A CORPUS ANALYSIS OF THE GRAMMATICAL BEHAVIOUR OF ENGLISH LOANWORDS IN THE JAPANESE LANGUAGE

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Keith Barrs
School of Arts
University of Leicester

September 2018

Abstract

A Corpus Analysis of the Grammatical Behaviour of English Loanwords in the Japanese Language

Keith Barrs

Reflecting a long history of contact between the Japanese and English languages, a large number of English loanwords have become integrated into the general, everyday Japanese language. This study is a corpus analysis of the grammatical behaviour of frequently-used English loanwords in contemporary Japanese. It addresses a previous lack of research in the area by providing the first, large-scale, empirically-grounded account of such grammatical behaviour. Framed within a lexico-grammatical view of language, a sample of over 500 English loanwords were analysed within their naturally-occurring linguistic contexts in a large Japanese corpus. For this, corpus analysis software was used to generate a 'word sketch' for each loanword showing their most frequent grammatical relationships and their most salient collocates in each relationship. The word sketches were collated into a database of over 5000 grammatical relationships and then compared to a database of over 1000 grammatical relationships of native and Sino-Japanese words. The comparison revealed a marked pattern of behaviour of the loanwords, with a large number strongly favouring a compound noun grammatical relationship. A subsequent analysis of the most salient collocates of a sub-sample of the loanwords found that the more strongly a loanword favoured the compound noun grammatical relationship, the more strongly and exclusively it collocated with other loanwords rather than with native and/or Sino-Japanese words. In accounting for this behaviour, these loanwords appear to be 'non-catachrestic innovations' (Onysko and Winter-Froemel, 2011), a category of loanwords which are seen to be the pragmatically marked lexical choices in a language. With these findings, this study contributes to a more thorough empirically-grounded understanding of the interaction between the Japanese and English languages, as well as to a reconceptualisation of the grammatical integration of loanwords in a language.

Acknowledgements

The journey is the reward.

If you were with me on this journey, thank you.

Typographical Conventions

Transcription of Japanese Words

The general convention used in this study to write out Japanese words and their transliterations and translations is:

行きます *ikimasu* 'to go'

Where possible, the use of the words in the Japanese scripts of *kanji*, *hiragana*, and *katakana* has been kept to a minimum, and just the Romanised version is given in italics (i.e. *ikimasu*) followed by the English translation in single quotation marks (i.e. 'to go'). Single quotation marks are also used to highlight special uses of English vocabulary, such as the term 'word sketch', as well as words from other languages, such as Dutch '*gomu*'.

When Romanising Japanese words, an adapted version of the Hepburn transcription system is used. This adapted version replaces macrons with a double vowel. For example, the Hepburn form of the loanword *kā*, from the English word 'car', is written in this thesis as *kaa*. This is to aid with the reading of Japanese words.

Transcription of Grammatical Relationships (Gramrels)

The Sketch Grammar used to create word sketches in the Sketch Engine software uses a mix of Japanese script and Latin alphabet to label the grammatical relationships, or what the software more commonly terms 'gramrels', for example:

を verb

In this study the gramrels using Japanese script have been Romanised and the Japanese part written in italics. Furthermore, to distinguish them from surrounding words they will be written enclosed in square brackets, for example:

[*o* verb]

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Research Context

This study explores the grammatical behaviour of English loanwords in the Japanese language. It investigates the grammatical relationships in which frequently-used English loanwords occur in natural, contemporary written Japanese language and investigates factors that can be seen to account for patterns of behaviour in the relationships. Since the start of Japan's contact with the English language at the beginning of the 17th century, and especially after its American-led reconstruction at the end of World War Two, a substantial number of English words have been thoroughly integrated into the Japanese language. As they undergo phonological adaptations for their use in spoken discourse and orthographical adaptations for use in written discourse, these English words have also become integrated into the grammatical structure of Japanese and employed in areas ranging from news reports to song lyrics, and government white papers to children's story books (Daulton, 2008; Seargeant, 2009; Stanlaw, 2004). The lexical borrowing from English has been so extensive that around ten percent of the general, everyday Japanese vocabulary is constituted of English loanwords (Daulton, 2008; Irwin, 2011; Kay, 1995; Loveday, 1996; Otake, 2008; Stanlaw, 2004); this is a particularly extraordinary percentage considering that not more than one percent of the population habitually uses a language other than Japanese (Irwin, 2011, p. 2).

As a resident of Japan, I encounter English loanwords every day. They are in my private life in the public places I visit, on signs in supermarkets and on menus in restaurants; and they are in my professional life, pre-installed into the native lexicons of the Japanese university students to whom I teach English. I became interested in loanwords as soon as I started studying the Japanese language, having learnt that they are typically written in *katakana*, a script primarily reserved for writing words borrowed from languages other than Chinese.[1] This has the advantage of making them easily identifiable in the written Japanese language. I also