other related forms were 7.12% and 6.81% of all forms in shoujo-manga and shounen-manga, respectively, and there were *no* significant differences between female and male usages overall, suggesting that the distributions are similar for male and female characters. While both of these forms were categorized as moderately masculine, their usage appears to have become largely neutral, leading some to call for their categorical reinterpretation (Nishinuma, Hayashi and Yabe, 2006). However, *da* forms were still more common in male speech, averaging 3.67% of all words compared to 0.52%; this difference was statistically significant ( $t(18) = 3.00$ ,  $p = .008^*$ ). It may be reasonable to suggest that it is still in the midst of change, hence the imbalance of frequencies. If it were truly a *masculine* form, it would be surprising to see it being so frequent in female speech, but if it were still in the midst of change, one would expect both that female characters would use them but that that rate would be lower than male characters, as it is here.

Things were more complicated, however, with the remaining three categories. While both genres used strongly masculine forms least—there were no statistical differences between the genres, and with the exception of *GinTama*, they both used them fairly infrequently—shoujo-manga used moderately feminine forms more frequently than strongly feminine forms, whereas shounen-manga used strongly feminine forms more frequently than moderately feminine forms. The difference between shoujo-manga and shounen-manga as regards strongly feminine forms was statistically significant ( $t(8) = 2.52$ ,  $p = .03603^*$ ). Amongst the shoujo-manga series, *RabuKon* seemed to be somewhat higher (105 instances [7.37% of female SFPs]); however, this is likely because its use of the Kansai dialect. It appears that *wa* was used more commonly than other series, but in the Kansai

dialect it may not be appropriate to treat *wa* as a strongly feminine form; some other forms such as *wa-na* are in fact treated as neutral by some (e.g., Sturtz Sreetharan (2004)). If *wa* was reinterpreted in *RabuKon* using this conceptualization, the percentage of SF forms would be even lower, making the difference between shoujo-manga and shounen-manga even clearer.

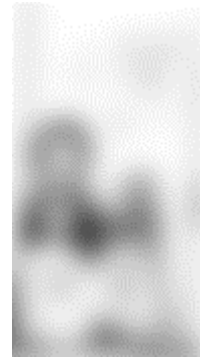
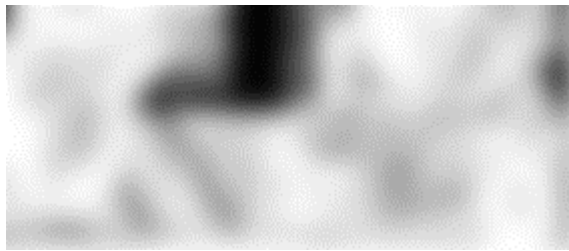
As this suggests, there are also important differences between series. Within shoujo-manga, *Nodame Cantabile* and *RabuKon* were both somewhat exceptional in their relatively low usage of moderately masculine forms (17.18% and 8.57%, respectively) and comparatively high usage of neutral forms (71.97% and 77.39%, respectively). While the peculiarities of *RabuKon* can mostly be explained by the use of the Kansai dialect, the characteristics of *Nodame* might be thought to be related to (1) the limited number of female characters and (2) the settings of the story. Female characters in *Nodame* actually account for fewer *Lines* than male characters; in fact, with the exception of Nodame herself, there really are no major female characters. For all that Nodame herself has often been seen as having strange ways of talking—look at sites such as the *Nodame Weird Cries Dictionary* (Nodame kisei dikushonari, 2004)—she actually accounts for a small percentage of *Lines* (23.84%) compared with the main female characters in the other shoujo-manga. This may affect speech patterns, in terms of giving characters the opportunities to use non-obligatory gendered SPF forms. As I also argued for personal pronouns, because *Nodame* is set in a college environment, and the characters somewhat older than in the other series, one might also expect somewhat more neutral styles befitting less egalitarian relationships. Being set in college, the characters are meeting less informally and less equally than they would have been in college, and just as personal pronouns appear to become somewhat less gender marked as women enter college (Hishikari, 2007), we might expect that marked SFPs also become somewhat less favored. Note that while *Nana* also features a somewhat older cast of characters, they are also much more familiar to each other, being either friends from childhood, romantic partners or roommates.

Two shounen-manga series are also somewhat exceptional. First, *Death Note* patterned similarly to *Nodame Cantabile*, with comparatively less feminine forms and mildly masculine forms, with close to 80% of all SFPs neutral. Of all the shounen-manga series, *Death Note* featured the lowest percentage of female *Lines*, and it also featured many public scenes; it might be that these factors affected the results as I argued that they did for *Nodame*. Second, *GinTama* featured more strongly-masculine forms (13.73%) than any other series. Looking at the data by individual characters, *GinTama* features three characters who use the majority of the 77 strongly masculine forms: Otose (13), Kagura (18), and Tatsumi (40). As noted in the pronouns discussion, Tatsumi is the main character in one story-arc; Otose and Kagura, however, are both regular recurring characters. All three have in common that they have very distinct personalities. Where Tatsumi is a wanna-be firefighter, Otose is a strong, older woman from the *shita-machi* old Edo town and Kagura, who looks like a cute girl, is

really an alien assassin. Their use of masculine forms might be part of their characterization as non-stereotypical women; in particular, Kagura also uses other forms (*aru*) that mark her as a foreigner, suggesting that these might be part of a larger process of defining her character.

**Figure 60 (Right):** Example of a female character using a mildly masculine SFE (*n-da-yo-ne*) in *Kimi ni Todoke* (1:134)

**Figure 61 (Left):** Example of female characters using strongly masculine SFEs (*ze*) in *Nana* (2:135)



**Figure 62:** Example of a female character using a neutral SFE (*ka-na*) in *Naruto* (2:177)



**Figure 63 (Left):** Example of a female character using a mildly feminine SFE (*mon*) in *Rabukon* (1:92)

**Figure 64 (Right):** Example of female characters using a strongly feminine SFE (*wa*) in *Detective Conan* (3:46)





**Male characters:** Like female characters, male characters in both genres used primarily neutral forms (shoujo-manga: 50.98%; shounen-manga: 40.89%), followed by numerically moderately masculine forms. In fact, both genres had exactly the same order for all types, and fit the gender lines closely, following by strongly masculine forms, moderately feminine forms and strongly feminine forms (Table 36); Figure 65~Figure 69). However, it would be incorrect to assume that this means that shoujo-manga and shounen-manga were identical. Shoujo-manga seemed to use more neutral forms and fewer moderately or strongly masculine forms; however, these were not statistically significant, suggesting that they are primarily differences between individual series.

**Table 36:** Sentence final expressions used by male characters

Genre	Series	SF	MF	N	MM	SM	Total	% words					
Shoujo-manga	<i>Bokura</i>	1	0.11%	24	2.71%	419	<b>47.29%</b>	317	<u>35.78%</u>	125	14.11%	886	12.23%
	<i>Kimi ni Todoke</i>	2	0.28%	20	2.76%	295	<u>40.69%</u>	316	<b>43.59%</b>	92	12.69%	725	12.74%
	<i>Nana</i>	5	0.43%	21	1.79%	481	<u>41.11%</u>	566	<b>48.38%</b>	97	8.29%	1,170	11.25%
	<i>Nodame Cantabile</i>	31	2.20%	27	1.91%	656	<b>46.52%</b>	546	<u>38.72%</u>	150	10.64%	1,410	11.27%
	<i>Rabukon</i>	6	0.47%	56	4.40%	935	<b>73.39%</b>	99	<u>7.77%</u>	178	13.97%	1,274	11.59%
	Total	45	0.82%	148	2.71%	2,786	<b>50.98%</b>	1,844	<u>33.74%</u>	642	11.75%	5,465	11.67%
Shounen-manga	<i>DeathNote</i>	1	0.04%	15	0.62%	1510	<b>62.47%</b>	742	<u>30.70%</u>	149	6.16%	2,417	7.55%
	<i>GinTama</i>	17	0.58%	51	1.75%	869	<u>29.83%</u>	1258	<b>43.19%</b>	718	24.65%	2,913	12.32%
	<i>Detective Conan</i>	2	0.10%	35	1.78%	898	<b>45.68%</b>	782	<u>39.78%</u>	249	12.67%	1,966	7.34%
	<i>Naruto</i>	3	0.18%	20	1.23%	692	<b>42.61%</b>	706	<u>43.47%</u>	203	12.50%	1,624	8.83%
	<i>OnePiece</i>	1	0.03%	11	0.37%	902	<u>30.14%</u>	1317	<b>44.00%</b>	762	25.46%	2,993	11.49%
	Total	24	0.20%	132	1.11%	4,871	<b>40.89%</b>	4,805	<u>40.33%</u>	2,081	17.47%	11,913	9.39%
Bold is the largest group; <u>underscore</u> the second largest group; and <i>italics</i> the third.													

**Bold** is the largest group; underscore the second largest group; and *italics* the third.

There are two interesting points to mention in regards to the use of feminine forms. First, while strongly feminine SFPs were very uncommon with male characters in shounen-manga, coming in at 0.20% of all the SFPs seen, they were somewhat more common in shoujo-manga (0.82%). However, when they were used they were mostly okama characters using *kashira*, *wa-yo*, *no-yo*, etc. (Figure 69), a patterning similar to that found by Suzuki (2001), who reported that while all SFPs were used by both sexes to some degree in television dramas, male characters only used strongly feminine forms if they were “gay” characters. In this sense, the use of strongly feminine forms by male characters is almost exclusively for the creation of stereotyped individuals, and can be thought of being yakuwari-go-like.

Secondly, moderately feminine SFPs were more frequent in both genres, but were more frequent in shoujo-manga, at 2.71% of all SFPs vs. 1.11% in shounen-manga. This difference was statistically significant ( $t(8) = 4.65, p = .0017^*$ ). Just as female characters in shoujo-manga appear through their use of fewer strongly feminine SFPs to be somewhat less traditionally *feminine*, male characters in shoujo-manga may appear to be somewhat *more* feminine than their counterparts in shounen-manga. However, this may also be another sign of linguistic change. While more attention has been given to the neutralization of women’s speech, there have also

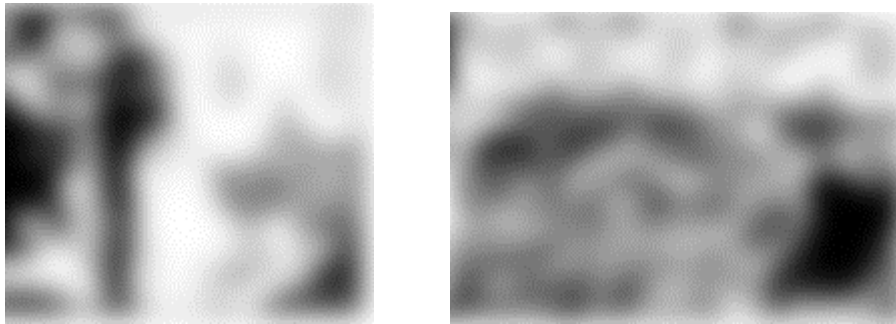
been reports that men's speech is more diverse than has been thought. Ohara (2004) reports, for example, that men sometimes utilize high pitches during social interactions, which has normally been thought to be primarily feminine. The use of moderately feminine forms by male characters in shoujo-manga may suggest that they are becoming less exclusively feminine, just as the moderately masculine forms were shown to be, and are potentially a sign of linguistic realism.

Finally, strongly masculine SFPs were commonly seen in both genres, coming to 11.75% in shoujo-manga and 17.47% in shounen-manga. As noted above, this brought SM SFPs to the third most common type of SFP; however, this usage appears to be much higher than what would be anticipated in natural speech. Like what I have observed in the corpus data, Sturtz Sreetharan (2004) found that neutral forms were by far the most common type of SFP used; however, she also found that both strongly masculine and strongly feminine forms were rare, with strongly masculine forms appearing to make up just 1% of all SFPs. However, there are two caveats to be made: First, Sturtz Sreetharan focused on speakers of the Kansai dialect, which is said to be less gendered to begin with, and second, she also found that younger speakers used SM SFPs much more frequently. Differences in dialect and age—most of the characters in manga being comparatively young—may account for some of these differences. Nonetheless, this high percentage seems to be a departure from naturalistic speech, and appears to be more prominent in shounen-manga than in shoujo-manga.

That said, in comparison with female characters, there were fewer differences between individual series. As with female characters, both *RabuKon* and *Death Note* favored more neutral forms than other series (73.39% and 62.47%, respectively). As regards *RabuKon*, one can again predict this from its use of the Kansai dialect. However, the reasons for *Death Note*'s consistent preference towards neutral forms for both sexes is less clear. As I pointed out in the discussion of personal pronouns, *Death Note* has comparatively more *public* scenes than the other series in the genre, meaning that the characters interact with a variety of different people and often in a semi-official status, such as in the capacity of being a head detective, or being the head detective's son, as is the case with Light Yagami, the main character. Since gendered speech is more associated with informal speech, with formality particularly decreasing the markedness of men's speech as masculine—according to Ide (1990), women's speech starts off more formal than men's does—the abundance of neutral forms may be a reflection of certain contextual issues. Finally, both *Nana* and *Kimi ni Todoke* favored moderately masculine forms over neutral forms (48.38% and 41.11%, 43.59% and 40.69%, respectively); while there are no clear reasons for this, one might suggest something of the opposite for these two series as for with *Death Note*; that is, that male characters tend to be more *masculine* in their speech patterns because the interaction between strangers and non-

strangers is relatively low. This does not account for the comparatively low percentage of masculine forms in *Bokura ga Ita*, however, suggesting that other issues are also at play.

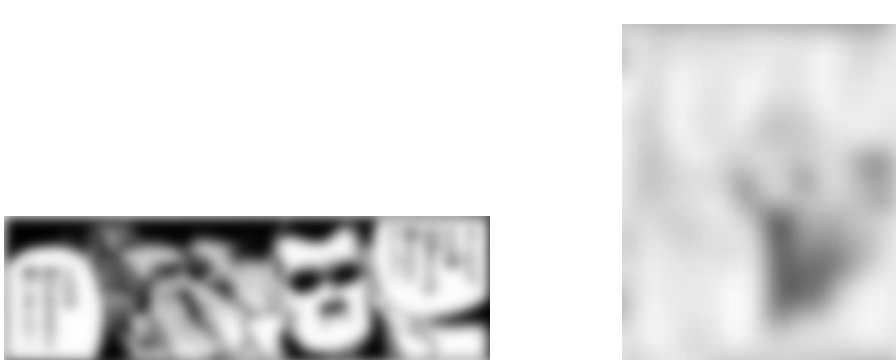
**Figure 65 (Left):** Example of a male character using a moderately masculine SFE (*yo*) in *Death Note* (3:62)  
**Figure 66 (Right):** Example of male characters using strongly masculine SFEs (*zo*; *nai*→*nee*; the command form) in *One Piece* (1:10)



**Figure 67:** Example of a male character using a neutral SFE (*jan*) in *Kimi ni Todoke* (2:14)



**Figure 68:** Example of a male character using a mildly feminine SFE (*desho*) in *Detective Conan* (2:91)  
**Figure 69:** Example of a male character using a strongly feminine SFE (*kashira*) in *Nodame Cantabile* (2:27)

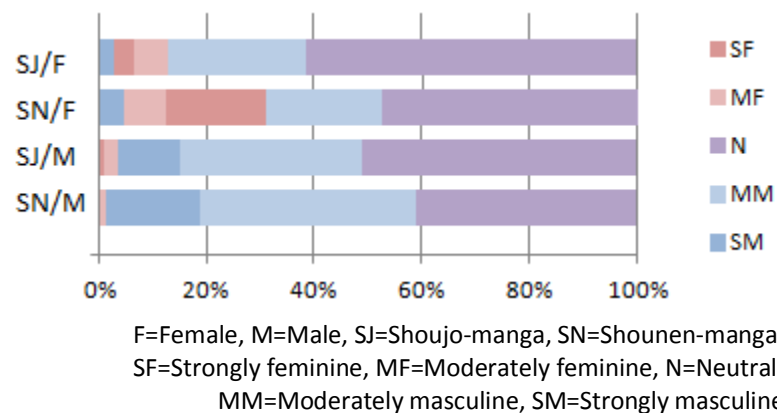


## 4.5 Discussion

### 4.5.1 Differences in characters by genres

Reviewing the distribution found above, there was generally little out-of-sex usage of personal pronouns, for either genre or for either person. In comparison, SFPs showed a more interesting distribution overall. All characters in all genres used primarily neutral forms, followed by moderately masculine forms. As I noted in the results for female characters, many of the moderately masculine forms may be better considered neutral forms now, so this may not be surprising. Male characters also followed the same pattern in both genres, using, in order of most to least, strongly masculine forms, moderately feminine forms and strongly feminine forms. However, while the patterns seem superficially similar, the real distributions are not as balanced. While male characters all used the same types of SFPs in the same orders, they used strongly masculine forms much more commonly in shounen-manga (Graph 10). Furthermore, while female characters in shounen-manga also followed the same pattern for neutral and moderately masculine forms, strongly feminine forms accounted for a significantly higher percentage than in shoujo-manga, so that the order changes from N-MM-MF-SF-SM to N-MM-SF-MF-SM.

**Graph 10:** Ordered distribution of most common SFPs by gender and genre



Thus, we see that the strongly masculine and strongly feminine forms account for a larger percentage of male and female speech, respectively, in shounen-manga than they do in shoujo-manga, making shounen-manga more marked for gender. The differences in the speech patterns of female characters are particularly stark given their distance from real life speech: Compare the data found in Table 35 with Okamoto's (1995) and

Philip's (2001) data in Table 37, which clearly show that the usage of SF forms drop considerably in concert with the age of speakers, but SM and MM forms increase amongst younger speakers. While shounen-manga are nowhere near as conservative as textbooks, where female and male characters fail to move out of expected gender roles at all (Kawasaki & McDougall, 2003), the lower percentage of SM SFPs by male characters in shoujo-manga seems more consistent with their low occurrence rates that Sturtz Sreetharan (2004) reports.

**Table 37:** SFP distributions in other studies

Study		Philips (2001)		Okamoto (1995)			Kawasaki & McDougall (2003)	
Data type		Natural speech		Natural speech			Textbooks	
Sex		Women		Women			Women	Men
Age		Younger	Older	18-23	27-34	45-57	-	-
# speakers		6	6	7	3	4	-	-
SFP type	SF	1.83	19.33	4.00	12.00	28.00	98.33	0
	MF	11.50	23.50	10.00	13.00	23.00		
	N	73.67	50.83	57.00	62.00	44.00	1.67	83.54
	MM	13.00	6.33	24.00	14.00	6.00	0	16.46
	SM	0.00	0.00	5.00	0.00	0.00		

#### 4.5.2 Potential explanations: Author interference, character roles and personal distance

The data force one to ask why characters in shounen-manga pattern so differently from shoujo-manga. In the discussion below I focus on female characters, as the differences between the genres and real-life speech is most obvious for them; and I look at three possible reasons. One perhaps unsurprising reason is that of author-related issues surrounding shounen-manga as comics by and for boys. I also offer two possibilities which are character and/or narrative driven. The first centers on the differences in roles female characters in the two genres play. Related to this point, the second possibility is that the personal distance between characters may also be responsible for some of these differences. Below, these hypotheses are discussed in greater detail.

**Author-related issues:** The first reason for the differences may be possible interference due to the sex of the authors themselves. In examining the use of feminine SFPs in television dramas, Mizumoto (2006) found that

male screenwriters tended to use more feminine SFPs for its female characters than do female screenwriters. Mizumoto also found that male screenwriters have female characters who are in a wider age group use them than female screenwriters do, thus finding that young female characters in male-written series use them where characters of similar ages in female-written series generally do not. The same point may hold for manga, as men tend to write for men, and vice-versa. This is certainly the case here, as all the shoujo-manga series were written by women, and the shounen-manga by men—with the possible exception of *Death Note*, which is co-authored by a male-artist and a writer who has not made his-or-her identity public. It may be that male writers have female characters use more strongly feminine SFPs because (1) as non-users of feminine speech patterns themselves, male writers may be less sensitive to the current differences in usages among particularly young women's speech, and (2) they may have a more conservative internalized view of women's language.

To some degree, this may be related to the preferences of the speakers themselves; Endo (1997, pp. 176-177), in analyzing the results of a 1995 language survey by the Japanese Agency for Cultural Affairs, found that older speakers and male speakers are more likely to prefer gendered differences in speech. While the male authors are by no means elderly, they are generally older (Table 38) than the characters they write about, who are, like their audience, mostly young men and women; and so Endo's results seem to strike true here, too. However, while it is tempting to say that this is what men view as being *ideal* female speech, it may be hasty to assume that this is necessarily an issue of usage values. Rather, as Suzuki (2001, p. 92) writes, written spoken language is a reflection of "what people have in their heads", and the language patterns seen here may simply be those that male speakers are most aware of. While the authors may be *trying* to write naturally, they are forced to imagine what would be realistic for individuals not only of the opposite sex, but also who are several generations below them in age. They may thus not be as aware of recent linguistic changes, a point which is as pertinent to the female writers as it is to the male writers.

**Table 38:** Age of authors

Genre	Series	Author	Year of Birth
Shoujo-manga	<i>Bokura ga Ita</i>	Obata, Yuki	?
	<i>Kimi ni Todoke</i>	Shiina, Karuho	1975
	<i>Nana</i>	Yazawa, Ai	1967
	<i>Nodame Cantabile</i>	Ninomiya, Tomoko	1969
	<i>Rabukon</i>	Nakahara, Aya	1973
	<i>Death Note</i>	Obata, Takeshi Ohba, Tsugumi	1969 ?
Shounen-manga	<i>GinTama</i>	Sorachi, Hideaki	1979
	<i>Meitantei Konan</i>	Aoyama, Gosho	1963
	<i>Naruto</i>	Kishimoto, Masashi	1974
	<i>One Piece</i>	Oda, Eiichiro	1975

**Differences in roles:** In considering the second reason—the differences in the roles female characters play in the genres—it is necessary to recall the results from the first study on the distribution of *Lines* from Section 4.2 above. With the dramatic skew of *Lines* in shounen-manga, female characters clearly have less page-time, and are less visible than male-characters overall. Given this fact, it seems natural to assume that female characters do not play as great a role in shounen-manga as they do in shoujo-manga, or as male characters do in shoujo-manga, either. Indeed, they are primarily secondary characters, such as people on the streets, intermittent characters, etc., and while three out of four series have female semi-main characters, they and their relationships with the main characters are not as central as other male characters are in shounen-manga, or even as male secondary characters are in shoujo-manga. With female characters less central to the plot than male characters, they may be more susceptible to using stereotypical yakuwari-go speech patterns. As I noted in Section 4.2.3, there is a *cost* involved in using yakuwari-go, which explains why shoujo-manga female characters, who play more central roles in the narrative, use less yakuwari-go like speech. On the other hand, because the mostly secondary female characters in shounen-manga are not subject to the same demands as main characters—such as being able to relate to them—realism may simply not be as important. For these types of characters, filling certain roles may trump realism. As Kinsui (2003) demonstrated in his analysis of yakuwari-go, stereotypically female speech patterns form an easy way to build characters like “the haughty upper-class woman”; and when a fully-fleshed character is not necessary, using yakuwari-go based characters may be more economic for authors. Note that this would also be relevant for male characters in shounen-manga, who also use more strongly gender-marked SFPs: The abundance of male characters, especially extras, means that there is that much a need to create many different characters, leading to the importance of economical speech patterns.

Of course, some of this may also be a result of narrative differences. Where shoujo-manga are said to be highly psychological and focused on characters' inner sides (Natsume (1997), Schodt (1996)), shounen-manga might be said to focus more on action than on individual relationships (Ôtsuka, 1994). While there appears to be good reason to not fully accept this schema for today's shoujo-manga and shounen-manga—as I discussed in Chapter 2, *Thoughts*, which Ôtsuka saw as an important part of the psychology of shoujo-manga, no longer appear to be a point of significant difference—recalling the results from the morpheme analysis in Chapter 3, shoujo-manga appeared to use names and personal pronouns more commonly than shounen-manga, which featured more theme-oriented nouns, leading one to argue that shoujo-manga were more interested in the personal relations between characters than shounen-manga. With that in mind, it may be that narratives in shoujo-manga, being more focused on interpersonal relationships, may not permit characters to use stock speech patterns: Readers must be able to identify with the characters they see, which is associated with the realism they feel for the characters (Keen, 2006), which means that realism expressed here in their speech patterns is also likely crucial. However, given the right conditions, i.e., characters one is not *expected* to relate to—or possibly even actively dislike—even characters in shoujo-manga will utilize such stock speech, as has been described by Takahashi (2009) for the villainess in *Life*.

***Personal distance between characters:*** Related to this second point, it may be that, as secondary characters, female characters in shounen-manga are too distant, relationship-wise, to use masculine forms of speech. As has likely become obvious from the results and their analyses, the use of gendered speech in Japanese is not exclusively about gender. As Sullivan (2006) notes, research on personal pronouns and SFPs in Japanese has started to move towards *indexicality* in recent years, meaning that those elements that have been discussed as being *gendered* should not be thought of as signs of gender exclusively, but rather index gender indirectly amongst factors. Thus, “. . . the meaning of gendered language is not static, but multiple and ultimately dependent upon the context of the speech act (Sullivan, 2006, p. 54).” As a result, the usage of personal pronoun and SFP choice is not just about whether or not the speaker is female or male, but about what larger interactive and/or other processes they may be seeking to mark or establish by means of the indexing they use in their language.

With these points in mind, feminine speech patterns are often associated with politeness, and they tend to use more masculine forms when at home or with close friends, etc. (McGloin (2005)). Since female characters in shounen-manga are usually in more public positions, such as clients (*Detective Conan*) or area villagers and townspeople (*One Piece*, *Naruto*), not only are their relationships to the main characters not important, they are often shallow or non-existent, and there is very little interaction *between* female characters. Thus we may think



that they are not in a position to talk casually with other characters in more relaxed speech patterns. In contrast, female characters in shoujo-manga are usually of the same age and in familiar settings: In the corpus here, two are centered upon relationships at the main characters' high school (*Rabukon*, *KimiTodo*), one on college friends (*Nodame*) and one on roommates, band members and college friends (*Nana*). Thus female characters are often already very familiar, or are in situations where they could be considered equals, and this familiarity may be what permits more casual *rough* speaking. As Nakajima (1997) found with the use of masculine question-forming SFPs, masculine speech amongst young women is more likely to come out amongst other young women of equal status; given this, one can expect more masculine speech in shoujo-manga as such relationships are the norm therein.

#### 4.5.3 Masculine personal pronouns and girls

One question that remains unanswered at this juncture is the perception by ordinary people in daily life of manga's personal pronoun usage. While masculine SFPs were not uncommon, and masculine second person pronouns infrequent but not unseen, first person pronouns used by female characters were extremely rare, suggesting a gap in acceptable gender-jumping similar to what Suzuki (2001) found for television series. While popular belief would have it that manga are why some girls are choosing to use masculine first person pronouns, only the character Tatsumi in *GinTama* used masculine first person pronouns. Her usage was in fact extremely marked: Tatsumi lives in the exclusively male environment of firefighting, having been taken under the wing of the head firefighter when she lost her family in a blaze. For her, the use of the masculine *ore* appears to be a way of making herself male, similar to how she wears men's clothes and has a male hairstyle (Figure 58). Not being accepted as an equal, her only recourse is to show that she is no different from men. Thus, *ore* for Tatsumi is essentially the adoption of a male identity, and is only one part of her larger efforts towards that end.

Yet the usage of masculine first person pronouns amongst girls in real life appears to have only superficially to do with adopting a male identity. Some have reported that the usage of *boku* and *ore* not as a way of *becoming* male, but rather as a way of avoiding what is entailed by the overtly-feminine personal pronouns. The association of *atashi* in particular with femininity and female roles has been said to have been avoided (see particularly Miyazaki (2004) for an interesting case study on junior high school girls' usage of *boku* and *ore*). The choices should perhaps be recognized not necessarily as female/male, but as female/other/male, with the female choices representing the looming responsibilities of adult women: Childcare, housework,

sexuality. In comparison, Tatsumi's usage of *ore* is mostly stereotypical in that it is used primarily to paint her as a masculine character; and far from being a positive image, it is a part of the humor of *GinTama*, where she is presented as an anomaly. In this sense, Tatsumi's usage, rather than a representation of linguistic change, could be considered another type of yakuwari-go.

In the context of my studies' findings, what seems more likely is that manga are what is being influenced by real world changes, and so show a lag in adopting them, as Yamane (1986) found with cute handwriting and manga. Just as I discussed for SFPs above, the age gap between authors and their characters may also be important. As non-users of those patterns, authors may not be in touch with such changes, as Mizumoto (2006) suggested for television dramas. This may also be the reason that the popular *uchi* and *jibun* are not present. While I focused on whether female characters in manga were using masculine forms, with the assumption that if manga were influential because they differed from natural speech, it would be because there were female characters who were speaking out of gender, this may have been a false assumption. Indeed, it may not be female characters who are the issue. While boys tend to read more restrictedly, girls read a variety of works, including boy-oriented series (Allen & Ingulsrud, 2005). Female readers may identify more with male characters, as the psychologist Takashi Tomida suggests in an interview (Shinomoto, 2008), particularly given the low percentage of female *Lines* in boy-oriented series. While such suggestions taken too far step into the realm of folk linguistics, begging one to claim that since *ore* is for men, if a woman uses it, she must wish to be a man, the lack of clear data here does suggest the value of taking a wider view on what may be influential.

## 4.6 Conclusions

In summarizing the results of the three studies described here, there appear to be three main points to bring home: First, shounen-manga feature more characters than shoujo-manga, but are largely skewed towards male characters; second, out-of-sex usage of first person pronouns is rare in either genre for either sex, and is generally used to create stereotyped characters when they do occur, although out-of-sex second person pronouns occur somewhat more frequently; and finally, characters, and particularly female characters, are more strongly marked for gender in shounen-manga than they are in shoujo-manga. Focusing in on the third point in particular, I argued that differences in the representation of female characters could be related to author interference, differences in the roles that female characters play in shoujo-manga and shounen-manga, and the personal distance between characters.

There are of course issues that I was not able to deal with here that would improve future analyses. First, while many of the main characters tend to be of similar ages, there could be generational differences as seen by Ueno (2006). While the data do not currently include information about the age of speakers, this is something that might be added in the future. Second, while most of my focus in regard to the second and third studies was on the distributional differences between genres, there are clearly differences between series, as noted with *RabuKon*. Third, as with the categorization of the semi-ambiguous SFP *wa*, the gender categories that I adopted in the second and third studies are not absolute settings. Reviewing the categorization of the SFPs, and possibly utilizing additional data such as the animated versions of the series could be helpful to achieve a better image of how they are being used here. Finally, since speech patterns are also dependent on speakers' relationships with their listeners, new tags about who is being spoken to could be informative.

Although these issues must be taken up again in a future forum, the differences in how female characters are being depicted in the genres leave us with an interesting question to consider. While the reasons I suggest above for the gap in female speech patterns harken back to the results for the first study on the distribution of *Lines*, one must ask why should there be such a large gap in the first place?

Just as shounen-manga are comics for-and-by-boys, shoujo-manga are also said to be characterized by being “for young-women, by young women” as Fujimoto (2008) describes them, with both their readers and their writers female. If one thought this would mean that women dominate in shoujo-manga, however, one would find oneself proved incorrect. Shoujo-manga are clearly not an exclusively women's world: if anything, as in *Nodame* or *Nana*, where the number of male characters is actually higher than that of female characters, even shoujo-manga are not just about women. This is a major difference, however with for-boys, by-boys, *about-boys* shounen-manga; one reason might be partially due to differences in plot settings.

As came up earlier, three out of five of the shounen-manga examined here are fantasy, with the remaining two in real-life settings with sci-fi elements; all five are, however, all action-oriented. In comparison, all five shoujo-manga are set in the real world (four at school, one at school and work), a fact which may leave no choice but to feature both sexes. That is to say, real-life settings may not allow them to cut out male characters as they are a crucial part of establishing realism; in a fantasy world, one can replay things as one pleases. Yet while many workplaces are gender-stratified, and many schools single-sex, it is hard to truly imagine an ordinary setting where one has absolutely no interaction with men. The focus on interpersonal relationships that appears to be central to manga also puts pressure on the need to depict both sexes equally. Indeed, as Fujimoto (2008, p. 22) describes in detail, shoujo-manga has long centered on the development of the main female character through her relationship with men, such that even if she is “ugly” or “awkward” or “no good”, some man will love her; and while there appear to have been changes towards less clichéd plot lines in recent (see Shamoon (2004) for an overview of this topic), to a large degree, the depiction of relationships is still fairly conservative, with girls often taking on traditional roles associated with motherhood in blossoming shoujo romances (Choo, 2008). In this way, interaction in shoujo-manga by nature requires male characters—which is not the case in more action-oriented shounen-manga. If anything, the trend in shoujo-manga is to also remove female characters, such as in boys-love series.

At the same time, one is forced to wonder what this means for people who read out of gender. While genres have traditionally been determined in manga by market segments defined by the gender of their readerships, reading trends seem to be changing in recent years. In particular, female readers seem to not restrict themselves to shoujo-manga titles, but often read, and even prefer, shounen-manga titles (Allen & Ingulsrud (2005), Ogino (2001)). Given that, as has been seen here, female characters tend to play extremely secondary roles in shounen-manga series, and that they tend to use stereotyped speech patterns different from readers’ likely own usages, it seems important to think about how one might interpret what it is that female readers get out of reading shounen-manga. While this is an issue which cannot be properly dealt with here, it is one which becomes more important when considering the social meaning and influence of the peculiar gendered speech patterns seen in manga.

## 5. Conclusions

### 5.1 Overview

#### 5.1.1 Summary of Chapters 1 to 4

The major goal of this thesis was to expand the general understanding of language in manga, which has hitherto been looked at as primarily a visual media. To do so, I constructed a corpus of popular manga, with an equal distribution of shoujo-manga and shounen-manga. As I outlined in Chapter 1, the corpus was structured so as to include all of the text seen in manga, organizing them into 8 different categories: *Lines*, *Thoughts*, *Narration*, *Onomatopoeia*, *Background Text*, *Background Lines/Thoughts*, *Comments* and *Titles*. As I argued, the unique structure of the corpus thus allows one to utilize it for any numerous types of queries, and is not limited to its ability to contribute to manga research alone.

I then began the data analysis in Chapter 2, where I examined how language in manga is characterized structurally. By looking at the distribution of the categories above, I showed that while *Lines* form the core of text in all of the series included in the corpus, individual series were distinguished by their use of the other, secondary categories. The data also showed that there was no significant differences between shoujo-manga and shounen-manga in regards to the number of *Thoughts* seen, thus not supporting arguments by Natsume (1997) and others that shoujo-manga are more “psychologically-oriented” than shounen-manga, which should result in more monologue-like, thought-like texts. Rather, I agreed with Ôtsuka’s analysis that shoujo-manga may no longer be described so clearly in this way, and instead focused on two categories, *Background Lines/Thoughts* and *Comments*, which proved to be more common in shoujo-manga than in shounen-manga.

I then took a step back to consider why these two text categories, both distinguished by their visual presentation, needed to be differentiated from *Lines*, *Thoughts* and *Narration*, with which they have very similar narrative roles. While suggesting that these texts are more prominent in shoujo-manga because of the emergence of a special authorial presence in shoujo-manga during the late 1960s and 1970s, I argued that these texts change the reading experience by visually encoding data for readers into more and less essential reading, allowing them to choose between different reading experiences. As handwritten text that speaks more strongly of a writer than type does, I also suggested that authorial comments are especially able to create a sense of community.

With these points in mind, Chapter 3 focused on the *content* of language in manga by looking at the orthographic styles and lexical characteristics of manga. In particular, I conducted three major studies, on (1) the orthographic characteristics of manga and irregular furigana, (2) the readability of manga and (3) a morpheme

analysis of manga's lexical characteristics. In the first study, I looked at the distribution of the different scripts, demonstrating that one of the characteristics of manga is their use of symbols. While this differentiates manga from other genres, it also has certain similarities with other recent media, so that manga's orthographic patterns helps create and locate a manga style of writing. In the second study, I used tools developed by Sato, Matsuyoshi & Kondoh (2008) and the Readability Research Laboratory (2010) to examine how easy manga are to read. While they graded fairly low, at a 5 to 6 grade level on average, they are only a few grade levels lower than Nishimura's (2011) results for keitai-shousetsu and conventional novels.

I then followed these studies with a morpheme analysis of *Lines*. After reviewing the issues involved in conducting morpheme analyses in Japanese, I used the results from the previous section as a guideline for standardizing the text to make it more appropriate for analysis. I then summarized the data by the distribution of parts of speech seen; the usage of different words by their origin (Sino-Japanese kango, native Japanese wago, borrowed foreign words, etc.); and frequency charts for the top-10 nouns, verbs, adjectives and adverbs. The results will point to additional commonalities with spoken speech in the distribution of wago and interjections. I also pointed out an interesting tendency for shounen-manga to use words related to their narrative themes, as opposed to shoujo-manga, whose frequency lists for nouns pointed instead towards the importance of interpersonal relationships.

Chapter 4 continued to move the focus one layer deeper by looking at specific language patterns seen and how they are used to create patterns; it particularly built up on the last chapter by utilizing the data produced by the morpheme analysis. Japanese language has traditionally been described as being gendered, but in recent years it is said to be going through a process of neutralization. Manga are often said to be influential in such changes, but reports differ on the speech patterns of characters in manga, with some describing highly stereotypical yakuwari-go patterns (Kinsui (2007), Takahashi (2009)), and others describing patterns similar to realistic speech (Ueno (2006), Aizawa (2003)). In response, I conducted three studies related to gender and characterization in manga: first, the distribution of speech between male and female characters, second, the use of personal pronouns, and third, the use of gendered sentence final expressions. While personal pronouns showed relatively conservative gendered patterns, sentence final expressions showed a major gap between shoujo-manga and shounen-manga. More specifically, female characters used many more strongly feminine expressions in shounen-manga. As became clear from the first study, shounen-manga is deeply skewed towards male characters, with very few central female characters, and in addition to more standard reasons such as author interference, I also argued that the peripherality of female characters in shounen-manga made them more susceptible to stereotyped speech patterns.

### 5.1.2 Genre, manga and language

Aside from the more general goal of establishing the characteristics of language in manga, another major goal of this thesis was to consider how language could be used to reconsider genre in manga. Several points within each of the studies suggested that linguistic characteristics were one effective way of looking at genre. Summarizing the results, one could say that in general, shoujo-manga were typified by the diversity of the text types it used; its use of vocabulary suggesting interpersonal relations; their focus on a core group of characters, with balance between male and female characters; and more realistic speech patterns. On the other hand, shounen-manga were typified by using more kanji and being correspondingly more difficult to read; more diversity in the vocabulary used, with more technical or topical vocabulary; featuring more characters, with less balance between the genders; and stereotypical speech patterns which could be considered *yakuwari-go*, particularly for female characters.

While I argued that many of these issues could be explained by differences in plot, it is also possible to state these differences in a different fashion. Specifically, whereas shoujo-manga appears to use structural methods to their advantage—here, typified by their diverse use of text types, but also typical of their special use of frames (Takeuchi O. , 2005)—shounen-manga utilize linguistic methods to its advantage, actively calling upon linguistic stereotypes to create economic characters as well as to actively move forward their narratives. While this means that shounen-manga has shown itself to be somewhat more difficult to read, and, with their use of stereotyped speech patterns, an interesting subject of research, that does not mean that shoujo-manga is inferior. Indeed, as Ôtsuka (1994) has argued, shoujo-manga have actually been criticized for being *difficult* to read: As opposed to difficulty in reading in shounen-manga, however, this is not so much a function of the actual language being seen, but rather how it is presented within the text. While the particulars of how this is balanced out cannot be given full attention in this particular venue, the corpus data has thus shown that language can be used as an effective marker for redefining genre in manga.

### 5.1.3 Future issues

One of the advantages of a corpus is the ability to reuse the data for additional studies depending on the topics of interest: Now that the data has been collected, research topics are only a matter of one's imagination. However, in reviewing the results here, there are still certain issues that would make the corpus vastly more informative. In particular, there are two points that must be left to be considered in a different forum in the future: first, the validity of the statistical analyses described here, and second, the actual reading patterns, habits and impact of manga.

One of the main goals of this thesis was to offer a quantitative analysis of language in manga to complement and expand upon the primarily qualitative analyses that have been conducted hitherto now. Creating a corpus was a major undertaking, and it has clearly offered many interesting points to consider and a great potential for continued future research beyond the studies described here. However, while I have conducted statistical analyses wherever possible, it should be noted that while the corpus comes to a large amount of data, it is still in some ways limited in its statistical power. That is to say, while the amount of text overall is considerable, 10 series, or 5 per each genre, leads to the minimum amount available to conduct statistical analyses. It is unclear how the results might change given more series, but it is not unreasonable to expect that there would be differences. As I discussed in Chapter 3 on the morpheme analysis and lexical items, the results for shounen-manga seem to be largely a result of their peculiar plots, and it may be that adding additional series that are less restricted in plot for both shoujo-manga and shounen-manga would change the results—or, perhaps, confirm that these are valid differences.

Another issue that I was not able to approach here was how manga are actually read, and how readers interact with them. These are points that came up along the way when thinking about the impact of the linguistic structures and patterns seen in manga. One of the themes of Chapter 2, on how different categories of text can potentially create different reading experiences, begs one to think about how manga are really read, and how such hypotheses could be tested experimentally. The discussion in Chapter 4 on characterization through language patterns was started partially with the question of manga's popular linguistic impact in mind, and it, too, would benefit from more input on the reading and linguistic habits of young people. While this was outside the scope of this thesis, it is a point which would be well worth considering in future studies using the corpus data. Given the amount of data collection they would require on top of the already labor-intensive basic corpus construction and morpheme analysis, neither of these points could be properly assessed in this thesis, but they are issues that I plan to approach in future research.



## 5.2 Education and manga

Keeping these issues in mind, it seems appropriate to end the discussion here by considering one way in which the corpus could be used proactively. As I mentioned in Chapter 1, another one of my goals was to approach what kind of a reading experience manga offer, and as I suggested, one field that the data could be useful for is Japanese language education. In recent years, the popularity of Japanese popular culture—especially manga and anime—has been seen as a common reason for studying Japanese as a second language. According to some accounts, as much as 70 to 80 percent of Japanese learners chose to study Japanese because of their interest in manga and anime (Kumano, 2010, p. 89). These trends have resulted in an increase in interest in using manga and anime as tools for studying Japanese, as demonstrated by the Japan Foundation's making researching and developing manga and anime learning tools one of its 2007-2012 mid-term goals (Kumano & Hirokaga, 2008). The importance of interest in popular culture is especially felt given the decline of the influence of other motivating factors such as the economy for Japanese learning (Wasabi Brothers Trading Company, 1998).

As was one of the motivating factors for starting this study, however, with many facets of manga's linguistic characteristics yet to be examined, it has been difficult to assess how one could effectively use manga in the classroom. In this final section, I will grapple with these questions by exploring some of the claimed advantages and disadvantages of using manga in language education, and give them context by introducing some of the characteristics found in the corpus. Taking a critical stance towards the assumed ease and naturalness of learning through manga, I focus on how two points that have come up in the discussions thus far: One, how orthographic styles differ in manga from other standard texts (Chapter 3), and two, how gendered speech patterns help create characters (Chapter 4). Through this final look at manga, I will argue that manga's linguistic characteristics necessitate more consideration before using them, and demonstrate how the data obtained here can help with that.

### 5.2.1 Ways and advantages of using manga for learning

One might generally observe two major approaches when considering how manga might be utilized for educational purposes. The first is what might be called a content-oriented approach, where manga are used to teach a certain theme or subject based upon the stories that it depicts. Two types of manga can be utilized within the content-oriented approach. The first is *kyouiku-manga* (manga for education), or manga specifically written

for learning a given topic like history or economics. The second are authentic texts, which can be used to study Japanese culture and society through the narratives they depict. The second approach to using manga in education is language-oriented; here, manga are used as a linguistic resource for studying Japanese itself. As the focus of this paper is on how manga may be used as a linguistic resource, this paper concerns primarily this latter approach, but it must be remembered that even when using manga for Japanese study, one can of course be informed by the content; that is to say, the use of manga with a language focus offers the chance to study Japanese within a certain cultural context.

This cultural context is in fact one of the major advantages suggested for using manga—and anime—for learning. Manga are often said to be a reflection of the “real” Japan, particularly because they are produced for domestic consumption (Wasabi Brothers Trading Company, 1998) and they are often said to be a good way of studying the “Japanese condition” (Murakami Y. , 2008). Manga are also a convenient medium as they are cheap and easy to obtain (Larose, 1993 in Okazaki (1993, p. 48)). As a medium based on paper alone, even less technologically-equipped classrooms can use them (Murakami Y. , 2008). They can also be fun for students, because the text is easy to deal with (Larose, 1993 in Okazaki (1993, p. 48)), and because students can get into them easily (Okazaki, 1993). The familiarity students have with the material may allow learners to become more engaged than with text-only materials (Kaneko, 2008). Finally, as a visual medium, manga have a strong sense of context, as the drawings make who is speaking to whom obvious and the situations more concrete (Kaneko, 2008).

On the other hand, students are not always positive about using manga in the classroom. Some may be turned off by it, especially in countries where they are not popular (Murakami Y. , 2008), or if they are not interested in them, such as Makino (2010) found with anime. Even amongst fans, using popular culture in the classroom may make some students feel their private space is being encroached upon. Manga can also be costly in comparison to the amount of input they provide, as the number of words per page is low (Murakami Y. , 2008). Gaps between the interests of teachers and students can make it effortful to look for works that students will enjoy (Murakami Y. , 2008). When poor choices are made, they can result in a lack of engagement with the materials, as evidenced by the way Doraemon fell flat as the Japanese Cultural Ambassador in Europe and North America (Kelts, 2010). The linguistic dependency in manga is also low: Manga’s rich non-linguistic visual context may allow students to not pay close attention to language (Murakami Y. , 2008). The content may also not always be appropriate, as violent and sexual themes are not uncommon.

Perhaps most importantly, it is unclear how to use manga appropriately, as there are no established methods for classroom use (Murakami Y. , 2008). As I have pointed out previously, however, while using manga effectively means making appropriate text selection through understanding their characteristics, possibilities,

and limitations (Chinami, 2005), this has not been an easy task given the lack of comprehensive research on language in manga. As I discussed in Chapter 1, because manga are a mixed, visual-linguistic media, there has been a tendency to overlook manga's linguistic properties, favoring its visual structures as more essential (Nakamura T. , 2006). Looking at what this thesis has made clear thus far, two of the perceived advantages of manga—their ease of reading and the naturalness of their speech patterns—are not as clear as has been presumed.

While I review those points in more detail below, note that one way of attempting to overcome this information gap in some manga-for-learning textbooks has been to use original, made-for-textbook manga, thus making them a form of *kyouiku-manga*; however, the use of made-for-textbook manga is likely also due to the expense of acquiring the rights for popular series—a problem experienced by *MangaJin*, a now defunct manga-for-learning magazine (Wasabi Brothers Trading Company, 1998). However, such made-for-textbook manga are consciously different from authentic texts in many ways. As controlled works, they are less complicated and diverse, as they do not continue at long-length like today's narrative-driven story-manga. For example, one textbook, *Japanese in MangaLand* (Bernabe, 2003), introduces a new grammatical point each chapter, and shows examples of them with one-frame original manga before giving practice problems. However, with only a few free-standing frames of manga per lesson, the bulk of learning actually takes place through traditional methods.

While there are positive points to such approaches, the use of non-authentic manga may miss the point: The manga students are interested in are authentic series, and they want the skills necessary to read those texts, skills which may not be as necessary for made-for-textbook manga. Of course, some tools have been developed with authentic manga in mind. *Anime-Manga.jp*, a website aiming “to give Japanese learners and anime/manga fans from all over the world an opportunity to learn Japanese in an enjoyable way, using the anime/manga they enjoy as a gateway to their studies (Japan Foundation Japanese-Language Institute Kansai, 2010),” was developed by analyzing the language in popular manga and anime series. Structurally, it features games, quizzes, and scene explanations using vocabulary and kanji for different genres, such as love or ninja stories. However, the website is not actually based upon reading manga, and while it uses language found in manga, the tasks do not involve actually reading manga, making it different in nature from using long narrative manga for teaching.

## **5.2.2 Possible problems made clear by the corpus**

Clearly, more knowledge about manga's linguistic characteristics is necessary, and the data from this corpus could go a long way to filling in that gap. Two of the issues I brought up in this thesis are of particular interest:

manga's use of non-standard orthography and gendered speech patterns. In regard to the first point, one of the key characteristics of language in manga is how it describes spoken speech—and audible sound—through not just written words but also symbols (Numata, 1989). As I showed in Chapter 3, this is largely realized through the use of non-standard orthography, which can be interpreted partially as a way to create orality (e.g., Groensteen (2007, p. 12)). While there has been a tendency to overlook orthography in linguistics research, it may affect second-language learning (Cook & Bassetti, 2005), and manga's unusual orthographic conventions may affect how well learners can process the language available as input. Three points can be given as potential problems when thinking about how manga's orthographic characteristics might affect second language learners.

First is how the text is structurally very broken. While I previously mentioned this as a technical issue for morpheme analysis in Section 3.4, it also places a burden on readers who must make accurate connections between speech bubbles. Second is the distinct punctuation: Many of the unusual non-standard symbols and styles would simply not be seen within Japanese language classrooms, which focus on standard language. Third, script use is often varied and inconsistent, such as I described with how the transformation of /nai/ to [ne:] is expressed (Example 2, p. 25).

Admittedly, the effect of these points on how manga are read is yet unclear; since line breaks and spaces seem to correspond with dependent phrases and deleted particles, they may be an accessible way to engage with naturalistic particle-dropping speech. The use of the syllabaries over kanji may also be easier for beginner readers. Some points may be intuitive, like long-vowel marker bousen with hiragana, as it is an extension of its existing usage. However, spaces do not necessarily make Japanese easier to read, and may actually be redundant (Sainioa, Hyona, Bingushib, & Bertrama, 2007), and variation in scripts may make it difficult to recognize words derivable to single forms. Although manga may have rated lower than conventional novels for ease, the unique styles adopted in manga mean that one cannot assume that they will be easy for the beginner learner, and orthographic issues may contribute to manga literacy in that they must be learned.

As for the second point, as I discussed in Chapter 4, traditional gendered speech has changed greatly in recently years, causing it to be a point of argument in Japanese language education. While textbooks are often more conservative than real speech (Kawasaki & McDougall, 2003), and beginner students may not be able to use them appropriately, advanced students may not seem natural without them (Ogawa, 2006). Given the claims that manga are “naturalistic”, they might be an ideal medium to learn gendered speech as readers can *see* the characters using it: That is, the drawings could potentially help inform readers of the images and different types of people associated the speech patterns they learn.

However, as showed in detail, gendered speech in manga paints a complicated picture in terms of *realism*. While personal pronouns demonstrated distributions not too far off real-life speech, they also lacked

certain forms like *uchi* and *jibun* that are gaining in popularity. SFP usages, however, differed greatly by genre, with shounen-manga using much more markedly gendered language than shoujo-manga, which had similar patterns of SFP usage to those reported for young girls.

In this sense, manga seem to show many different types of gendered speech, and it is clearly incorrect to generalize that all manga are realistic in regards to gendered speech. While manga may aim for realism, as fiction, one must expect some level of stereotyped speech forms, especially as they may be important in character development. Genre appears to be an important factor in determining how common stereotyped forms are, with shounen-manga particularly prone to using them. This suggests that simply choosing series may not be enough: It may also be important to consider genre. These results also suggest that it may be necessary to reconsider what is meant by *realistic*: Being unrealistic in one way does not mean it is not in others. For example, as text meant to represent conversation, the low overall occurrence rate of first person pronouns in *Lines* (3,072 instances over 33,236 entries spoken by male or female characters (9.24%)) means that many sentences likely do not have an overt subject, a characteristic of spoken Japanese.

### 5.2.3 Rethinking manga in the classroom

While using manga in the classroom might get students engaged in new ways, their linguistic landscapes may not be as clear-cut as has been supposed. It may be necessary to actively deal with how text is presented, and to reconsider what makes it realistic and its nature as a language model for students. Knowing more about the linguistic characteristics of different genres and series may prove helpful in making appropriate selections of series for different levels and purposes, and may also be helpful in thinking about what kind of knowledge may be necessary to process them. Taking the time to talk about these points with students could be a good opportunity to think together about how and why texts differ, and as a corpus of popular manga, the characteristics identified in the current study could be helpful to understanding how to work with students' interests. Furthermore, comics being an international medium, the points raised here may also be applicable to languages other than Japanese, using either manga in translation or non-Japanese comics.

As a preliminary step in thinking about manga's linguistic characteristics, there are still many important questions left to consider. In particular, the vocabulary and grammatical forms which manga readers will meet should be examined to determine how difficult manga are and what makes them realistic (or not). Generally, the language seen in shoujo-manga appear to be somewhat more oriented towards everyday life situations, and given

that they ranked somewhat lower for readability, they might be a good choice for use in the classroom. Using shounen-manga to teach might require making lesson plans that deal specifically with the more technical vocabulary that is seen there—although this could be used as an advantage, as it could be an opportunity to learn more specialized language,

Indeed, there are already several such possibilities that could actively use the issues above for productive and positive purposes. Discussing manga's orthographic styles could become a chance to talk about how writers choose which scripts to use, and what the effects of such choices are stylistically. One could compare them with popular mediums like blogs, texting or keitai-shousetsu, which also use unusual orthography (Akitsuki (2009), Coates (2010), Nishimura (2003), (2011)). In addition, as Chinami (2005, p. 137) notes, because the unusual orthography one sees in manga is reflective of the linguistic changes that occur in spoken speech, such orthographic peculiarities could, if dealt with appropriately, become an opportunity for introspection on the phonetics and spoken speech.

Utilizing manga in class could also provide an opportunity to consider the dynamism involved in using gendered speech. Instead of treating them as a static feminine/masculine dichotomy, manga could be used to show the different factors that go into choosing speech patterns. The differences seen in shounen-manga and shoujo-manga may be particularly useful for considering how attitudes and context change gendered speech patterns. While specific lesson plans must be considered in a different forum, I believe that the data from the corpus could be one useful way of thinking about how manga could be potentially used for educational purposes. One thing that is certain is that knowing more about manga may prove useful to teachers, whether or not they are to be incorporated into the curriculum: With many students reading manga on their own, teachers should be prepared for the questions they will surely bring to class.

## 5.3 Final words

In ending the discussion with manga and education, the discussion has come full circle, as it cumulates with specific details on how the data collected could contribute to other efforts beyond this thesis. While the discussion in this thesis began with the observation that not very much is known about language in manga, through corpus-based, empirical observations about the structural and lexical characteristics of manga, these five chapters have gone a long way to filling in those gaps in our understanding. Of course, designing and maintaining a corpus is always a work in progress, and there are still issues that should be accounted for, as I mentioned above. Situated as manga are as a multimodal, mixed medium, however, the corpus described here is not just helpful for manga studies alone. While manga may seem obscure at first, when re-evaluated in light of their special linguistic characteristics, they clearly form a unique and relevant resource, that can be used not just for manga research, but also for the research of larger linguistic question. In designing the corpus, I also laid out plans that may be helpful for those approaching similar media, from other cartoon forms to things like advertisements, which also use text in unusual ways. In that sense, the lessons learnt from the corpus and various studies described here may be of use to other individuals interested in non-traditional sources of text, and I hope that this may be a resource to others.

## 6. Appendix

### 6.1 Corpus samples

**Sample 1:** Pages from *Bokura ga Ita* (2:14~15) [raw manga]





Sample 2: Pages from *Bokura ga Ita* (2:14~15) [corpus data]

Genre	Series	Data #	Vol.	Page	Frame	Cat.	In-F #	Sex	Speaker	Ch-T	Text+&	Char#1
Shoujo-manga	Bokura	2125	2	14	1	1	1	2	七美	I	ご&&ごめんね&&遅れて	11
Shoujo-manga	Bokura	2126	2	14	1	1	2	2	髪&&の長い子	I	走って&&来たの？	8
Shoujo-manga	Bokura	2127	2	14	1	1	3	2	みずちん	I	え&&ナナ なんか&&今日すっごく&&かわいくない？	24
Shoujo-manga	Bokura	2128	2	14	2	1	1	2	七美	I	・・・そ&&そお？	9
Shoujo-manga	Bokura	2129	2	14	2	1	2	2	みずちん	I	うん&&&ゆかた&&似合うよ&&びっくり	17
Shoujo-manga	Bokura	2130	2	14	2	1	3	2	髪&&の長い子	I	その&&ゆかたの色&&ピッタリ	13
Shoujo-manga	Bokura	2131	2	14	2	7	1	3	不明語り手	0	山吹色と&&ピンク	8
Shoujo-manga	Bokura	2132	2	14	3	1	1	2	七美	I	あ&&&いとこの&&お姉ちゃんちまで&&借りに行ってきたの	26
Shoujo-manga	Bokura	2133	2	14	3	2	1	2	七美	I	え&&本当に	5
Shoujo-manga	Bokura	2134	2	14	3	1	2	2	七美	I	あたしの&&中 1 に作った&&やつで&&子供っぽい&&柄だった&&から	30
Shoujo-manga	Bokura	2135	2	14	3	1	3	2	みずちん	I	へー	4
Shoujo-manga	Bokura	2136	2	14	3	1	4	2	髪&&の長い子	I	チークも&&すっごい&&かわいー	14
Shoujo-manga	Bokura	2137	2	14	3	2	2	2	七美	I	似合ってる&&かな	8
Shoujo-manga	Bokura	2138	2	14	4	1	1	2	髪&&の長い子	I	湯上り&&みたいー	10
Shoujo-manga	Bokura	2139	2	14	4	6	1	2	七美	I	あつ これ&&アブリコット	12
Shoujo-manga	Bokura	2140	2	14	4	2	1	2	七美	I	大丈夫&&かなあ	7
Shoujo-manga	Bokura	2141	2	14	5	1	1	1	クラスの男子B	E	おー&&&ゆかただ	8
Shoujo-manga	Bokura	2142	2	14	5	1	2	1	クラスの男子A	E	ゆかた&&軍団	6
Shoujo-manga	Bokura	2143	2	14	5	4	1	0	0	0	ビューー	5
Shoujo-manga	Bokura	2144	2	14	6	1	1	2	みずちん・髪&&の長い子	E	男子&&来た	5
Shoujo-manga	Bokura	2145	2	14	6	4	1	0	0	0	ドキ	2
Shoujo-manga	Bokura	2146	2	15	1	1	1	2	みずちん	I	あれ&&&矢野&&は？	9
Shoujo-manga	Bokura	2147	2	15	1	1	2	1	クラスの男子A	E	あー&&&それがさ&&連絡つかなくて	16
Shoujo-manga	Bokura	2148	2	15	2	1	1	2	七美	I	・・・え	4
Shoujo-manga	Bokura	2149	2	15	2	1	2	2	七美	I	来ないの？	5
Shoujo-manga	Bokura	2150	2	15	2	1	3	1	クラスの男子A	E	いや&&&多分 地元の&&やつらと&&来てんじや&&ねえ？	25
Shoujo-manga	Bokura	2151	2	15	2	6	1	1	クラスの男子A	E	メール&&入れといたし	10
Shoujo-manga	Bokura	2152	2	15	2	1	4	1	クラスの男子A	E	その辺で&&会うかも	9
Shoujo-manga	Bokura	2153	2	15	3	2	1	2	七美	I	・・・来ると	6
Shoujo-manga	Bokura	2154	2	15	3	2	2	2	七美	I	いいな	3
Shoujo-manga	Bokura	2155	2	15	4	1	1	1	クラスの男子A	E	高橋&&別人みてー	8
Shoujo-manga	Bokura	2156	2	15	4	1	1	2	七美	I	え？	2
Shoujo-manga	Bokura	2157	2	15	4	1	2	2	みずちん	I	ねー	2
Shoujo-manga	Bokura	2158	2	15	4	1	3	2	みずちん	I	このコ ゆかた&&似合うでしょー	15
Shoujo-manga	Bokura	2159	2	15	5	1	1	1	クラスの男子	I	童顔 チビは&&意外と和装が&&似合うという	20
Shoujo-manga	Bokura	2160	2	15	5	1	2	1	クラスの男子	I	普段の高橋知つてると&&めっちゃめっちゃサギな&&気がすんな	26
Shoujo-manga	Bokura	2161	2	15	5	6	1	2	七美	I	えへへ&&そんなあ	8
Shoujo-manga	Bokura	2162	2	15	5	2	1	2	七美	I	ああ&&惜しい&&あたし	10
Shoujo-manga	Bokura	2163	2	15	5	2	2	2	七美	I	ゆかたが&&制服だったら&&モテモテ人生&&だったかも	24
Shoujo-manga	Bokura	2164	2	15	5	1	3	1	クラスの男子	I	サギか	3
Shoujo-manga	Bokura	2165	2	15	5	1	4	1	クラスの男子	I	サギだよ	4
Shoujo-manga	Bokura	2166	2	15	7	2	1	2	七美	I	ああ&&もう&&矢野！！	10
Shoujo-manga	Bokura	2167	2	15	7	2	2	2	七美	I	何やってん&&のさー！！	13
Shoujo-manga	Bokura	2168	2	15	7	1	1	2	七美	I	あ&&&ヒヨコ	6
Shoujo-manga	Bokura	2169	2	15	7	5	1	0	0	0	ひよこ 1 羽 4 0 0 円	11
Shoujo-manga	Bokura	2170	2	15	8	1	1	2	髪&&の長い子	I	カワイイ	4
Shoujo-manga	Bokura	2171	2	15	8	1	2	2	みずちん	I	欲しいね	4
Shoujo-manga	Bokura	2172	2	15	8	1	3	2	七美	I	でも大きくなると&&ニワトリになると&&いう欠点が	23
Shoujo-manga	Bokura	2173	2	15	9	1	1	2	七美	I	ヨーヨー&&やろー	8

### Sample 3: Pages from *Bokura ga Ita* (2:14) [morpheme analysis data]

Genre	Series	Data #	Mor. #	Sex	Speaker	Type	Bou.	Main form	Phonetic	Root-phon.	Root	POF	Con. C.	F. of C.	Form	E-Type
Shoujo	Bokura	2125	6775	2	七美	I	B	ご	ゴ	ゴ	ご	感動詞-カム			ゴ	他
Shoujo	Bokura	2125	6776	2	七美	I	I	ごめん	ゴメン	ゴメン	御免	名詞-普通名詞-一般			ゴメン	漢
Shoujo	Bokura	2125	6777	2	七美	I	I	ね	ネ	ネ	ね	助詞-終助詞			ネ	和
Shoujo	Bokura	2125	6778	2	七美	I	I	遅れ	オクレ	オクレル	遅れる	動詞-非自立可能	下一段-ラ行-一般	連用形-一般	オクレル	和
Shoujo	Bokura	2125	6779	2	七美	I	I	て	テ	テ	て	助詞-接続助詞			テ	和
Shoujo	Bokura	2126	6780	2	七美	I	I	走っ	ハシッ	ハシル	走る	動詞-一般	五段-ラ行-一般	連用形-促音便	ハシル	和
Shoujo	Bokura	2126	6781	2	七美	I	I	来	テ	テ	て	助詞-接続助詞			テ	和
Shoujo	Bokura	2126	6782	2	七美	I	I	来	キ	クル	来る	動詞-非自立可能	カ行変格	連用形-一般	クル	和
Shoujo	Bokura	2126	6783	2	七美	I	I	た	タ	タ	た	助動詞	助動詞-タ	連体形-一般	タ	和
Shoujo	Bokura	2126	6784	2	七美	I	I	の	ノ	ノ	の	助詞-終助詞			ノ	和
Shoujo	Bokura	2126	6785	2	七美	I	I	？			？	補助記号-句点				記号
Shoujo	Bokura	2127	6786	2	みずちん	I	B	え	エ	エー	えー	感動詞-フィラー			エ	和
Shoujo	Bokura	2127	6787	2	みずちん	I	I	ナナ	ナナ	ナナ	ナナミ	名詞-固有有名詞-人名-あだ名			ナナ	和
Shoujo	Bokura	2127	6788	2	みずちん	I	I					空白				記号
Shoujo	Bokura	2127	6789	2	みずちん	I	I	なんか	ナンカ	ナンカ	なんか	助詞-副助詞			ナンカ	和
Shoujo	Bokura	2127	6790	2	みずちん	I	I	今日	キョー	キョウ	今日	名詞-普通名詞-副詞可能			キョウ	和
Shoujo	Bokura	2127	6791	2	みずちん	I	I	すっごく	スッゴク	スゴイ	凄い	形容詞-一般	形容詞	連用形-一般	スッゴイ	和
Shoujo	Bokura	2127	6792	2	みずちん	I	I	かわいく	カワイク	カワイイ	可愛い	形容詞-一般	形容詞	連用形-一般	カワイイ	和
Shoujo	Bokura	2127	6793	2	みずちん	I	I	ない	ナイ	ナイ	無い	形容詞-非自立可能	形容詞	終止形-一般	ナイ	和
Shoujo	Bokura	2127	6794	2	みずちん	I	I	？			？	補助記号-句点				記号
Shoujo	Bokura	2128	6795	2	七美	I	B	．．			．．	補助記号-一般				記号
Shoujo	Bokura	2128	6796	2	七美	I	I	そ	ソ	ソ	そ	感動詞-カム			ソ	他
Shoujo	Bokura	2128	6797	2	七美	I	I	そお	ソー	ソウ	そう	副詞			ソウ	和
Shoujo	Bokura	2128	6798	2	七美	I	I	？			？	補助記号-句点				記号
Shoujo	Bokura	2129	6799	2	みずちん	I	B	うん	ウン	ウン	うん	感動詞-一般			ウン	和
Shoujo	Bokura	2129	6800	2	みずちん	I	I	ゆかた	ユカタ	ユカタ	浴衣	名詞-普通名詞-一般			ユカタ	和
Shoujo	Bokura	2129	6801	2	みずちん	I	I	似合う	ニアウ	ニアウ	似合う	動詞-一般	五段-ワア行-一般	終止形-一般	ニアウ	和
Shoujo	Bokura	2129	6802	2	みずちん	I	I	よ	ヨ	ヨ	よ	助詞-終助詞			ヨ	和
Shoujo	Bokura	2129	6803	2	みずちん	I	I	びっくり	ビックリ	ビックリ	びっくり	名詞-普通名詞-サ変可能			ビックリ	和
Shoujo	Bokura	2130	6804	2	七美	I	B	その	ソノ	ソノ	其の	連体詞			ソノ	和
Shoujo	Bokura	2130	6805	2	七美	I	I	ゆかた	ユカタ	ユカタ	浴衣	名詞-普通名詞-一般			ユカタ	和
Shoujo	Bokura	2130	6806	2	七美	I	I	の	ノ	ノ	の	助詞-格助詞			ノ	和
Shoujo	Bokura	2130	6807	2	七美	I	I	色	イロ	イロ	色	名詞-普通名詞-一般			イロ	和
Shoujo	Bokura	2130	6808	2	七美	I	I	ピッタリ	ピッタリ	ピッタリ	ぴったり	副詞			ピッタリ	和
Shoujo	Bokura	2132	6809	2	七美	I	B	あ	ア	アッ	あっ	感動詞-一般			ア	和
Shoujo	Bokura	2132	6810	2	七美	I	I	いとこ	イトコ	イトコ	従兄弟	名詞-普通名詞-一般			イトコ	和
Shoujo	Bokura	2132	6811	2	七美	I	I	の	ノ	ノ	の	助詞-格助詞			ノ	和
Shoujo	Bokura	2132	6812	2	七美	I	I	お	オ	オ	御	接頭辞			オ	和
Shoujo	Bokura	2132	6813	2	七美	I	I	姉	ネー	ネエ	姉	名詞-普通名詞-一般			ネエ	和
Shoujo	Bokura	2132	6814	2	七美	I	I	ちゃん	チャン	チャン	ちゃん	接尾辞-名詞的-一般			チャン	和
Shoujo	Bokura	2132	6815	2	七美	I	I	ち	チ	ウチ	家	名詞-普通名詞-一般			チ	和
Shoujo	Bokura	2132	6816	2	七美	I	I	まで	マデ	マデ	まで	助詞-副助詞			マデ	和
Shoujo	Bokura	2132	6817	2	七美	I	I	借り	カリ	カリル	借りる	動詞-一般	上一段-ラ行	連用形-一般	カリル	和
Shoujo	Bokura	2132	6818	2	七美	I	I	に	ニ	ニ	に	助詞-格助詞			ニ	和
Shoujo	Bokura	2132	6819	2	七美	I	I	行っ	イク	イク	行く	動詞-非自立可能	五段-カ行-イク	連用形-促音便	イク	和
Shoujo	Bokura	2132	6820	2	七美	I	I	て	テ	テ	て	助詞-接続助詞			テ	和
Shoujo	Bokura	2132	6821	2	七美	I	I	き	キ	クル	来る	動詞-非自立可能	カ行変格	連用形-一般	クル	和
Shoujo	Bokura	2132	6822	2	七美	I	I	た	タ	タ	た	助動詞	助動詞-タ	連体形-一般	タ	和
Shoujo	Bokura	2132	6823	2	七美	I	I	の	ノ	ノ	の	助詞-終助詞			ノ	和
Shoujo	Bokura	2134	6824	2	七美	I	B	あたし	アタシ	アタシ	あたし	代名詞			アタシ	和
Shoujo	Bokura	2134	6825	2	七美	I	I	の	ノ	ノ	の	助詞-格助詞			ノ	和
Shoujo	Bokura	2134	6826	2	七美	I	I	中1	チュウイチ	チュウイチ	中一	名詞-普通名詞-一般			チュウイチ	漢
Shoujo	Bokura	2134	6827	2	七美	I	I	に	ニ	ニ	に	助詞-格助詞			ニ	和
Shoujo	Bokura	2134	6828	2	七美	I	I	作っ	ツクッ	ツクル	作る	動詞-一般	五段-ラ行-一般	連用形-促音便	ツクル	和
Shoujo	Bokura	2134	6829	2	七美	I	I	た	タ	タ	た	助動詞	助動詞-タ	連体形-一般	タ	和
Shoujo	Bokura	2134	6830	2	七美	I	I	やつ	ヤツ	ヤツ	奴	名詞-普通名詞-一般			ヤツ	和
Shoujo	Bokura	2134	6831	2	七美	I	I	で	デ	デ	で	助詞-格助詞			デ	和
Shoujo	Bokura	2134	6832	2	七美	I	I	子供	コドモ	コドモ	子供	名詞-普通名詞-一般			コドモ	和
Shoujo	Bokura	2134	6833	2	七美	I	I	っぽい	ツポイ	ボイ	ばい	接尾辞-形容詞的	形容詞	連体形-一般	ツポイ	和
Shoujo	Bokura	2134	6834	2	七美	I	I	柄	ガラ	ガラ	柄	名詞-普通名詞-一般			ガラ	和
Shoujo	Bokura	2134	6835	2	七美	I	I	だっ	ダッ	ダ	だ	助動詞	助動詞-ダ	連用形-促音便	ダ	和
Shoujo	Bokura	2134	6836	2	七美	I	I	た	タ	タ	た	助動詞	助動詞-タ	終止形-一般	タ	和
Shoujo	Bokura	2134	6837	2	七美	I	I	から	カラ	カラ	から	助詞-接続助詞			カラ	和
Shoujo	Bokura	2135	6838	2	みずちん	I	B	へー	ヘー	ヘー	へー	感動詞-一般			ヘー	和
Shoujo	Bokura	2136	6839	2	七美	I	B	チーク	チーク	チーク	チーク-cheek	名詞-普通名詞-サ変可能			チーク	外
Shoujo	Bokura	2136	6840	2	七美	I	I	も	モ	モ	も	助詞-係助詞			モ	和
Shoujo	Bokura	2136	6841	2	七美	I	I	すっごい	スッゴイ	スゴイ	凄い	形容詞-一般	形容詞	終止形-一般	スッゴイ	和
Shoujo	Bokura	2136	6842	2	七美	I	I	かわいー	カワイー	カワイイ	可愛い	形容詞-一般	形容詞	終止形-一般	カワイー	和
Shoujo	Bokura	2138	6843	2	七美	I	B	湯上り	ユアガリ	ユアガリ	湯上がり	名詞-普通名詞-一般			ユアガリ	和
Shoujo	Bokura	2138	6844	2	七美	I	I	みたい	ミタイ	ミタイ	みたい	形状詞-助動詞語幹			ミタイ	和
Shoujo	Bokura	2141	6845	1	クラスの男子B	E	B	おー	オー	オー	おー	感動詞-一般			オー	和
Shoujo	Bokura	2141	6846	1	クラスの男子B	E	I	ゆかた	ユカタ	ユカタ	浴衣	名詞-普通名詞-一般			ユカタ	和
Shoujo	Bokura	2141	6847	1	クラスの男子B	E	I	だ	ダ	ダ	だ	助動詞	助動詞-ダ	終止形-一般	ダ	和
Shoujo	Bokura	2142	6848	1	クラスの男子A	E	B	ゆかた	ユカタ	ユカタ	浴衣	名詞-普通名詞-一般			ユカタ	和
Shoujo	Bokura	2142	6849	1	クラスの男子A	E	I	軍団	グンダン	グンダン	軍団	名詞-普通名詞-一般			グンダン	漢
Shoujo	Bokura	2144	6850	2	みずちん・七美の長い子	E	B	男子	ダンシ	ダンシ	男子	名詞-普通名詞-一般			ダンシ	漢
Shoujo	Bokura	2144	6851	2	みずちん・七美の長い子	E	I	来	キ	クル	来る	動詞-非自立可能	カ行変格	連用形-一般	クル	和
Shoujo	Bokura	2144	6852	2	みずちん・七美の長い子	E	I	た	タ	タ	た	助動詞	助動詞-タ	終止形-一般	タ	和

# Sample 4: Pages from *Bokura ga Ita* (2:15) [morpheme analysis data]

Genre	Series	Data #	Mor. #	Sex	Speaker	Type	Bou.	Main form	Phonetic	Root-phon.	Root	POF	Con. C.	F. of C.	Form	E-Type
Shoujo	Bokura	2146	6853	2	みずちん	I	B	あれ	アレ	アレ	あれ	感動詞一般			アレ	和
Shoujo	Bokura	2146	6854	2	みずちん	I	I	矢野	ヤノ	ヤノ	ヤノ	名詞 固有名詞 人名 姓			ヤノ	和
Shoujo	Bokura	2146	6855	2	みずちん	I	I	は	ハ	ハ	は	助詞 係助詞			ハ	和
Shoujo	Bokura	2146	6856	2	みずちん	I	I	？	？	？	？	補助記号 句点				記号
Shoujo	Bokura	2147	6857	1	クラスの男子 A	E	B	あー	アー	アー	あー	感動詞一般			アー	和
Shoujo	Bokura	2147	6858	1	クラスの男子 A	E	I	それ	ソレ	ソレ	其れ	代名詞			ソレ	和
Shoujo	Bokura	2147	6859	1	クラスの男子 A	E	I	が	ガ	ガ	が	助詞 格助詞			ガ	和
Shoujo	Bokura	2147	6860	1	クラスの男子 A	E	I	さ	サ	サ	さ	助詞 終助詞			サ	和
Shoujo	Bokura	2147	6861	1	クラスの男子 A	E	I	連絡	レンラク	レンラク	連絡	名詞 普通名詞 抽象可能			レンラク	和
Shoujo	Bokura	2147	6862	1	クラスの男子 A	E	I	つか	ツク	ツク	付く	動詞 非自立可能	五段 力行一般	未然形一般	ツク	和
Shoujo	Bokura	2147	6863	1	クラスの男子 A	E	I	なく	ナク	ナイ	ない	助動詞	助動詞 ナイ	連用形一般	ナイ	和
Shoujo	Bokura	2147	6864	1	クラスの男子 A	E	I	て	テ	テ	て	助詞 接続助詞			テ	和
Shoujo	Bokura	2148	6865	2	七美	I	B	・・・	・・・	・・・	・・・	補助記号 一般				記号
Shoujo	Bokura	2148	6866	2	七美	I	I	え	エ	エー	えー	感動詞 フィラー			エ	和
Shoujo	Bokura	2149	6867	2	七美	I	B	来	コ	クル	来る	動詞 非自立可能	力行変格	未然形一般	クル	和
Shoujo	Bokura	2149	6868	2	七美	I	I	ない	ナイ	ナイ	ない	助動詞	助動詞 ナイ	連体形一般	ナイ	和
Shoujo	Bokura	2149	6869	2	七美	I	I	の	ノ	ノ	の	助詞 終助詞			ノ	和
Shoujo	Bokura	2149	6870	2	七美	I	I	？	？	？	？	補助記号 句点				記号
Shoujo	Bokura	2150	6871	1	クラスの男子 A	E	B	いや	イヤ	イヤ	否	感動詞一般			イヤ	和
Shoujo	Bokura	2150	6872	1	クラスの男子 A	E	I	多分	タブン	タブン	多分	副詞			タブン	漢
Shoujo	Bokura	2150	6873	1	クラスの男子 A	E	I					空白				記号
Shoujo	Bokura	2150	6874	1	クラスの男子 A	E	I	地元	ジモト	ジモト	地元	名詞 普通名詞 一般			ジモト	混
Shoujo	Bokura	2150	6875	1	クラスの男子 A	E	I	の	ノ	ノ	の	助詞 格助詞			ノ	和
Shoujo	Bokura	2150	6876	1	クラスの男子 A	E	I	やつ	ヤツ	ヤツ	奴	名詞 普通名詞 一般			ヤツ	和
Shoujo	Bokura	2150	6877	1	クラスの男子 A	E	I	ら	ラ	ラ	等	接尾辞 名詞 一般			ラ	和
Shoujo	Bokura	2150	6878	1	クラスの男子 A	E	I	と	ト	ト	と	助詞 格助詞			ト	和
Shoujo	Bokura	2150	6879	1	クラスの男子 A	E	I	来	キ	クル	来る	動詞 非自立可能	力行変格	連用形一般	クル	和
Shoujo	Bokura	2150	6880	1	クラスの男子 A	E	I	て	テ	テル	てる	助動詞	下一段 タ行	連体形 省略	テル	和
Shoujo	Bokura	2150	6881	1	クラスの男子 A	E	I	ん	ン	ン	の	助詞 準体助詞			ン	和
Shoujo	Bokura	2150	6882	1	クラスの男子 A	E	I	じゃ	ジャ	だ	だ	助動詞	助動詞 ダ	連用形 融合	ダ	和
Shoujo	Bokura	2150	6883	1	クラスの男子 A	E	I	ねえ	ネー	ナイ	無い	形容詞 非自立可能	形容詞	終止形一般	ナイ	和
Shoujo	Bokura	2150	6884	1	クラスの男子 A	E	I	？	？	？	？	補助記号 句点				記号
Shoujo	Bokura	2152	6885	1	クラスの男子 A	E	B	その	ソノ	ソノ	其の	連体詞			ソノ	和
Shoujo	Bokura	2152	6886	1	クラスの男子 A	E	I	辺	ヘン	ヘン	辺	名詞 普通名詞 一般			ヘン	漢
Shoujo	Bokura	2152	6887	1	クラスの男子 A	E	I	で	デ	デ	で	助詞 格助詞			デ	和
Shoujo	Bokura	2152	6888	1	クラスの男子 A	E	I	会う	アウ	アウ	会う	動詞 一般	五段 ワア行一般	終止形一般	アウ	和
Shoujo	Bokura	2152	6889	1	クラスの男子 A	E	I	かも	カモ	カモ	かも	助詞 係助詞			カモ	和
Shoujo	Bokura	2155	6890	1	クラスの男子 A	E	B	高橋	タカハシ	タカハシ	タカハシ	名詞 固有名詞 人名 姓			タカハシ	漢
Shoujo	Bokura	2155	6891	1	クラスの男子 A	E	I	別人	ベツジン	ベツジン	別人	名詞 普通名詞 一般			ベツジン	漢
Shoujo	Bokura	2155	6892	1	クラスの男子 A	E	I	みてー	ミタイ	ミタイ	みたい	形状詞 助動詞 語幹			ミタイ	和
Shoujo	Bokura	2156	6893	2	七美	I	B	え	エ	エー	えー	感動詞 フィラー			エ	和
Shoujo	Bokura	2156	6894	2	七美	I	I	？	？	？	？	補助記号 句点				記号
Shoujo	Bokura	2157	6895	2	みずちん	I	B	ねー	ネー	ネエ	ねえ	感動詞一般			ネエ	和
Shoujo	Bokura	2158	6896	2	みずちん	I	B	この	コノ	コノ	此の	連体詞			コノ	和
Shoujo	Bokura	2158	6897	2	みずちん	I	I	コ	コ	コ	子	名詞 普通名詞 一般			コ	和
Shoujo	Bokura	2158	6898	2	みずちん	I	I					空白				記号
Shoujo	Bokura	2158	6899	2	みずちん	I	I	ゆかた	ユカタ	ユカタ	浴衣	名詞 普通名詞 一般	五段 ワア行一般	終止形一般	ユカタ	和
Shoujo	Bokura	2158	6900	2	みずちん	I	I	似合う	ニアウ	ニアウ	似合う	動詞 一般	助動詞	意志推量形	ニアウ	和
Shoujo	Bokura	2158	6901	2	みずちん	I	I	でしょー	デショー	デス	です	助動詞	助動詞 デス		デス	和
Shoujo	Bokura	2159	6902	1	クラスの男子	I	B	童顔	ドウガン	ドウガン	童顔	名詞 普通名詞 一般			ドウガン	漢
Shoujo	Bokura	2159	6903	1	クラスの男子	I	I					空白				記号
Shoujo	Bokura	2159	6904	1	クラスの男子	I	I	ちび	チビ	チビ	ちび	名詞 普通名詞 形状詞 可能			チビ	和
Shoujo	Bokura	2159	6905	1	クラスの男子	I	I	は	ハ	ハ	は	助詞 係助詞			ハ	和
Shoujo	Bokura	2159	6906	1	クラスの男子	I	I	意外	イガイ	イガイ	意外	形状詞 一般			イガイ	漢
Shoujo	Bokura	2159	6907	1	クラスの男子	I	I	と	ト	ト	と	助詞 格助詞			ト	和
Shoujo	Bokura	2159	6908	1	クラスの男子	I	I	和装	ワソウ	ワソウ	和装	名詞 普通名詞 一般			ワソウ	漢
Shoujo	Bokura	2159	6909	1	クラスの男子	I	I	が	ガ	ガ	が	助詞 格助詞			ガ	和
Shoujo	Bokura	2159	6910	1	クラスの男子	I	I	似合う	ニアウ	ニアウ	似合う	動詞 一般	五段 ワア行一般	終止形一般	ニアウ	和
Shoujo	Bokura	2159	6911	1	クラスの男子	I	I	と	ト	ト	と	助詞 格助詞			ト	和
Shoujo	Bokura	2159	6912	1	クラスの男子	I	I	いう	ユウ	イウ	言う	動詞 一般	五段 ワア行 イウ	終止形一般	イウ	和
Shoujo	Bokura	2160	6913	1	クラスの男子	I	B	普段	フダン	フダン	普段	名詞 普通名詞 副詞 可能			フダン	漢
Shoujo	Bokura	2160	6914	1	クラスの男子	I	I	の	ノ	ノ	の	助詞 格助詞			ノ	和
Shoujo	Bokura	2160	6915	1	クラスの男子	I	I	高橋	タカハシ	タカハシ	タカハシ	名詞 固有名詞 人名 姓			タカハシ	漢
Shoujo	Bokura	2160	6916	1	クラスの男子	I	I	知っ	シル	シル	知る	動詞 一般	五段 ラ行 一般	連用形 促音便	シル	和
Shoujo	Bokura	2160	6917	1	クラスの男子	I	I	てる	テル	テル	てる	助動詞	下一段 タ行	終止形一般	テル	和
Shoujo	Bokura	2160	6918	1	クラスの男子	I	I	と	ト	ト	と	助詞 接続助詞			ト	和
Shoujo	Bokura	2160	6919	1	クラスの男子	I	I	めっちゃめっちゃ	メチャメチャ	メチャメチャ	目茶目茶	副詞			メチャメチャ	和
Shoujo	Bokura	2160	6920	1	クラスの男子	I	I	サギ	サギ	サギ	詐欺	名詞 普通名詞 一般			サギ	漢
Shoujo	Bokura	2160	6921	1	クラスの男子	I	I	な	ナ	ダ	だ	助動詞	助動詞 ダ	連体形一般	ダ	和
Shoujo	Bokura	2160	6922	1	クラスの男子	I	I	気	キ	キ	気	名詞 普通名詞 一般			キ	漢
Shoujo	Bokura	2160	6923	1	クラスの男子	I	I	が	ガ	ガ	が	助詞 格助詞			ガ	和
Shoujo	Bokura	2160	6924	1	クラスの男子	I	I	ずん	スン	スル	ゐる	動詞 非自立可能	サ行変格	終止形 撥音便	スル	和
Shoujo	Bokura	2160	6925	1	クラスの男子	I	I	な	ナ	サギ	な	助詞 終助詞			ナ	和
Shoujo	Bokura	2164	6926	1	クラスの男子	I	B	サギ	サギ	サギ	罵	名詞 普通名詞 一般			サギ	和
Shoujo	Bokura	2164	6927	1	クラスの男子	I	I	か	カ	サギ	か	助詞 終助詞			カ	和
Shoujo	Bokura	2165	6928	1	クラスの男子	I	B	サギ	サギ	サギ	罵	名詞 普通名詞 一般			サギ	和
Shoujo	Bokura	2165	6929	1	クラスの男子	I	I	だ	ダ	ダ	だ	助動詞	助動詞 ダ	終止形一般	ダ	和
Shoujo	Bokura	2165	6930	1	クラスの男子	I	I	よ	ヨ	ヨ	よ	助詞 終助詞			ヨ	和
Shoujo	Bokura	2168	6931	2	七美	I	B	あ	ア	アッ	あっ	感動詞 一般			ア	和
Shoujo	Bokura	2168	6932	2	七美	I	I	ヒヨコ	ヒヨコ	ヒヨコ	雛	名詞 普通名詞 一般			ヒヨコ	和
Shoujo	Bokura	2170	6933	2	髪の毛い子	I	B	カワイイ	カワイー	カワイ	可愛い	形容詞 一般	形容詞	終止形一般	カワイイ	和
Shoujo	Bokura	2171	6934	2	みずちん	I	B	欲しい	ホシイ	ホシイ	欲しい	形容詞 非自立可能	形容詞	終止形一般	ホシイ	和
Shoujo	Bokura	2171	6935	2	みずちん	I	I	ね	ネ	ネ	ね	助詞 終助詞			ネ	和
Shoujo	Bokura	2172	6936	2	七美	I	B	でも	デモ	デモ	でも	接続詞			デモ	和
Shoujo	Bokura	2172	6937	2	七美	I	I	大きい	オオキク	オオキイ	大きい	形容詞 一般	形容詞	連用形一般	オオキイ	和
Shoujo	Bokura	2172	6938	2	七美	I	I	なる	ナル	ナル	成る	動詞 非自立可能	五段 ラ行 一般	終止形一般	ナル	和
Shoujo	Bokura	2172	6939	2	七美	I	I	と	ト	ト	と	助詞 接続助詞			ト	和
Shoujo	Bokura	2172	6940	2	七美	I	I	ニワトリ	ニワトリ	ニワトリ	鶏	名詞 普通名詞 一般			ニワトリ	和
Shoujo	Bokura	2172	6941	2	七美	I	I	に	ニ	ニ	に	助詞 格助詞			ニ	和
Shoujo	Bokura	2172	6942	2	七美	I	I	なる	ナル	ナル	成る	動詞 非自立可能	五段 ラ行 一般	終止形一般	ナル	和
Shoujo	Bokura	2172	6943	2	七美	I	I	と	ト	ト	と	助詞 格助詞			ト	和
Shoujo	Bokura	2172	6944	2	七美	I	I	いう	ユウ	イウ	言う	動詞 一般	五段 ワア行 イウ	連体形一般	イウ	和
Shoujo	Bokura	2172	6945	2	七美	I	I	欠点	ケッテン	ケッテン	欠点	名詞 普通名詞 一般			ケッテン	漢
Shoujo	Bokura	2172	6946	2	七美	I	I	が	ガ	ガ	が	助詞 格助詞			ガ	和
Shoujo	Bokura	2173	6947	2	七美	I	B	ヨーヨー	ヨーヨー	ヨーヨー	ヨーヨー	名詞 固有名詞 人名 一般			ヨーヨー	漢
Shoujo	Bokura	2173	6948	2	七美	I	I	やろー	ヤロー	ヤロー	野郎	名詞 普通名詞 一般			ヤロー	漢

**Sample 5:** Pages from *GinTama* (3:90~91) [raw manga]



## Sample 6: Pages from *GinTama* (3:90~91) [corpus data]

Genre	Series	Data #	Vol.	Page	Frame	Cat.	In-F #	Sex	Speaker	Ch-T	Text+&	Char#1
Shounen-manga	Gintama	3716	3	90	1	4	1	0	0	0	ザン	2
Shounen-manga	Gintama	3717	3	90	3	4	1	0	0	0	ズウン	3
Shounen-manga	Gintama	3718	3	90	4	1	1	1	坂田	!	心配するな&峰打ちだ	10
Shounen-manga	Gintama	3719	3	90	4	1	2	1	坂田	!	まァ これに&懲りたら もう&俺にからむのは&止めるこつたな	30
Shounen-manga	Gintama	3720	3	90	4	1	3	1	土方	!	てめェ さっきから&グーしか出して&ねーじゃねーか&ナメてんのか!!	34
Shounen-manga	Gintama	3721	3	91	1	4	1	0	叫ぶ	!	ギャー---&ギャー---	11
Shounen-manga	Gintama	3722	3	91	1	1	1	1	山崎	!	お互い&妙な上司がいて&大変ですね	17
Shounen-manga	Gintama	3723	3	91	2	1	1	1	山崎	!	一緒に&飲みましょーか	11
Shounen-manga	Gintama	3724	3	91	2	1	2	1	山崎	!	グチを&着にして	8
Shounen-manga	Gintama	3725	3	91	3	2	1	1	志村	!	---父上	5
Shounen-manga	Gintama	3726	3	91	3	2	2	1	志村	!	やっぱり&花見は大勢&でやるもの&ですね	20
Shounen-manga	Gintama	3727	3	91	4	2	1	1	志村	!	こんな&楽しい&お花見は	12
Shounen-manga	Gintama	3728	3	91	5	1	1	1	坂田	!	アレ?&どこどこ?	9

## Sample 7: Pages from *GinTama* (3:90~91) [morpheme analysis data]

Genre	Series	Data #	Mor. #	Sex	Speaker	Type	Bou.	Main form	Phonetic	Root-phon.	Root	POF	Con. C.	F. of C.	Form	E-Type
Shounen	GinTama	3715	22002	1	坂田	!	B	どっ	トツ	トル	取る	動詞一般	五段-ラ行一般	連用形-促音便	トル	和
Shounen	GinTama	3715	22003	1	坂田	!	!	た	タ	タ	た	助動詞	助動詞-タ	連体形一般	タ	和
Shounen	GinTama	3715	22004	1	坂田	!	!	!	!	!	!	補助記号-句点				記号
Shounen	GinTama	3718	22005	1	坂田	!	B	心配	シンバイ	シンバイ	心配	名詞-普通名詞-サ変形状詞可能			シンバイ	漢
Shounen	GinTama	3718	22006	1	坂田	!	!	する	スル	スル	為る	動詞-非自立可能	サ行変格	終止形一般	スル	和
Shounen	GinTama	3718	22007	1	坂田	!	!	な	ナ	ナ	な	助詞-終助詞		命令形	ナ	和
Shounen	GinTama	3718	22008	1	坂田	!	!	峰打ち	ミネウチ	ミネウチ	峰打ち	名詞-普通名詞一般			ミネウチ	和
Shounen	GinTama	3718	22009	1	坂田	!	!	だ	ダ	ダ	だ	助動詞	助動詞-ダ	終止形一般	ダ	和
Shounen	GinTama	3719	22010	1	坂田	!	B	まァ	マー	マー	まー	感動詞一般			マァ	和
Shounen	GinTama	3719	22011	1	坂田	!	!	!	!	!	!	空白				記号
Shounen	GinTama	3719	22012	1	坂田	!	!	これ	コレ	コレ	此れ	代名詞			コレ	和
Shounen	GinTama	3719	22013	1	坂田	!	!	に	ニ	ニ	に	助詞-格助詞			ニ	和
Shounen	GinTama	3719	22014	1	坂田	!	!	懲り	コリル	コリル	懲りる	動詞一般	上一段-ラ行	連用形一般	コリル	和
Shounen	GinTama	3719	22015	1	坂田	!	!	たら	タラ	タ	た	助動詞	助動詞-タ	仮定形一般	タ	和
Shounen	GinTama	3719	22016	1	坂田	!	!	!	!	!	!	空白				記号
Shounen	GinTama	3719	22017	1	坂田	!	!	もう	モウ	モウ	もう	副詞			モウ	和
Shounen	GinTama	3719	22018	1	坂田	!	!	俺	オレ	オレ	俺	代名詞			オレ	和
Shounen	GinTama	3719	22019	1	坂田	!	!	に	ニ	ニ	に	助詞-格助詞			ニ	和
Shounen	GinTama	3719	22020	1	坂田	!	!	からむ	カラム	カラム	絡む	動詞一般	五段-マ行	連体形一般	カラム	和
Shounen	GinTama	3719	22021	1	坂田	!	!	の	ノ	ノ	の	助詞-準体助詞			ノ	和
Shounen	GinTama	3719	22022	1	坂田	!	!	は	ワ	ハ	は	助詞-係助詞			ハ	和
Shounen	GinTama	3719	22023	1	坂田	!	!	止める	トメル	トメル	止める	動詞一般	下一段-マ行	連体形一般	トメル	和
Shounen	GinTama	3719	22024	1	坂田	!	!	こつた	コッタ	コトハ	事は	名詞-普通名詞一般			コトハ	和
Shounen	GinTama	3719	22025	1	坂田	!	!	な	ナ	ナ	な	助詞-終助詞			ナ	和
Shounen	GinTama	3720	22026	1	土方	!	B	てめェ	テメー	テマエ	てまえ	代名詞			テメエ	和
Shounen	GinTama	3720	22027	1	土方	!	!	!	!	!	!	空白				記号
Shounen	GinTama	3720	22028	1	土方	!	!	さっき	サッキ	サキ	先	名詞-普通名詞-副詞可能			サッキ	和
Shounen	GinTama	3720	22029	1	土方	!	!	から	カラ	カラ	から	助詞-格助詞			カラ	和
Shounen	GinTama	3720	22030	1	土方	!	!	グー	グー	グッド	グッド	名詞-普通名詞-形状詞可能			グー	外
Shounen	GinTama	3720	22031	1	土方	!	!	しか	シカ	シカ	しか	助詞-副助詞			シカ	和
Shounen	GinTama	3720	22032	1	土方	!	!	出し	ダシ	ダス	出す	動詞-非自立可能	五段-サ行	連用形一般	ダス	和
Shounen	GinTama	3720	22033	1	土方	!	!	て	テル	テル	てる	助動詞	下一段-タ行	未然形一般	テル	和
Shounen	GinTama	3720	22034	1	土方	!	!	ねー	ネー	ナイ	ない	助動詞	助動詞-ナイ	終止形一般	ナイ	和
Shounen	GinTama	3720	22035	1	土方	!	!	じゃ	ジャ	ダ	だ	助動詞	助動詞-ダ	連用形-融合	ダ	和
Shounen	GinTama	3720	22036	1	土方	!	!	ねー	ネー	ナイ	無い	形容詞-非自立可能	形容詞	終止形一般	ナイ	和
Shounen	GinTama	3720	22037	1	土方	!	!	か	カ	カ	か	助詞-終助詞			カ	和
Shounen	GinTama	3720	22038	1	土方	!	!	ナメ	ナメ	ナメル	當める	動詞一般	下一段-マ行	連用形一般	ナメル	和
Shounen	GinTama	3720	22039	1	土方	!	!	てん	テン	テル	てる	助動詞	下一段-タ行	連体形-撥音便	テル	和
Shounen	GinTama	3720	22040	1	土方	!	!	の	ノ	ノ	の	助詞-終助詞			ノ	和
Shounen	GinTama	3720	22041	1	土方	!	!	か	カ	カ	か	助詞-終助詞			カ	和
Shounen	GinTama	3720	22042	1	土方	!	!	!	!	!	!	補助記号-句点				記号
Shounen	GinTama	3722	22043	1	山崎	!	B	お	オ	オ	御	接頭辞			オ	和
Shounen	GinTama	3722	22044	1	山崎	!	!	互い	タガイ	タガイ	互い	名詞-普通名詞一般			タガイ	和
Shounen	GinTama	3722	22045	1	山崎	!	!	妙	ミョウ	ミョウ	妙	名詞-普通名詞-形状詞可能			ミョウ	漢
Shounen	GinTama	3722	22046	1	山崎	!	!	な	ナ	ダ	だ	助動詞	助動詞-ダ	連体形一般	ダ	和
Shounen	GinTama	3722	22047	1	山崎	!	!	上司	ジョーシ	ジョウシ	上司	名詞-普通名詞一般			ジョウシ	漢
Shounen	GinTama	3722	22048	1	山崎	!	!	が	ガ	ガ	が	助詞-格助詞			ガ	和
Shounen	GinTama	3722	22049	1	山崎	!	!	い	イル	イル	居る	動詞-非自立可能	上一段-ア行	連用形一般	イル	和
Shounen	GinTama	3722	22050	1	山崎	!	!	て	テ	テ	て	助詞-接続助詞			テ	和
Shounen	GinTama	3722	22051	1	山崎	!	!	大変	タイヘン	タイヘン	大変	形状詞一般			タイヘン	漢
Shounen	GinTama	3722	22052	1	山崎	!	!	です	デス	デス	です	助動詞	助動詞-デス	終止形一般	デス	和
Shounen	GinTama	3722	22053	1	山崎	!	!	ね	ネ	ネ	ね	助詞-終助詞			ネ	和
Shounen	GinTama	3723	22054	1	山崎	!	B	一緒	イッショ	イッショ	一緒	名詞-普通名詞-サ変可能			イッショ	漢
Shounen	GinTama	3723	22055	1	山崎	!	!	に	ニ	ニ	に	助詞-格助詞			ニ	和
Shounen	GinTama	3723	22056	1	山崎	!	!	飲み	ノミ	ノム	飲む	動詞一般	五段-マ行	連用形一般	ノム	和
Shounen	GinTama	3723	22057	1	山崎	!	!	ましょー	マショー	マス	ます	助動詞	助動詞-マス	意志推量形	マス	和
Shounen	GinTama	3723	22058	1	山崎	!	!	か	カ	カ	か	助詞-終助詞			カ	和
Shounen	GinTama	3724	22059	1	山崎	!	B	グチ	グチ	グチ	愚痴	名詞-普通名詞一般			グチ	漢
Shounen	GinTama	3724	22060	1	山崎	!	!	を	オ	ヲ	を	助詞-格助詞			ヲ	和
Shounen	GinTama	3724	22061	1	山崎	!	!	肴	サカナ	サカナ	肴	名詞-普通名詞一般			サカナ	和
Shounen	GinTama	3724	22062	1	山崎	!	!	に	ニ	ニ	に	助詞-格助詞			ニ	和
Shounen	GinTama	3724	22063	1	山崎	!	!	し	シ	スル	為る	動詞-非自立可能	サ行変格	連用形一般	スル	和
Shounen	GinTama	3724	22064	1	山崎	!	!	て	テ	テ	て	助詞-接続助詞			テ	和
Shounen	GinTama	3728	22065	1	坂田	!	B	アレ	アレ	アレ	あれ	感動詞一般			アレ	和
Shounen	GinTama	3728	22066	1	坂田	!	!	?	?	?	?	補助記号-句点				記号
Shounen	GinTama	3728	22067	1	坂田	!	!	ここ	ココ	ココ	此处	代名詞			ココ	和
Shounen	GinTama	3728	22068	1	坂田	!	!	どこ	ドコ	ドコ	何処	代名詞			ドコ	和
Shounen	GinTama	3728	22069	1	坂田	!	!	?	?	?	?	補助記号-句点				記号

**Sample 8:** Pages from *RabuKon* (1:18~19) [raw manga]



### Sample 9: Pages from *RabuKon* (1:18~19) [corpus data]

Genre	Series	Data #	Vol.	Page	Frame	Cat.	In-F #	Sex	Speaker	Ch-T	Text+&	Char#1
Shoujo-manga	RabuKon	185	1	18	1	1	1	2	小泉	I	あ あ&&は&&ハイ・・・・・	16
Shoujo-manga	RabuKon	186	1	18	1	4	1	0	0	0	カタン	3
Shoujo-manga	RabuKon	187	1	18	2	1	1	1	大谷	I	小泉の後ろ&なんか座ったら&黒板見えへんで	21
Shoujo-manga	RabuKon	188	1	18	2	1	2	2	小泉	I	うるさいね&大谷！！	11
Shoujo-manga	RabuKon	189	1	18	3	1	1	2	小泉	I	す・・・・・&&すんません&でかくて・・・・・	25
Shoujo-manga	RabuKon	190	1	18	3	4	1	0	笑う	0	やはは	3
Shoujo-manga	RabuKon	191	1	18	3	4	1	0	0	0	キィ	2
Shoujo-manga	RabuKon	192	1	18	4	1	1	1	鈴木	I	大丈夫&&見えるから	10
Shoujo-manga	RabuKon	193	1	18	5	4	1	0	0	0	さくっ	3
Shoujo-manga	RabuKon	194	1	18	6	4	1	0	0	0	くる	2
Shoujo-manga	RabuKon	195	1	18	6	5	1	0	0	0	L e t o u t & m y p e n t u p & f r u s t	24
Shoujo-manga	RabuKon	196	1	18	7	2	1	2	小泉	I	好き！！	4
Shoujo-manga	RabuKon	197	1	18	7	7	1	3	不明語り手	0	早い	2
Shoujo-manga	RabuKon	198	1	18	7	2	2	2	小泉	I	きつと&あたしより&背高いん&やわ	17
Shoujo-manga	RabuKon	199	1	18	7	5	1	0	0	0	u s t & o n	6
Shoujo-manga	RabuKon	200	1	19	1	1	1	1	先生	I	5組一&&鈴木一	8
Shoujo-manga	RabuKon	201	1	19	1	1	2	1	鈴木	I	はい	2
Shoujo-manga	RabuKon	202	1	19	1	2	1	2	小泉	I	5組の&鈴木くん&かあ・・・・・	17
Shoujo-manga	RabuKon	203	1	19	1	1	3	1	大谷	I	・・・・・	9
Shoujo-manga	RabuKon	204	1	19	2	4	1	0	0	0	ニヤリ	3
Shoujo-manga	RabuKon	205	1	19	3	2	1	2	小泉	I	・・・・・&・・・・・	31
Shoujo-manga	RabuKon	206	1	19	4	2	1	2	小泉	I	ええわ〜〜	6
Shoujo-manga	RabuKon	207	1	19	4	1	1	2	のぶちゃん	I	どないしたん&リサ	9
Shoujo-manga	RabuKon	208	1	19	4	1	2	2	千春	I	さあ	2
Shoujo-manga	RabuKon	209	1	19	4	4	1	0	0	0	くわ	2
Shoujo-manga	RabuKon	210	1	19	5	4	1	0	0	0	タン	2
Shoujo-manga	RabuKon	211	1	19	5	1	1	1	大谷	I	よーよー&姉ちゃん	9
Shoujo-manga	RabuKon	212	1	19	6	1	1	1	大谷	I	ちょっと話&&あんねん	11
Shoujo-manga	RabuKon	213	1	19	6	1	1	2	小泉	I	帰り&顔貸せや	7
Shoujo-manga	RabuKon	214	1	20	1	1	1	2	小泉	I	ちびっこ&ギャングか？	11
Shoujo-manga	RabuKon	215	1	20	1	6	1	2	小泉	I	迫力ね〜〜	6
Shoujo-manga	RabuKon	216	1	20	1	1	2	1	大谷	I	やかましわ&&立った	10



# Sample 10: Pages from *RabuKon* (1:18~19) [morpheme analysis data]

Genre	Series	Data #	Mor. #	Sex	Speaker	Type	Bou.	Main form	Phonetic	Root-phon.	Root	POF	Con. C.	F. of C.	Form	E-Type
Shoujo	RabuKon	185	756	2	小泉	I	B	あ	ア	アツ	あっ	感動詞一般			ア	和
Shoujo	RabuKon	185	757	2	小泉	I	I					空白				記号
Shoujo	RabuKon	185	758	2	小泉	I	I	あ	ア	アツ	あっ	感動詞一般			ア	和
Shoujo	RabuKon	185	759	2	小泉	I	I	は	ハ	ハ	は	感動詞-カム			ハ	他
Shoujo	RabuKon	185	760	2	小泉	I	I	ハイ	ハイ	ハイ	はい	感動詞一般			ハイ	和
Shoujo	RabuKon	185	761	2	小泉	I	I	・・・			・・・	補助記号一般				記号
Shoujo	RabuKon	187	762	1	大谷	I	B	小泉	コイズミ	コイズミ	コイズミ	名詞-固有名詞-人名-姓			コイズミ	固
Shoujo	RabuKon	187	763	1	大谷	I	I	の	ノ	ノ	の	助詞-格助詞			ノ	和
Shoujo	RabuKon	187	764	1	大谷	I	I	後ろ	ウシロ	ウシロ	後ろ	名詞-普通名詞一般			ウシロ	和
Shoujo	RabuKon	187	765	1	大谷	I	I	なんか	ナンカ	ナンカ	なんか	助詞-副助詞			ナンカ	和
Shoujo	RabuKon	187	766	1	大谷	I	I	座っ	スワツ	スワル	座る	動詞一般	五段-ラ行一般	連用形-促音便	スワル	和
Shoujo	RabuKon	187	767	1	大谷	I	I	たら	タラ	タ	た	助動詞	助動詞-タ	仮定形一般	タ	和
Shoujo	RabuKon	187	768	1	大谷	I	I	黒板	コクバン	コクバン	黒板	名詞-普通名詞一般			コクバン	漢
Shoujo	RabuKon	187	769	1	大谷	I	I	見え	ミエ	ミエル	見える	動詞一般	下一段-ア行	未然形一般	ミエル	和
Shoujo	RabuKon	187	770	1	大谷	I	I	へん	ヘン	ヘン	へん	助動詞	助動詞-ヘン	終止形一般	ヘン	和
Shoujo	RabuKon	187	771	1	大谷	I	I	で	デ	デ	で	助詞-終助詞			デ	和
Shoujo	RabuKon	188	772	2	小泉	I	B	うるさい	ウルサイ	ウルサイ	煩い	形容詞一般	形容詞	終止形一般	ウルサイ	和
Shoujo	RabuKon	188	773	2	小泉	I	I	ねん	ネン	ネン	ねん	助詞-終助詞			ネン	和
Shoujo	RabuKon	188	774	2	小泉	I	I	大谷	オオタニ	オオタニ	オオタニ	名詞-固有名詞-人名-姓			オオタニ	固
Shoujo	RabuKon	188	775	2	小泉	I	I	!			!	補助記号-句点				記号
Shoujo	RabuKon	189	776	2	小泉	I	B	す	ス	ス	す	感動詞-カム			ス	他
Shoujo	RabuKon	189	777	2	小泉	I	I	・・・			・・・	補助記号一般				記号
Shoujo	RabuKon	189	778	2	小泉	I	I	すん	スン	スム	済む	動詞一般	五段-マ行	連用形-撥音便	スム	和
Shoujo	RabuKon	189	779	2	小泉	I	I	ませ	マセ	マス	ます	助動詞	助動詞-マス	未然形一般	マス	和
Shoujo	RabuKon	189	780	2	小泉	I	I	ん	ン	ズ	ず	助動詞	助動詞-ヌ	終止形-撥音便	ヌ	和
Shoujo	RabuKon	189	781	2	小泉	I	I	でかく	デカク	デカイ	でかい	形容詞一般	形容詞	連用形一般	デカイ	不明
Shoujo	RabuKon	189	782	2	小泉	I	I	て	テ	テ	て	助詞-接続助詞			テ	和
Shoujo	RabuKon	189	783	2	小泉	I	I	・・・			・・・	補助記号一般				記号
Shoujo	RabuKon	192	784	1	鈴木	I	B	大丈夫	ダイジョーブ	ダイジョーブ	大丈夫	形状詞一般			ダイジョーブ	漢
Shoujo	RabuKon	192	785	1	鈴木	I	I	見える	ミエル	ミエル	見える	動詞一般	下一段-ア行	終止形一般	ミエル	和
Shoujo	RabuKon	192	786	1	鈴木	I	I	から	カラ	カラ	から	助詞-接続助詞			カラ	和
Shoujo	RabuKon	200	787	1	先生	I	B	5	ゴ	ゴ	五	名詞-数詞			ゴ	漢
Shoujo	RabuKon	200	788	1	先生	I	I	組	グミ	クミ	組	名詞-普通名詞一般			クミ	和
Shoujo	RabuKon	200	789	1	先生	I	I	鈴木	スズキ	スズキ	スズキ	名詞-固有名詞-人名-姓			スズキ	固
Shoujo	RabuKon	201	790	1	鈴木	I	B	はい	ハイ	ハイ	はい	感動詞一般			ハイ	和
Shoujo	RabuKon	203	791	1	大谷	I	B	・・・			・・・	補助記号一般				記号
Shoujo	RabuKon	207	792	2	のぶちゃん	I	B	どない	ドナイ	ドナイ	どない	形状詞一般			ドナイ	和
Shoujo	RabuKon	207	793	2	のぶちゃん	I	I	し	シ	スル	为る	動詞-非自立可能	サ行変格	連用形一般	スル	和
Shoujo	RabuKon	207	794	2	のぶちゃん	I	I	た	タ	タ	た	助動詞	助動詞-タ	連体形一般	タ	和
Shoujo	RabuKon	207	795	2	のぶちゃん	I	I	ん	ン	ノ	の	助詞-終助詞			ン	和
Shoujo	RabuKon	207	796	2	のぶちゃん	I	I	リサ	リサ	リサ	リサ	名詞-固有名詞-人名-名			リサ	固
Shoujo	RabuKon	208	797	2	千春	I	B	さあ	サー	サア	さあ	感動詞一般			サア	和
Shoujo	RabuKon	211	798	1	大谷	I	B	よー	ヨー	ヨー	よー	感動詞一般			ヨー	和
Shoujo	RabuKon	211	799	1	大谷	I	I	よー	ヨー	ヨー	よー	感動詞一般			ヨー	和
Shoujo	RabuKon	211	800	1	大谷	I	I	姉	ネー	ネエ	姉	名詞-普通名詞一般			ネエ	和
Shoujo	RabuKon	211	801	1	大谷	I	I	ちゃん	チャン	チャン	ちゃん	接尾辞-名詞一般			チャン	和
Shoujo	RabuKon	212	802	1	大谷	I	B	ちよつと	チョット	チョット	一寸	副詞			チョット	和
Shoujo	RabuKon	212	803	1	大谷	I	I	話	ハナシ	ハナシ	話	名詞-普通名詞-サ変可能			ハナシ	和
Shoujo	RabuKon	212	804	1	大谷	I	I	が	ガ	ガ	が	助詞-格助詞			ガ	和
Shoujo	RabuKon	212	805	1	大谷	I	I	あん	アン	アル	有る	動詞-非自立可能	五段-ラ行一般	連体形-撥音便	アル	和
Shoujo	RabuKon	212	806	1	大谷	I	I	ねん	ネン	ネン	ねん	助詞-終助詞			ネン	和
Shoujo	RabuKon	213	807	2	小泉	I	B	帰り	カエリ	カエリ	帰り	名詞-普通名詞一般			カエリ	和
Shoujo	RabuKon	213	808	2	小泉	I	I	顔	ツラ	ツラ	顔	名詞-普通名詞一般			ツラ	和
Shoujo	RabuKon	213	809	2	小泉	I	I	貸せ	カセ	カス	貸す	動詞一般	五段-サ行	命令形	カス	和
Shoujo	RabuKon	213	810	2	小泉	I	I	や	ヤ	ヤ	や	助詞-終助詞			ヤ	和
Shoujo	RabuKon	214	811	2	小泉	I	B	ちびっこ	チビッコ	チビッコ	ちびっこ	名詞-普通名詞一般			チビッコ	和
Shoujo	RabuKon	214	812	2	小泉	I	I	ギャング	ギャング	ギャング	ギャング	名詞-普通名詞一般			ギャング	外
Shoujo	RabuKon	214	813	2	小泉	I	I	か	カ	カ	か	助詞-終助詞			カ	和
Shoujo	RabuKon	214	814	2	小泉	I	I	?			?	補助記号-句点				記号
Shoujo	RabuKon	216	815	1	大谷	I	B	やかまし	ヤカマシ	ヤカマシイ	喧しい	形容詞一般	形容詞	語幹一般	ヤカマシイ	和
Shoujo	RabuKon	216	816	1	大谷	I	I	わ	ワ	ワ	わ	助詞-終助詞			ワ	和
Shoujo	RabuKon	216	817	1	大谷	I	I	立つ	タツ	タツ	立つ	動詞一般	五段-タ行	連用形-促音便	タツ	和
Shoujo	RabuKon	216	818	1	大谷	I	I	た	タ	タ	た	助動詞	助動詞-タ	終止形一般	タ	和



**Sample 11:** Pages from *Naruto* (3:82~83) [raw manga]



# Sample 12: Pages from *Naruto* (3:82~83) [corpus data]

Genre	Series	Data #	Vol.	Page	Frame	Cat.	In-F #	Sex	Speaker	Ch-T	Text+&	Char#1
Shounen-manga	Naruto	4254	3	82	1	1	1	1	ナルト	1	な——んちやってエ——！！	16
Shounen-manga	Naruto	4255	3	82	1	4	1	0	0	0	ビタッ！	4
Shounen-manga	Naruto	4256	3	82	1	4	2	0	0	0	ニツ！	4
Shounen-manga	Naruto	4257	3	82	3	1	1	1	ナルト	1	ハッハー——！！&ひっかった！！&ひっかった！！	23
Shounen-manga	Naruto	4258	3	82	3	1	2	1	カカシ	1	……	6
Shounen-manga	Naruto	4259	3	82	3	1	3	2	サクラ	1	……	6
Shounen-manga	Naruto	4260	3	82	3	6	1	1	ナルト	1	ハッハー！	5
Shounen-manga	Naruto	4261	3	82	4	1	1	2	サクラ	1	びっくりする&じゃない バカ！	15
Shounen-manga	Naruto	4262	3	82	4	7	1	3	不明語り手	0	内なる&サクラ	7
Shounen-manga	Naruto	4263	3	82	5	2	2	2	サクラ	1	あとで殺す！	6
Shounen-manga	Naruto	4264	3	82	5	6	1	2	サクラ	1	しゃーんなろ！	7
Shounen-manga	Naruto	4265	3	82	7	1	1	1	ナルト	1	あ！	2
Shounen-manga	Naruto	4266	3	82	7	4	1	0	0	0	フツ…	4
Shounen-manga	Naruto	4267	3	82	8	1	1	1	カカシ	1	って…&あ！	9
Shounen-manga	Naruto	4268	3	82	9	1	1	2	サクラ	1	あ`あー！！	7
Shounen-manga	Naruto	4269	3	82	9	1	2	2	サクラ	1	ナルトのバカ！！&&調子ぶっこいてる&からよ！	24
Shounen-manga	Naruto	4270	3	82	9	4	1	0	0	0	ガン	3
Shounen-manga	Naruto	4271	3	83	1	1	1	1	ナルト	1	あ`あ`あ`&&ぎゃあああ&&あ！	18
Shounen-manga	Naruto	4272	3	83	2	4	1	0	0	0	ガッ	2
Shounen-manga	Naruto	4273	3	83	3	4	1	0	0	0	ガッ！	3
Shounen-manga	Naruto	4274	3	83	4	1	2	1	サスケ	1	この&ウスランカチが…	14
Shounen-manga	Naruto	4275	3	83	4	4	1	0	0	0	ドンッ	3
Shounen-manga	Naruto	4276	3	83	5	1	1	1	サスケ	1	サスケェ！！	6
Shounen-manga	Naruto	4277	3	83	6	1	1	2	サクラ	1	キヤ——&&さすが サスケ君！&しびれるウー！	24
Shounen-manga	Naruto	4278	3	83	7	1	1	1	カカシ	1	……	9
Shounen-manga	Naruto	4279	3	83	7	2	1	1	カカシ	1	こいつら&よく成長して&やがる…	18

### Sample 13: Pages from *Naruto* (3:82~83) [morpheme analysis data]

Genre	Series	Data #	Mor. #	Sex	Speaker	Type	Bou.	Main form	Phonetic	Root-phon.	Root	POF	Con. C.	F. of C.	Form	E-Type
Shounen	Naruto	4254	16388	1	ナルト	I	B	なん	ナン	ナニ	何	代名詞			ナン	和
Shounen	Naruto	4254	16389	1	ナルト	I	I	ちやつ	チャツ	テガウ	違う	動詞-一般	五段-フア行-一般	連用形-促音便	チャウ	和
Shounen	Naruto	4254	16390	1	ナルト	I	I	て	テ	テ	て	助詞-接続助詞			テ	和
Shounen	Naruto	4254	16391	1	ナルト	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4257	16392	1	ナルト	I	B	ハツハー	ハツハー	ハハ	はは	感動詞-一般			ハツハー	和
Shounen	Naruto	4257	16393	1	ナルト	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4257	16394	1	ナルト	I	I	ひっかかつ	ヒツカカツ	ヒツカカル	引っ掛かる	動詞-一般	五段-ラ行-一般	連用形-促音便	ヒツカカル	和
Shounen	Naruto	4257	16395	1	ナルト	I	I	た	タ	タ	た	助動詞	助動詞-タ	終止形-一般	タ	和
Shounen	Naruto	4257	16396	1	ナルト	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4257	16397	1	ナルト	I	I	ひっかかつ	ヒツカカツ	ヒツカカル	引っ掛かる	動詞-一般	五段-ラ行-一般	連用形-促音便	ヒツカカル	和
Shounen	Naruto	4257	16398	1	ナルト	I	I	た	タ	タ	た	助動詞	助動詞-タ	終止形-一般	タ	和
Shounen	Naruto	4257	16399	1	ナルト	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4258	16400	1	カカシ	I	B	...			...	補助記号-一般				記号
Shounen	Naruto	4259	16401	2	サクラ	I	B	...			...	補助記号-一般				記号
Shounen	Naruto	4261	16402	2	サクラ	I	B	びっくり	ビツクリ	ビツクリ	びっくり	名詞-普通名詞-サ変可能			ビツクリ	和
Shounen	Naruto	4261	16403	2	サクラ	I	I	する	スル	スル	ある	動詞-非自立可能	サ行変格	終止形-一般	スル	和
Shounen	Naruto	4261	16404	2	サクラ	I	I	じゃ	ジャ	ダ	だ	助動詞	助動詞-ダ	連用形-融合	ダ	和
Shounen	Naruto	4261	16405	2	サクラ	I	I	ない	ナイ	ナイ	無い	形容詞-非自立可能	形容詞	終止形-一般	ナイ	和
Shounen	Naruto	4261	16406	2	サクラ	I	I					空白				記号
Shounen	Naruto	4261	16407	2	サクラ	I	I	バカ	バカ	バカ	馬鹿	名詞-普通名詞-形状詞可能			バカ	不明
Shounen	Naruto	4261	16408	2	サクラ	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4265	16409	1	ナルト	I	B	あ	ア	アツ	あつ	感動詞-一般			ア	和
Shounen	Naruto	4265	16410	1	ナルト	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4267	16411	1	カカシ	I	B	って	ツテ	ツテ	って	感動詞-一般			ツテ	和
Shounen	Naruto	4267	16412	1	カカシ	I	I	...			...	補助記号-一般				記号
Shounen	Naruto	4267	16413	1	カカシ	I	I	あ`	ア`	ア`ア`	あ`あ`	感動詞-一般			ア`	和
Shounen	Naruto	4267	16414	1	カカシ	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4268	16415	2	サクラ	I	B	あ`あ`ー	ア`ア`ー	ア`ア`	あ`あ`	感動詞-一般			ア`ア`ー	和
Shounen	Naruto	4268	16416	2	サクラ	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4269	16417	2	サクラ	I	B	ナルト	ナルト	ナルト	ナルト	名詞-固有名詞-人名-名			ナルト	固
Shounen	Naruto	4269	16418	2	サクラ	I	I	の	ノ	ノ	の	助詞-格助詞			ノ	和
Shounen	Naruto	4269	16419	2	サクラ	I	I	バカ	バカ	バカ	馬鹿	名詞-普通名詞-形状詞可能			バカ	不明
Shounen	Naruto	4269	16420	2	サクラ	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4269	16421	2	サクラ	I	I	調子	チョウシ	チョウシ	調子	名詞-普通名詞-一般			チョウシ	漢
Shounen	Naruto	4269	16422	2	サクラ	I	I	ぶっ	ブツ	ブツ	打つ	動詞-一般	五段-タ行	連用形-促音便	ブツ	和
Shounen	Naruto	4269	16423	2	サクラ	I	I	こい	コイ	コク	こく	動詞-一般	五段-カ行-一般	連用形-イ音便	コク	和
Shounen	Naruto	4269	16424	2	サクラ	I	I	てる	テル	テル	てる	助動詞	下一段-タ行	終止形-一般	テル	和
Shounen	Naruto	4269	16425	2	サクラ	I	I	から	カラ	カラ	から	助詞-接続助詞			カラ	和
Shounen	Naruto	4269	16426	2	サクラ	I	I	よ	ヨ	ヨ	よ	助詞-終助詞			ヨウ	和
Shounen	Naruto	4269	16427	2	サクラ	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4271	16428	1	ナルト	I	B	あ`あ`あ`	ア`ア`ア`	ア`ア`	あ`あ`	感動詞-一般			ア`ア`ア`	和
Shounen	Naruto	4271	16429	1	ナルト	I	I	ぎやああああ	ギヤアアアア	ギヤア	ギヤア	感動詞-一般			ギヤアアアア	和
Shounen	Naruto	4271	16430	1	ナルト	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4274	16431	1	サスケ	I	B	この	コノ	コノ	此の	連体詞			コノ	和
Shounen	Naruto	4274	16432	1	サスケ	I	I	ウストランカチ	ウストランカチ	ウストランカチ	薄らとんかち	名詞-普通名詞-一般			ウストランカチ	和
Shounen	Naruto	4274	16433	1	サスケ	I	I	が	ガ	ガ	が	助詞-格助詞			ガ	和
Shounen	Naruto	4274	16434	1	サスケ	I	I	...			...	補助記号-一般				記号
Shounen	Naruto	4276	16435	1	サスケ	I	B	サスケ	サスケ	サスケ	サスケ	名詞-固有名詞-人名-名			サスケ	固
Shounen	Naruto	4276	16436	1	サスケ	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4277	16437	2	サクラ	I	B	キヤア	キヤア	キヤア	キヤア	感動詞-一般			キヤア	和
Shounen	Naruto	4277	16438	2	サクラ	I	I	さすが	サスガ	サスガ	流石	副詞			サスガ	和
Shounen	Naruto	4277	16439	2	サクラ	I	I	!			!	空白				記号
Shounen	Naruto	4277	16440	2	サクラ	I	I	サスケ	サスケ	サスケ	サスケ	名詞-固有名詞-人名-名			サスケ	固
Shounen	Naruto	4277	16441	2	サクラ	I	I	君	クン	クン	君	接尾辞-名詞的-一般			クン	漢
Shounen	Naruto	4277	16442	2	サクラ	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4277	16443	2	サクラ	I	I	しびれる	シビレル	シビレル	痺れる	動詞-一般	下一段-ラ行-一般	連体形-一般	シビレル	和
Shounen	Naruto	4277	16444	2	サクラ	I	I	!			!	補助記号-句点				記号
Shounen	Naruto	4278	16445	1	カカシ	I	B	...			...	補助記号-一般				記号

## 6.2 Gloss

Word	Definition
<i>1PP</i>	See <i>first person pronoun</i>
<i>2PP</i>	See <i>second person pronoun</i>
<i>3PP</i>	See <i>third person pronoun</i>
<i>4-koma manga</i>	4-frame gag manga; while series may continue for a long time, they are not plot-driven, and feature minimal character or narrative development
<i>ateji</i>	kanji used only for phonetic purposes, regardless of their meanings, or kanji used with non-established, irregular readings
<i>bousen</i>	a Japanese orthographic symbol used with katakana that shows that a vowel is long (ー)
<i>chou-bousen</i>	a Japanese orthographic symbol found in comics that consists of a long line, demonstrating either long vowels, or pauses. Registered in the corpus as — — —
<i>chou-namisen</i>	a Japanese orthographic symbol found in comics that consists of a long squiggly line, demonstrating either long vowels or pauses. Registered in the corpus as ~ ~ ~
<i>dakuon-ten</i>	a Japanese orthographic symbol used with the kana that shows that a syllable is voiced (゛)
<i>emoji</i>	literally <i>picture letters</i> ; visual, non-phonetic symbols used within computerized type, such as happy faces
<i>e-monogatari</i>	literally <i>picture stories</i> ; a form of narrative popular in the early 20th century that used equal parts pictures and text, with the text less integrated into the pictures than in manga
<i>entry</i>	one line of text data in the corpus, defined as any block of text, such as text within an individual speech bubble or a string of onomatopoeic expressions
<i>first person pronoun</i>	personal pronouns referring to the speaker ( <i>I</i> or <i>we</i> in English). Japanese allows for multiple forms, commonly differentiated by the speaker's sex and relationship with their listener(s), such as <i>watashi</i> , <i>ore</i> and <i>atashi</i>
<i>furigana</i>	glosses in the margins of Japanese text, usually giving the phonetic readings of kanji
<i>fuushuu manga</i>	political satire manga
<i>gairaigo</i>	words of foreign origin

<i>genbun'icchi</i>	the unification of spoken and written language; a movement during the Meiji period to reform written language so that it would be more similar to spoken language
<i>hentai shoujo-moji</i>	extra-round forms of script that became popular amongst young girls in the 1980s
<i>hiragana</i>	along with katakana, one of the two Japanese syllabaries. Mostly used for function words and words of Japanese origin
<i>hyoujungo</i>	standard Japanese language
<i>jidou-manga</i>	manga for young children
<i>jinmeiyou-kanji</i>	kanji for personal names; an approved list of 861 kanji by the Ministry of Education for use in personal and place names only
<i>JMPA</i>	Japanese Magazine Publisher Association
<i>josei-muke komikkushi</i>	comic magazines aimed for older female readers. See <i>redisu-manga</i>
<i>jouyou-kanji</i>	kanji for everyday use; an approved list of 2,136 kanji by the Ministry of Education specifying what kanji are necessary for everyday use of Japanese
<i>kana</i>	the two syllabaries of Japanese: <i>hiragana</i> and <i>katakana</i>
<i>kanbun</i>	Japanese texts which would be written in Chinese, grammatically, with rules to transform them into Japanese texts when read aloud
<i>kango</i>	words of Chinese origin
<i>kanji</i>	Chinese characters, which usually have at least one kun and one on reading
<i>Kansai-ben</i>	the dialect spoken in the Kansai region, including Osaka, Kobe and Kyoto
<i>katakana</i>	along with hiragana, one of the two Japanese syllabaries. Mostly used for foreign words, onomatopoeia and other sound words
<i>keitai-meeru</i>	e-mails written on cell-phones
<i>keitai-shousetsu</i>	novels written on and for cell-phones
<i>komikku-zasshi</i>	comic magazines, generally phonebook size, containing 10 or more stories by different authors

<i>kouhai</i>	a lower classmate
<i>kun readings</i>	readings for kanji that have a Japanese origin
<i>kuten</i>	the Japanese period ( 〰 )
<i>kyouiku-kanji</i>	the 1,006 kanji taught during the first six years of obligatory education in elementary school
<i>kyouiku-manga</i>	manga written for educational purposes
<i>manga-ka</i>	a manga-writer.
<i>nakaguro</i>	a Japanese orthographic symbol used to connect two words, or to show word boundaries ( ・ )
<i>namisen</i>	a Japanese orthographic symbol generally used to mean <i>from - to</i> , similar to the tilde; also used in comics to write long vowels ( ～ )
<i>NINJAL</i>	National Institute for Japanese Language and Linguistics
<i>odori-ji</i>	a Japanese orthographic symbol used to repeat the previous kanji ( 々 )
<i>okama</i>	a stereotypical, transsexual or gay man
<i>on readings</i>	readings for kanji that have a Chinese origin
<i>onomatopoeia</i>	a mimetic or imitative word that express sounds or which use sound symbolism to express moods, atmospheres or actions
<i>personal pronoun</i>	the class of pronouns used to refer to the speaker, or to whom or about whom the speaker is referring to ( <i>I, you, they</i> )
<i>PP</i>	see <i>personal pronoun</i>
<i>rediisu-manga</i>	literally <i>ladies-manga</i> , or manga for adult women
<i>roomaji</i>	roman alphabet letters
<i>second person pronoun</i>	personal pronouns referring to whom the speaker is talking to ( <i>you</i> in English). Japanese allows for multiple forms, commonly differentiated by the speaker's sex and relationship with their listener(s), such as <i>anata</i> or <i>omae</i>

<i>seinen-manga</i>	manga for adult men
<i>senpai</i>	an upper classmate, or someone who is not one's boss, but is one's senior at work
<i>sentence final expressions</i>	expressions seen at the end of sentences, including SFPs and verbal and adjectival forms
<i>sentence final particles</i>	non-obligatory particles which attach to the ends of sentences or phrases to add emphasis. SFPs
<i>SFE</i>	see <i>sentence final expressions</i>
<i>SFP</i>	see <i>sentence final particles</i>
<i>shoujo-manga</i>	manga for young girls, generally of late elementary to junior high school age
<i>shoujo-muke komikkushi</i>	comic magazines aimed for young female readers. See <i>shoujo-manga</i>
<i>shounen-manga</i>	manga for young boys, generally of late elementary to junior high school age
<i>shounen-muke komikkushi</i>	comic magazines aimed for young male readers. See <i>shounen-manga</i>
<i>shin-genbun'icchi</i>	"The new unification of spoken and written language", or a new style of text described by Satake (1980) that uses more symbols and is more emotionally expressive than standard texts. See <i>genbun'icchi</i>
<i>sutoorii-manga</i>	manga which are story-based, with narrative-driven plots which develop over time; compare with <i>4koma-manga</i>
<i>tankoubon</i>	A5-sized books, which collect 3 to 5 chapters of a series or different short stories by an individual author after they have been published in <i>komikku-zasshi</i>
<i>third person pronoun</i>	personal pronouns referring to whom or what the speaker is talking about ( <i>he, she, it, or they</i> in English)
<i>touten</i>	the Japanese comma (、)
<i>wago</i>	words of Japanese origin

## 6.3 Previous research

- Unser-Schutz, G. (2009, October). Developing a text-based corpus of the language of Japanese comics (manga). *American Association for Corpus Linguistics*. Edmonton, Canada: University of Alberta. [Conference paper]
- Unser-Schutz, G. (2010, March). Exploring the role of language in manga: text types, their usages and their distributions. *Popular Culture Association*. St. Louis, United States of America: Renaissance Grand Hotel St. Louis. [Conference paper]
- Unser-Schutz, G. (2010). Exploring the role of language in manga: Text types, their usages and their distributions. *The International Journal of Comic Art*, 12(2/3), 25-43.
- Unser-Schutz, G. (2010, September). Girls, boys and manga: Sentence final particles in Japanese comics for girls and boys. *International Gender and Language Association*. Tokyo, Japan: Tsuda Juku University. [Conference paper]
- Unser-Schutz, G. (2010). Girls, boys and manga: Sentence final particles in Japanese comics for girls and boys. *International Gender and Language Association 2010 Conference Proceedings* (pp. 288-306). International Gender and Language Association.
- Unser-Schutz, G. (2010, November). Manga as a linguistic resource for learning. *Japanese Association of Language Teachers*. Nagoya, Japan: Aichi Industry and Labor Center. [Conference paper]
- Unser-Schutz, G. (2010). Ninki manga koopasu de miru moji hyougen no bunrui ni tsuite: Manga no gengoteki tokuchou wo kousatsu suru. *Nihon Manga Gakkai 2010 Taikai Youshishuu*, p. 23. [In Japanese] [Conference paper]
- Unser-Schutz, G. (2010, January). Personal pronouns and gendered speech in popular manga (Japanese comics). *Linguistic Society of America*. Baltimore, United States of America: Hilton Baltimore. [Conference paper]
- Unser-Schutz, G. (2010). Personal pronouns and gendered speech in popular manga (Japanese Comics). *Linguistic Society of America Conference Long Abstracts* (pp. 1-4.). Washington, United States of America: Linguistic Society of America.
- Unser-Schutz, G. (2011, September). Balancing the scales: Analyzing the distribution of text amongst speakers in popular Japanese comics for girls and boys. *East Asian Popular Culture Association*. Taipei, Taiwan: Howard International House. [Conference paper]
- Unser-Schutz, G. (2011). Developing a text-based corpus of the language of Japanese comics (manga). In S. Rice, H. Baayen, & J. Newman, *Corpus-based Studies in Language Documentation, Use, and Learning*



- Corpus-based Studies in Language Documents* (pp. 213-238). Amsterdam, the Netherlands: Rodopi.
- Unser-Schutz, G. (2011). Language as the visual: Exploring the intersection of linguistic and visual language in manga. *Image&Narrative*, 12(1), 167-188.
- Unser-Schutz, G. (2011). Manga as a linguistic resource for learning. *JALT2010 Conference Proceedings* (pp. 453-465). Tokyo, Japan: JALT.
- Unser-Schutz, G. (2011). Manga as a linguistic resource for learning. *The Language Teacher*, 35(5), 20-21.
- Unser-Schutz, G. (2011). Manga ni okeru moji-hyougen no shikakuteki kubetsu to sono yakuwari ni tsuite: Tegaki no serifu/kangaegoto/komento wo rei ni. *Nihon Manga Gakkai 2011nen Taikai Youshishuu*, 22-23. [In Japanese] [Conference paper]
- Unser-Schutz, G. (2011, November). Reconsidering visual context through comics. *Japanese Association of Language Teachers*. Tokyo, Japan: National Olympics Memorial Center. [Conference paper]
- Unser-Schutz, G. (2012, October). Building and analyzing a corpus of Japanese manga: One attempt at dealing with non-standard text. *International Workshop on Corpus Linguistics and Endangered Dialects*. Tokyo, Japan: National Institute for Japanese Language and Linguistics. [Conference paper]
- Unser-Schutz, G. (2012). Kakikotoba toshite miru manga to sono hyouki-jou no tokuchou. *Nihon Manga-gakkai 2012-nen Taikai Youshishuu*, pp. 52-53. [In Japanese] [Conference paper]
- Unser-Schutz, G. (2012). Manga ni okeru moji-jouhou no bunrui to manga no yomikata: Tegaki no serifu/kangaegoto/komento ni chuumoku shite. *Manga Kenkyuu*, (18), 8-24. [In Japanese]
- Unser-Schutz, G. (2012). Reconsidering visual context through comics. *The Language Teacher*, 36(5), 20-21.
- Unser-Schutz, G. (2012). Reconsidering visual context through comics. *JALT2011 Conference Proceedings* (pp. 542-554). Tokyo, Japan: JALT.
- Unser-Schutz, G. (2012, August). Speaking out of character: Consistency in yakuwari-go 'role-playing' speech in a corpus of Japanese comics. *New Ways of Analyzing Variation in Asia-Pacific*. Tokyo, Japan: National Institute for Japanese Language and Linguistics. [Conference paper]

## 7. Bibliography

### 7.1 Referenced texts

- Abe, H. (2004). Lesbian bar talk in Shinjuku, Tokyo. In S. Okamoto, & J. S. Shibamoto-Smith, *Japanese Language, Gender, and Ideology: Cultural Models and Real People* (pp. 205-221). Oxford, United Kingdom: Oxford University Press.
- Abolfazl, S., & Omidvar, A. (2011). Internet and Arabic language. *Quarterly of Arabic Language*, 7(11), 57-70.
- Aizawa, M. (2003). Shoujo manga ni miru onna kotoba. *Meikai Nihongo*, 8, 85-99. [In Japanese]
- Akitsuki, K. (2006). Gendai shoujo-manga ni miru 'josei-kotoba'. *Gekkan-Gengo*, 35(3), 50-51. [In Japanese]
- Akitsuki, K. (2009). *Nihongo vijuaru-kei: Atarashii nihongo no kakikata*. Tokyo, Japan: Kadokawa Group Publishing. [In Japanese]
- Allen, K., & Ingulsrud, J. (2005). Reading manga: Patterns of personal literacies among adolescents. *Language and Education*, 19(4), 265-80.
- Allen, K., & Ingulsrud, J. (2007). Strategies used by children when reading manga. *Kanda Gaigo Daigaku Kiyou*, (20), 1-19.
- Allwood, J. (2008). Multimodal corpora. In A. Lüdling, & M. Kytö, *Corpus Linguistics: An International Handbook* (pp. 207-225). Berlin, Germany: Mouton de Gruyter.
- Androutsopoulos, J. K. (2000). Non-standard spellings in media texts: The case of German fanzines. *Journal of Sociolinguistics*, 4(4), 514-533.
- Anis, J. (2007). Neography: Unconventional spelling in French SMS text messages. In B. Danet, & S. C. Herring, *The Multilingual Internet: Language, Culture, and Communication Online* (pp. 87-115). Oxford, United Kingdom: Oxford University Press.
- Anonymous. (2006, September 26). *Josei wa daitei 'watashi' na no ni, dansei ga 'boku' 'ore' 'watashi' nado to tsukai wake suru no wa naze?* Retrieved November 30, 2012, from Yahoo Chiebukuro: [http://detail.chiebukuro.yahoo.co.jp/qa/question\\_detail/q116137481](http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q116137481) [In Japanese]
- Aronoff, M. (1985). Orthography and linguistic theory: The syntactic basis of Masoretic Hebrew punctuation. *Language*, 61(1), 28-72.
- Auster, P., Karasik, P., & Mazzucchelli, D. (2004). *City of Glass*. New York, United States: Picador.
- Bateman, J., Delin, J., & Renate, H. (2002). Multimodality and empiricism: methodological issues in the study of multimodal meaning-making. *GeM Report*, (1), 1-35.

- Bernabe, M. (2003). *Japanese in Mangaland: Learning The Basics*. Tokyo: Japan: Japan Publications Trading.
- Beronä, D. A. (2001). Pictures speak in comics without words: Pictorial principles in the work of Milt Gross, Hendrik Dorgathen, Eric Drooker, & Peter Kuper. In R. G. Varnum, *The Language of Comics: Word and Image*. Jackson, United States: University of Mississippi Press.
- Biber, D. (1993). The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities*, 26(5-6), 131-145.
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge, United Kingdom: Cambridge University Press.
- Bohn, W. B. (2001). *Modern Visual Poetry*. London, United Kingdom: Associated Universities Press.
- Bunka Shingi-kai. (2010, June 7). 'Kaitei jouyou-kanji-hyou' Bunka Shingi-kai toushin. Retrieved November 30, 2012, from Japanese Agency for Cultural Affairs: [http://www.mext.go.jp/b\\_menu/shingi/chousa/shotou/076/toushin/1298417.htm](http://www.mext.go.jp/b_menu/shingi/chousa/shotou/076/toushin/1298417.htm) [In Japanese]
- Buono, M. (2013). Reaching out through graphic novels. In C. Smallwood, & K. Becnel, *Library Services for Multicultural Patrons: Strategies to Encourage Library Use* (pp. 157-162). Plymouth, United Kingdom: Scarecrow Press, Inc.
- Chinami, K. (2001). Manga ni miru jendaa hyougen no kinou. *Nihongo to jendaa*, (3). Retrieved November 30, 2012, from [http://www.gender.jp/journal/no3/No3\\_2.html](http://www.gender.jp/journal/no3/No3_2.html) [In Japanese]
- Chinami, K. (2005). Manga de manabu nihongo. In M. Hinoshita, *Manga kenkyuu e no tobira* (pp. 133-153). Fukuoka, Japan: Azusa-shoin. [In Japanese]
- Choo, K. (2008). Girls return home: Portrayal of femininity in popular Japanese girls' *manga* and *anime* texts during the 1990s in *Hana yori Dango* and *Fruits Basket*. *Women: A Cultural Review*, 19(3), 275-296.
- Coates, S. (2010). Language of mobile phone novels: Japanese youth, media language and communicative practice. *Proceedings of the 18th Biennial Conference of the Asian Studies Association of Australia*, (pp. 1-13). Adelaide, Australia.
- Cohn, N. (2010). Japanese visual language: The structure of manga. In T. Johnson-Woods, *Manga: An Anthology of Global and Cultural Perspectives* (pp. 187-203). New York, United States of America: Continuum.
- Cook, V. J., & Bassetti, B. (2005). *Second Language Writing Systems*. Toronto, Canada: Multilingual Matters.
- Corkill, E. (2008, February 22). Manga makes it to the museum. *Japan Times*, p. <http://www.japantimes.co.jp/text/fv20080222a1.html>.
- Couch, C. (2000). The publication formats of comics, graphic novels, and tankobon. *Image & Narrative*, 1, <http://www.imageandnarrative.be/inarchive/narratology/chris Couch.htm>.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, (34), 213-238.

- DuBay, W. (2004). *The Principles of Readability*. Costa Mesa, United States: Impact Information.
- Eguchi, T. (2000). Postposing constructions in Japanese: Strategies for presenting information. *Faculty of Languages and Cultures, Kyushu University: Studies in Languages and Cultures*, 81-93.[In Japanese]
- Endo, O. (1997). *Onna no Kotoba no Bunkashi*. Tokyo, Japan: Gakuyo Shobo. [In Japanese]
- Endo, O. (2001). Onna no ko no 'boku/ore' wa okashikunai. In O. Endo, *Onna to Kotoba: Onna wa Kawatta ka, Nihongo wa Kawatta ka*. (pp. 30-39). Tokyo, Japan: Meiji-shoten. [In Japanese]
- Francis, W. N., & Kucera, H. (1979). *Brown Corpus manual*. Retrieved December 23, 2012, from <http://icame.uib.no/brown/bcm.html>
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6(2), 167-190.
- Fujimoto, Y. (2008). *Watashi no Ibasho wa Doko ni aru no?* Tokyo, Japan: Asahi Shinbun-shuppan. [In Japanese]
- Furui, S., Maekawa, K., & Isahara, H. (2000). A Japanese national project on spontaneous speech corpus and processing technology. *ISCA ITRW ASR2000 Automatic Speech Recognition: Challenges for the New Millenium*, (pp. 244-248). Paris, France.
- Fuse, H. (2004). *Manga wo Kaibou suru*. Tokyo, Japan: Chikuma-shobou. [In Japanese]
- Gilreath, C. T. (1993). Graphic cueing of text: The typographic and diagraphic dimensions. *Visible Language*, (3), 336-361.
- Goldberg, W. (2010). The manga phenomenon in America. In T. Johnson-Woods, *Manga: An Anthology of Global and Cultural Perspectives* (pp. 281-296). New York, United States: Continuum.
- Gotoo, H. (2007). Koopasu-gengogaku to nihongo kenkyuu. *Nihongokagaku*, (22), 45-58.[In Japanese]
- Gottlieb, N. (1995). *Kanji Politics: Language Policy and Japanese Script*. London, United Kingdom: Kegan Paul International.
- Gottlieb, N. (2005). *Language and Society in Japan*. Cambridge, United Kingdom: Cambridge University Press.
- Gottlieb, N. (2008). Japan: Language policy and planning in transition . In R. B. Kaplan, & R. B. Baldauf, *Language Planning and Policy in Asia: Japan, Nepal, Taiwan and Chinese Characters* (pp. 102-169). Bristol, United Kingdom: Multilingual Matters.
- Groensteen, T. (2007). *The System of Comics*. (B. Beaty, & N. Nguyen, Trans.) Jackson, United States: University Press of Mississippi.
- Hayakawa, G., & X(Ekkusu). (2009, May 13). *Manga suki dakara koso 'Kokuritsu Media Geijutsu Sougou Sentaa' ni wa igi ari*. Retrieved from Nikkei BPNet: <http://www.nikkeibp.co.jp/article/column/20090513/152101/?rt=nocnt> [In Japanese]
- Hayashi, O., Miyajima, T., Nomura, M., Egawa, K., Nakano, H., Sanada, S., & Satake, H. (1982). *Zusestu no*

- Nihongo*. Tokyo: Kadokawa-shoten. [In Japanese]
- Henry, A., & Roseberry, R. L. (2001). A narrow-angled corpus analysis of moves and strategies of the genre: 'Letter of Application'. *English for Specific Purposes*, 20(2), 153-167.
- Hishikari, C. (2007). A survey of the use of first-person pronouns by young people in the Tokyo metropolitan area. *Tokyo Woman's Christian University Studies in Language and Culture*, (16), 33-49.[In Japanese]
- ICV2. (2012, March 20). *Manga after Borders: An opportunity for comic stores?* Retrieved December 23, 2012, from <http://icv2.com/articles/news/22448.html>
- Ide, S. (1990). How and why do women speak more politely in Japanese? In S. Ide, & N. H. McGloin, *Aspects of Japanese Women's Language* (pp. 63-78). Tokyo, Japan: Kurosio Publishers.
- Iio, M. (2006). A study of the discourse marker *nanka* in junior college students' daily language. *The Toyo Review*, (38), 67-77. [In Japanese]
- Imano, S. (2009). *Furigana no Rekishi*. Tokyo, Japan: Shuueisha-shinsho. [In Japanese]
- Ingulsrud, J., & Allen, K. (2009). *Reading Japan Cool: Patterns of Manga Literacy and Discourse*. Plymouth, United States: Lexington Books.
- Inoue, K. (2009). Onomatopoeic neologisms in Japanese comics. *Journal of Language and Culture*, (18), 69-85. [In Japanese]
- Inoue, M. (2006). *Vicarious Language: Gender and Linguistic Modernity in Japan*. Berkley, United States: University of California Press.
- Ito, K. (2000). The manga culture in Japan. *Japan Studies Review*, (4), 1-16.
- Ito, K. (2005). A history of manga in the context of Japanese culture and society. *Journal of Popular Culture*, 38(3), 456-475.
- Itou, G. (2005). *Tezuka izzu Deddo: Hikarareta Manga-hyougenron e*. Tokyo, Japan: NTT-shuppan. [In Japanese]
- Izumitani, M. (1996). *Da* and *desu* as a sentence-final particle in Japanese. *Tokyo Ikashika-daigaku Kyouyougakubu Kenkyuu-kiyou*, (39), 1-18.[In Japanese]
- Japan External Trade Organization. (2012). *Manga and anime captivating U.S. audiences*. Retrieved December 23, 2012, from [http://www.jetro.org/manga\\_and\\_anime\\_captivating\\_u.s.\\_audiences](http://www.jetro.org/manga_and_anime_captivating_u.s._audiences)
- Japan Foundation Japanese-Language Institute Kansai. (2010). *About Anime/Manga no nihongo*. Retrieved December 20, 2010, from Anime-Manga.jp: [http://anime-manga.jp/about\\_anime-manga/eng/](http://anime-manga.jp/about_anime-manga/eng/)
- Japanese Agency for Cultural Affairs. (2001). *Heisei 12 nendo 'Kokugo ni kansuru yoron chousa' no kekka ni tsuite*. Retrieved 11 30, 2012, from [http://www.bunka.go.jp/kokugo\\_nihongo/yoronchousa/h12/kekka.html](http://www.bunka.go.jp/kokugo_nihongo/yoronchousa/h12/kekka.html) [In Japanese]

- Japanese Agency for Cultural Affairs. (2010). *Heisei 19 nendo 'Kokugo ni kansuru yoron chousa' no kekka nitsuite*. Retrieved 11 30, 2012, from Agency for Cultural Affairs: [http://www.bunka.go.jp/kokugo\\_nihongo/yoronchousa/h19/kekka.html](http://www.bunka.go.jp/kokugo_nihongo/yoronchousa/h19/kekka.html) [In Japanese]
- Japanese Magazine Publishers Association. (2012, August 8). '*Zasshi janru/kategori kubun' saishin-hyou <2012.08.08 koushinban>*'. Retrieved December 23, 2012, from <http://www.zakko.or.jp/subwin/genre.html> [In Japanese]
- Japanese Ministry of Education. (2012, November 13). *Jouyou kanji-hyou kaitei ni tomonau kyouiku-jou no taiou ni kansuru senmon kaigi (dai1kai): Gijiroku*. Retrieved November 30, 2012, from [http://www.mext.go.jp/b\\_menu/shingi/chousa/shotou/076/gijiroku/1298275.htm](http://www.mext.go.jp/b_menu/shingi/chousa/shotou/076/gijiroku/1298275.htm) [In Japanese]
- Jung, H. (2005). Nihongo to kankokugo no yakuwari-go no taishou: Taiyaku-sakuhin kara miru hon'yaku-jou no mondai wo chuushin ni. *Shakai Gengo-kagaku*, 8(1), 82-92. [In Japanese]
- Kaalep, H.-J., & Muischnek, K. (2011). Morphological analysis of a non-standard language variety. In G. Nešpore, & I. Skadiņa (Ed.), *NODALIDA 2011 Conference Proceedings*, (pp. 130-137). Retrieved November 30, 2012, from [http://dspace.utlib.ee/dspace/bitstream/handle/10062/17299/0Kaalep\\_Muischneek\\_36.pdf?sequence=1](http://dspace.utlib.ee/dspace/bitstream/handle/10062/17299/0Kaalep_Muischneek_36.pdf?sequence=1)
- Kabashima, R., & Oda, N. (2003). *Shouchan no Bouken*. Tokyo, Japan: Shougakukan Creative. [In Japanese]
- Kai, M. (1989). Manga no kotoba: Mondai teiki. *Nihongo-gaku*, (9), 33-38.[In Japanese]
- Kajino, S., & Podesva, R. J. (2007). Youthfulness, femininity, and the meaning of nonpronominal. *New Ways of Analyzing Variation (NWAV) 36*. Philadelphia, United States: University of Pennsylvania. [Conference paper]
- Kallen, J. L., & Dhonnacha, E. N. (2010). Language and inter-language in urban Irish and Japanese linguistic landscapes. In E. Shohamy, E. Ben-Rafael, & M. Barni, *Linguistic Landscape in the City* (pp. 19-36). Bristol, United Kingdom: Multilingual Matters.
- Kaneko, S. (2008). Jissen! Manga kyouzai wo tsukatta jugyou. *Gekkan Nihongo*, 21(5), 28. [In Japanese]
- Kashino, W. (2008). Kanji no shiyou wa, izen to hikaku shite hetteiru no deshou ka. In N. I. Linguistics, *Shin 'Kotoba' Shiriizu: Moji to Shakai* (pp. 69-71). Tokyo, Japan: Gyousei.[In Japanese]
- Kawasaki, K., & McDougall, K. (2003). Implications of representations of casual conversation: A case study in gender-associated sentence final particles. *Sekai no Nihongo-kyouiku Nihongo-kyouiku Ronshuu*, (13), 41-55.
- Keen, S. (2006). A theory of narrative empathy. *Narrative*, 14(3), 207-236.
- Kelts, R. (2010, June 11). Soft power useful, but Japan needs to find smarter approach. *Daily Yomiuri*. Retrieved November 30, 2012, from [http://4.bp.blogspot.com/\\_tNeTec-](http://4.bp.blogspot.com/_tNeTec-)

- \_HQs/TBefiWtpVRI/AAAAAABJY/LMKFwLclmxE/s1600/yomcom%28june%282010%29.jpg
- Kess, J. F., & Miyamoto, T. (1999). *The Japanese Mental Lexicon: Psycholinguistic Studies of Kana and Kanji Processing*. Amsterdam, the Netherlands: John Benjamins.
- Kessler, B., Numberg, G., & Schütze, H. (1997). Automatic detection of text genre. In C. Boitet, & P. Whitelock (Ed.), *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association* (pp. 32-38). Madrid, Spain: Morgan Kaufmann.
- Kindaichi, H. (1989). *The Japanese Language*. (U. Hirano, Trans.) Tokyo, Japan: Tuttle.
- Kinsella, S. (2000). *Adult Manga: Culture and Power in Contemporary Japan*. Richmond, United Kingdom: Curzon Press.
- Kinsui, S. (2003). *Vaacharu Nihongo: Yakuwari-go no Nazo*. Tokyo, Japan: Iwanami-shoten. [In Japanese]
- Kinsui, S. (2007). Kindai nihon manga no gengo. In S. Kinsui, *Yakuwari-go Kenkyuu no Chihei* (pp. 97-108). Tokyo, Japan: Kurosio Publishers. [In Japanese]
- Kishimoto, C. (2008). Webu-nikki ni mirareru hanashi-kotoba-teki buntai. *Nihongo-gaku*, 27(5), 168-176. [In Japanese]
- Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, (10), pp. 62-102.
- Kobayashi, M. (1993). Sedai to joseigo: Wakai sedai no kotoba no 'chuuseika' ni tsuite. *Nihongo-gaku*, 12(5), 181-192. [In Japanese]
- Kobayashi, M. (1999). Jishou/taishou wa chuuseika suru ka? In Gendai Nihongo Kenkyuukai, *Josei no Kotoba: Shokuba-hen* (pp. 113-137). Tokyo, Japan: Hitsuji-shobou. In Comments
- Kress, G. (2003). *Literacy in the New Media Age*. New York, United States: Routledge.
- Kudou, T. (2012, 2). *MeCab no kaihatsu keii*. Retrieved from <http://mecab.googlecode.com/svn/trunk/mecab/doc/feature.html> [In Japanese]
- Kumano, N. (2010). Nihongo-gakushuusha to anime/manga. *Hiroshima Daigaku Ryuugakusei Sentaa Kiyou*, (20), 89-103. [In Japanese]
- Kumano, N., & Hirokaga, M. (2008). An examination of 'anime and manga': Regional information overseas and Japanese-language materials. *Kokusai Kouryuu Kikin Nihongo Kyouiku Kiyou*, (4), 55-69. [In Japanese]
- Kurohashi & Kawahara Laboratory. (2012, May 12). *Nihongo keitaiso kaiseki shisutemu JUMAN*. Retrieved from <http://nlp.ist.i.kyoto-u.ac.jp/index.php?cmd=read&page=JUMAN&alias> [In Japanese]
- Kurosaki, S. (2011). Yakuwari-go kara kangaeru jishoushi 'washi' no hougensei to shutsugen jiki. *Seigakuin-daigaku Ronsou*, 23(2), 1-14. [In Japanese]
- Kyoto National Museum. (2009, 12 28). *Kanransha-suu toukei*. Retrieved 11 30, 2012, from

- <http://www.kyohaku.go.jp/jp/gaiyou/index.html> [In Japanese]
- Lavidor, M., Babkof, H., & Faust, M. (2001). Analysis of standard and non-standard visual word format in the two hemispheres. *Neuropsychologia*, 39(4), 430-439.
- Lee, W.-h., & Shaw, Y. (2006). A textual comparison of Japanese and Chinese editions of manga: Translation as cultural hybridization. *International Journal of Comic Art*, 8(2), 34-55.
- Leech, G. (2004). Recent grammatical change in English: Data, description, theory. In K. Aijmer, & B. Altenberg, *Language and Computers, Advances in Corpus Linguistics: Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)* (pp. 61-81). Amsterdam, the Netherlands: Rodopi.
- Leech, G., & Fallon, R. (1992). Computer corpora: What do they tell us about culture? *ICAME Journal*, (16), 1-22.
- MacDonald, H. (2012, October 11). *ICV2: Digital, comics sales up for 2012; manga down*. Retrieved December 23, 2012, from Publishers Weekly: <http://www.publishersweekly.com/pw/by-topic/industry-news/comics/article/54317-icv2-digital-comics-sales-up-for-2012-manga-down.html>
- Maekawa, K. (2003). Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of the ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, (pp. 7-12). Tokyo, Japan.
- Maekawa, K. (2008). Hanashi-kotoba to kaki-kotoba. *Gekkan Gengo*, 27(5), 23-33. [In Japanese]
- Maekawa, K. (2008). Kotonoha 'gendai nihongo kaki-kotoba kinkou koupasu' no kaihatsu. *Nihongo no Kenkyuu*, 4(1), 82-95. [In Japanese]
- Makino, S. (2010). Anime wo nihongo-kyouiku de dou tsukau beki ka. *Gekkan Nihongo*, 23(5), 30-31.[In Japanese]
- Marzá, N. E. (2009). *The Specialised Lexicographical Approach: A Step Further in Dictionary-making*. Bern, Switzerland: Peter Lang AG.
- Maynard, S. K. (1998). *Principles of Japanese Discourse: A Handbook*. Cambridge, United Kingdom: Cambridge University Press.
- Maynard, S. K. (2000). Speaking for the unspeakable: Expressive functions of *nan(i)* in Japanese discourse. *Journal of Pragmatics*, 32(8), 1209-1239.
- Maynard, S. K. (2002). *Linguistic Emotivity: Centrality of Place, the Topic-Comment Dynamic, and an Ideology of Pathos in Japanese Discourse*. Amsterdam, the Netherlands: John Benjamins.
- Maynard, S. K. (2008). *Maruchi-janru Danwa-ron*. Tokyo, Japan: Kurosio Publishers. [In Japanese]
- McCloud, S. (1994). *Understanding Comics: The Invisible Art*. New York, United States: Harper Paperbacks.



- McEnergy, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge, United Kingdom: Cambridge University Press.
- McGloin, N. H. (2005). Sex difference and sentence final particles. In N. Tsujimura, *Japanese Linguistics* (Vol. 3, pp. 222-238). Abingdon, United Kingdom: Routledge.
- Mealing, S. (2003). Value-added text: Where graphic design meets paralinguistics. *Visible Language*, (1), 45-58.
- Mihara, T., Nagamori, M., & Sugimoto, S. (2011). Model for representing story and visual structure of digital manga. *Jouhou Shori Gakkai Kenkyuu Houkoku*, 104(9), 1-8. [In Japanese]
- Misamisa091v. (2009, February 10). *Onna no ko na no ni ichininshou ga 'ore' ya 'boku' nado ni tsuite*. Retrieved November 30, 2012, from Yahoo Chieburo: [http://detail.chiebukuro.yahoo.co.jp/qa/question\\_detail/q1123032426](http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1123032426) [In Japanese]
- Miyazaki, A. (2004). Japanese junior high school girls' and boys' first-person pronoun use and their social world. In S. Okamoto, & J. S. Shibamoto-Smith, *Japanese Language, Gender, and Ideology: Cultural Models and Real People* (pp. 256-273). Oxford, United Kingdom: Oxford University Press.
- Mizumoto, T. (2006). Terebi-dorama to jissshakai ni okeru jousei-bunmatsushi shiyō no zure ni miru jendaa firutaa. In Nihongo Jendaa Gakkai, *Nihongo to Jendaa* (pp. 73-93). Tokyo, Japan: Hitsuji-shobou. [In Japanese]
- Mori, M. (2007). Graphic novels: Leading the way to teen literacy and leadership. *Indiana Libraries*, 26(3), 29-32.
- Morita, S. (2007). Shuujoshi/kantoujoshi no kubetsu wa hitsuyō ka. *Gekkan-Gengo*, 36(3), 44-52. [In Japanese]
- Murakami, S., & Bryce, M. (2009). Manga as an educational medium. *The International Journal of the Humanities*, 7(10), 47-55.
- Murakami, Y. (2008). Manga kyōzai no meritto/demeritto. *Gekkan Nihongo*, 21(5), 27. [In Japanese]
- Myslín, M., & Gries, S. T. (2010). k dixez? A corpus study of Spanish Internet orthography. *Literary and Linguistic Computing*, 25(1), 85-104.
- Nakajima, E. (1997). Gimon hyōgen no yōsō. In Gendai Nihongo Kenkyūkai, *Josei no Kotoba: Shokubahen* (pp. 59-82). Tokyo, Japan: Hitsuji-shobou. [In Japanese]
- Nakamura, M. (2007). 'Onna-kotoba' wa Tsukurareru. Tokyo, Japan: Hitsuji-shobou. [In Japanese]
- Nakamura, M. (2007). 'Sei' to Nihongo: Kotoba ga Tsukuru Onna to Otoko. Tokyo, Japan: Nihon Hōsō Shuppan Kyōkai. [In Japanese]
- Nakamura, T. (2006). Nobody's words: On the aspects of language in Japanese comics in the 1970s-80s. *Yamagata University Faculty of Literature & Social Sciences Annual Research Report*, (3), 29-44.[In

Japanese]

Nakano, H. (2009). *Manga Shinkaron: Kontentsu Bijunesu wa Manga kara Umareru!* Tokyo, Japan: Blues Interactions. [In Japanese]

Nakazawa, J. (2005). Development of *manga* (comic book) literacy in children. In D. W. Shwalb, J. Nakazawa, & B. J. Shwalb, *Applied Developmental Psychology: Theory, Practice and Research from Japan* (pp. 23-42). Greenwich, United States: Information Age Publishing.

Nanba, K. (2001). 'Shoujo' to iu dokusha. In K. Miyahara, & M. Ogino, *Manga no Shakaigaku* (pp. 188-220). Kyoto, Japan: Sekai Shisousha. [In Japanese]

Natsume, F. (1997). *Manga wa Naze Omoshiroi no ka*. Tokyo, Japan: Nihon Housou Shuppan Kyokai. [In Japanese]

Nihon Kanji Nouryoku Kentei Kyokai. (2012, October 18). *Kakkyuu no teido to shutsudai naiyou (Heisei 24 nendo kara kaitei)*. Retrieved November 2012, 2012, from <http://www.kanken.or.jp/teido/index.html> [In Japanese]

Nihongo Koopasu. (2004). *UniDic no kaiseki seido*. Retrieved 11 30, 2012, from [http://www.tokuteicorpus.jp/dist/modules/system/modules/menu/main.php?page\\_id=3463&op=change\\_page](http://www.tokuteicorpus.jp/dist/modules/system/modules/menu/main.php?page_id=3463&op=change_page) [In Japanese]

Nishida, T. (2011). Linguistics expression of "boku shoujo": The relation between "attributive expression" and role language. *Konan Joshi Daigaku Kenkyuu Kiyuu: Bungaku/Bunka-hen*, (48), 13-22.[In Japanese]

Nishimura, Y. (2003). Linguistic innovations and interactional features of casual online communication in Japanese. *Journal of Computer-Mediated Communication*, 9(1), 163-182.

Nishimura, Y. (2011). Japanese keitai novels and ideologies of literacy. In C. Thurlow, & K. Mroczek, *Digital Discourse: Language in the New Media* (pp. 86-109). Oxford, United Kingdom: Oxford University Press.

*Nodame kisei dikushonari*. (2004). Retrieved November 30, 2012, from Gyaboo! Japan: [http://akiba.geocities.jp/shirouto10x/gyaboo\\_jp/dictionary.html](http://akiba.geocities.jp/shirouto10x/gyaboo_jp/dictionary.html) [In Japanese]

Norricks, N. R. (2009). Interjections as pragmatic markers. *Journal of Pragmatics*, 41(5), 866-891.

Nozaki, H., & Shimizu, Y. (2000). Shinbun ni okeru kanji hindo tokusei no bunseki to NIE no tame no kanji gakushuu-hyou no kaihatsu. *Nihon Kyouiku-kougaku Zasshi*, 24(2), 121-132. [In Japanese]

Nozaki, H., Yokoyama, S., Isomoto, Y., & Yoneda, J. (1996). Moji-shiyuu ni kansuru keiryuu teki kenkyuu: Nihongo-kyouiku sien no kanten kara. *Nihon Kyouiku-kougaku Zasshi*, 20(3), 141-149. [In Japanese]

Numata, Y. (1989). Hyoukiron toshite mita manga no kotoba. *Nihongo-gaku*, 8(9), 53-59. [In Japanese]

Ogawa, S. (2006). Hanashi kotoba no shuujosi no danjo-sa no jissai to ishiki. In N. J. Gakkai, *Nihongo to Jendaa*

- (pp. 39-51). Tokyo, Japan: Hitsuji-shobou. [In Japanese]
- Ogino, M. (2001). Manga wo shakaigaku suru. In K. Miyahara, & M. Ogino, *Manga no Shakaigaku* (pp. 134-156). Kyoto, Japan: Sekai Shisousha. [In Japanese]
- Ohara, Y. (2004). Prosody and gender in workplace interaction: Exploring constraints and resources in the use of Japanese. In S. Okamoto, & J. S. Shibamoto Smith, *Japanese Language, Gender, and Ideology: Cultural Models and Real People* (pp. 220-240). Oxford, United Kingdom: Oxford University Press.
- Okamoto, S. (1995). 'Tasteless' Japanese: Less 'feminine' speech among young Japanese women. In K. Hall, & M. Bucholtz, *Gender Articulated: Language and the Socially Constructed Self* (pp. 297-325). New York, United States: Routledge.
- Okamoto, S. (2005). Social context, linguistic ideology, and indexical expressions in Japanese. In N. Tsujimura, *Japanese Linguistics* (Vol. 3, pp. 98-124). Abingdon, United Kingdom: Routledge.
- Okazaki, M. (1993). Dorama/manga ni yoru nihongo kyouiku. *Artes Liberales*, (53), 39-53.[In Japanese]
- Ôtsuka, E. (1989). *Shoujo Minzoku-gaku: Seikimatsu no Shinwa wo Tumugu 'Miko no Matsue'*. Tokyo, Japan: Koubunsha. [In Japanese]
- Ôtsuka, E. (1994). *Sengo-manga no Hyougen Kuukan: Kigouteki Shintai no Jubaku*. Kyoto, Japan: Houzoukan. [In Japanese]
- Ozaki, Y. (1999). Josei senyou no bunmatsu keishiki no ima. In Gendai Nihongo Kenkyuukai, *Josei no Kotoba: Shokuba-hen* (pp. 33-57). Tokyo, Japan: Hitsuji-shobou.[In Japanese]
- Ozaki, Y. (2002). 'Wakamono no keigo': Mikawa chuugaku no chousa kara. *Dai9Kai Kotoba Fooramu: Hanashi-kotoba no Yutakasa, Saihakken* (pp. 11-17). Mikawa-cho, Japan: National Institute for Japanese Language and Linguistics. [In Japanese]
- Pacea, O. (2009). New worlds, new words: On language change and word formation in Internet English and Romanian. *The Annals of Ovidius University Constanta-Philology*, 20, 87-102.
- Peterson, R. S. (2009). The acoustics of manga. In J. Heer, & K. Worcester, *A Comic Studies Reader* (pp. 163-172). Jackson, United States: University of Mississippi Press.
- Philips, M. (2001). Are Japanese women less feminine now? A study of sentence-final forms in Japanese women's conversation. In T. E. McAuley, *Language Change in East Asia* (pp. 47-69). Richmond, United Kingdom: Curzon.
- Pink, D. H. (2007, October 22). *Japan, ink: Inside the manga industrial complex*. Retrieved November 30, 2012, from Wired: [http://www.wired.com/techbiz/media/magazine/15-11/ff\\_manga](http://www.wired.com/techbiz/media/magazine/15-11/ff_manga)
- Pollman, J. (2001). Shaping sounds in comics. *International Journal of Comic Art*, 3(1), 9-21.
- Powell, D., & Dixon, M. (2011). Does SMS text messaging help or harm adults' knowledge of standard spelling?

- Journal of Computer Assisted Learning*, 27(1), 58–66.
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. *WCC '00 Proceedings of the Workshop on Comparing Corpora*, 16, pp. 1-16.
- Readability Research Laboratory. (2010). *Readability Research Laboratory*. Retrieved November 30, 2012, from Nihongo riidabirithii sokutei: <http://readability.nagaokaut.ac.jp/research/html/modules/tinyd0/2010> [In Japanese]
- Rommens, A. (2000). Manga Story-Telling/Showing. *Image & Narrative*, 1, <http://www.imageandnarrative.be/inarchive/narratology/aarnoudrommens.htm>.
- Sadanobu, T. (2008). Rikimu kenri, rikimu gimu. *Nihongo-gaku*, 27(5), 178-186. [In Japanese]
- Sainioa, M., Hyona, J., Bingushib, K., & Bertrama, R. (2007). The role of interword spacing in reading Japanese: An eye movement study. *Vision Research*, (47), 1575-1584.
- Sakurai, T. (2009). *Anime Bunka Gaikou*. Tokyo, Japan: Chikuma-shobou. [In Japanese]
- Sakurai, T. (2009). *Sekai Kawaii Kakumei*. Tokyo: Japan. [In Japanese]
- Satake, H. (1980). Wakamono-zasshi no kotoba: Shin-genbun'icchi-tai. *Gengo-seikatsu*, (343), 46-52.[In Japanese]
- Sato, S., Matsuyoshi, S., & Kondoh, Y. (2008). Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, & D. Tapias, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (pp. 654-660). European Language Resources Association (ELRA). Retrieved 30 2012, November, from <http://www.lrec-conf.org/proceedings/lrec2008/>
- Satoshi Sato Laboratory. (2008, August 6). 2008.08.06: *Obi2: Japanese Readability Analyzer*. Retrieved 30 2012, November, from [http://kotoba.nuee.nagoya-u.ac.jp/sc/obi2/index\\_e.html](http://kotoba.nuee.nagoya-u.ac.jp/sc/obi2/index_e.html)
- Satou, M. (2007). *Yominikui Namae wa Naze Fueta ka*. Tokyo, Japan: Yoshikawa-koubunkan. [In Japanese]
- Schodt, F. L. (1983). *Manga! Manga!: The World of Japanese Comics*. Tokyo, Japan: Kodansha International.
- Schodt, F. L. (1996). *Dreamland Japan: Writings on Modern Manga*. Berkeley, United States: Stone Bridge.
- Schodt, F. L. (2006, May 2). Manga fans take their Japanese to another level. *The Japan Times*. Retrieved November 30, 2012, from <http://www.japantimes.co.jp/text/ek20060502a1.html>
- Shamoon, D. (2004). Situating the shōjo in shōjo manga: Teenage girls, romance comics, and contemporary Japanese culture. In M. Wheeler MacWilliams, *Japanese Visual Culture: Explorations in the World of Manga and Anime* (pp. 137-154). Armonk, United States of America: East Gate.
- Shibamoto, J. (1987). The womanly woman: Manipulation of stereotypical and nonstereotypical features of Japanese female speech. In S. Urmston Philips, S. Steele, & C. Tanz, *Language, Gender and Sex in*

- Comparative Perspective* (pp. 26-49). Cambridge, United Kingdom: Cambridge University Press.
- Shibamoto, J. (2005). Women's speech in Japan. In N. Tsujimura, *Japanese Linguistics* (Vol. 3, pp. 181-222). Abingdon, United Kingdom: Routledge.
- Shibamoto, J. S., & Schmidt, D. L. (1996). Variability in written Japanese: Towards a sociolinguistics of script choice. *Visible Language*, 30(1), 46-71.
- Shibasaki, H., & Tamaoka, K. (2010). Kokugoka-kyoukasho wo moto ni shita shou/chuugakkou no bunshou nan'i gakunen hanteishiki no kouchiku. *Nihon Kyouiku Kougaku Kairon Bunshi*, 33(4), 449-458. [In Japanese]
- Shimizu, I. (1991). *Manga no Rekishi*. Tokyo, Japan: Iwanami. [In Japanese]
- Shinomoto, G. (2008, October 22). *Jibun no koto wo 'boku' 'ore' to yobu 'bokukko' 'ore-onna' no sesshikata manyuaru*. Retrieved November 30, 2012, from R25: [http://r25.yahoo.co.jp/fushigi/wxr\\_detail/?id=20080822-90004782-r25](http://r25.yahoo.co.jp/fushigi/wxr_detail/?id=20080822-90004782-r25) [In Japanese]
- Shirato, J., & Stapleton, P. (2007). Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from an empirical study in Japan. *Language Teaching Research*, 11(4), 393-412.
- Shortis, T. (2009). Revoicing txt: Spelling, vernacular orthography, and "unregimented writing". In S. Wheeler, *Connected Minds, Emerging Cultures: Cybercultures in Online Learning* (pp. 225-243). Charlotte, United States: Information Age Publisher.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation: Describing English Language*. Oxford, United Kingdom: Oxford University Press.
- Smiley, S. (2005). *Playwriting: The Structure of Action*. New Haven, United States: Yale University Press.
- Someya, H. (2002). Kanban no moji-hyouki. In Y. Hida, & T. Satou, *Nihongo Kouza Dai 6 Kan: Moji/Hyouki* (pp. 104-126). Tokyo: Meiji-shoin. [In Japanese]
- Stamatatos, E., Fakotakis, N. D., & Kokkinakis, G. K. (2000). Text genre detection using common word frequencies. *COLING '00 Proceedings of the 18th Conference on Computational Linguistics*. 2, pp. 808-814. Saarbrücken, Germany: Morgan Kaufmann.
- Street, B. (2009). The future of 'social literacies'. In M. Baynham, & M. Prinsloo, *The Future of Literacy Studies* (pp. 21-37). New York, United States of America: Palgrave MacMillan.
- Sturtz Sreetharan, C. (2004). Students, *sarariiman* (pl.), and seniors: Japanese men's use of 'manly' speech register. *Language in Society*, (33), 81-107.
- Su, H.-y. (2007). The multilingual and multiorthographic Taiwan-based Internet: Creative uses of writing systems on college-affiliated BBSs. In B. Danet, & S. C. Herring, *The Multilingual Internet: Language,*

- Culture, and Communication Online* (pp. 64-87). New York, United States: Oxford.
- Sullivan, K. (2006). Towards a new approach to men's speech research in the Japanese context: Investigating men's speech through gender and sexuality. *Shimonoseki Shiritsu Daigaku Ronshuu*, 49(3), 53-63.
- Suzuki, C. (2001). Onna no tsukau 'otoko no kotoba', otoko no tsukau 'onna no kotoba'. In O. Endo, *Onna to Kotoba: Onna wa Kawatta ka, Nihongo wa Kawatta ka* (pp. 91-99). Tokyo, Japan: Meiji-shoten. [In Japanese]
- Tada, I. (2001). Pasokon meeru to keitai meeru: 'Meeru no kata' kara no bunseki. *Nihongo-gaku*, 20(10), 44-53. [In Japanese]
- Takahashi, S. (2009). Akujo no 'yakuwari': shoujo manga 'raifu' ni miru shoujo no 'onna-kotoba'. *Gender and sexuality*, (4), 17-38. [In Japanese]
- Takayanagi, Y. (2005). *Moji no Dezain wo Yomu*. Tokyo, Japan: Socym. [In Japanese]
- Takekuma, K. (2006). Dejitaru manga no genzai. *Yuriika*, 38(1), 179-186. [In Japanese]
- Takeuchi, I. (2006). *Tezuka Osamu = Sutorii-manga no Kigen*. Tokyo, Japan: Kodansha. [In Japanese]
- Takeuchi, O. (2005). *Manga Hyougengaku Nyuumon*. Tokyo, Japan: Chikuma-shoten. [In Japanese]
- Tannen, D. (1980). Spoken/written language and the oral/literate continuum. *Proceedings of the Sixth Annual Meeting of the Berkeley Linguistics Society* (pp. 207-218). Berkeley, United States: University of California, Berkeley.
- Taruruda. (2007, June 25). *Jibun no koto wo boku tte iu onna no ko tte, mawari ni imasu ka?* Retrieved November 30, 2012, from Yahoo Chiebukuro: [http://detail.chiebukuro.yahoo.co.jp/qa/question\\_detail/q1411855674](http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1411855674) [In Japanese]
- Taylor, I., & Taylor, M. M. (1995). *Writing and Literacy in Chinese, Korean and Japanese*. Amsterdam, the Netherlands: John Benjamins.
- Tokuhisa, M., Murakami, J., & Ikehara, S. (2006). Construction and evaluation of text-dialog corpus with emotion tags focusing on facial expression in comics. In B. Gabrys, R. J. Howlett, & L. Jain, *Knowledge-based intelligent information and engineering systems* (pp. 715-724). Berlin, Germany: Springer-Verlag.
- Tsujimura, N. (2007). *An Introduction to Japanese Linguistics*. Oxford, United Kingdom: Blackwell Publishing.
- Twine, N. (1991). *Language and the Modern State: The Reform of Written Japanese*. New York, the United States of America: Routledge.
- Ueno, J. (2006). Shoujo and adult women: A linguistic analysis of gender identity in manga (Japanese comics). *Women and Language*, 29(1), 16-25.
- Waller, R. 1.-P.-2. (1980). Graphic aspects of complex texts: Typography as macro-punctuation. In P. A. Kolers, M. E. Wrolstad, & H. Bouma, *Processing of Visible Language* (Vol. 2, pp. 240-253). New York, United

States: Plenum Press.

- Wasabi Brothers Trading Company. (1998). *The MangaJin story*. Retrieved October 20, 2010, from <http://www.mangajin.com/mangajin/story>
- Werry, C. C. (1996). Linguistic and interactional features of Internet Relay Chat. In S. C. Herring, *Computer-mediated Communication: Linguistic, Social and Cross-cultural Perspectives* (pp. 47-64). Amsterdam, the Netherlands: John Benjamins.
- Whitman, J. (1999). Personal pronoun shift in Japanese: A case study in lexical change and point of view. In A. Kamio, K. Takami, & S. Kuno, *Function and Structure: In Honor of Susumu Kuno* (pp. 257-386). Amsterdam, the Netherlands: John Benjamins.
- Wilkerson, K. T., & Wilkerson, D. (2000). The gloss as poetics: Transcending the didactic. *Visible Language*, 34(3), 228-263.
- Yamaguchi, N. (2005). Komikku sekai no giongo/gitaigo. In T. Y. Kinenka, *Tsukishima Yutaka Hakushi Sanju Kinen Kokugo-gaku Ronshuu* (pp. 732-760). Tokyo, Japan: Kyuuko-shoin. [In Japanese]
- Yamane, K. (1986). *Hentai Shoujo Moji no Kenkyuu: Moji no Mukou ni Shoujo ga Mieru*. Tokyo, Japan: Kodansha. [In Japanese]
- Yumeno, K. (2011, May 20). *Dogura Magura*. Retrieved November 30, 2012, from Aozora-bunko: [http://www.aozora.gr.jp/cards/000096/files/2093\\_28841.html](http://www.aozora.gr.jp/cards/000096/files/2093_28841.html) [In Japanese]

## 7.2 Series used in corpus

- Aoyama, G. (1994-present), *Meitantei Konan*. Tokyo, Japan: Shougakukan.
- Oda, E. (1997-present), *One Piece*. Tokyo, Japan: Shueisha.
- Ohba, T. & T. Obata (2003-2006), *Death Note*. Tokyo, Japan: Shueisha.
- Kishimoto, M. (1999-present), *Naruto*. Tokyo, Japan: Shueisha.
- Nakahara, A. (2001-2007), *Rabukon*. Tokyo, Japan: Shueisha.
- Ninomiya, T. (2001-2010), *Nodame Kantaabire*. Tokyo, Japan: Koudansha.
- Obata, Y. (2002-2012), *Bokura ga Ita*. Tokyo, Japan: Shougakukan
- Shiina, K. (2005-present), *Kimi ni todoke*. Tokyo, Japan: Shueisha.
- Sorachi, H. (2004-present), *GinTama*. Tokyo, Japan: Shueisha
- Yazawa, A. (1999-present), *Nana*. Tokyo, Japan: Shueisha.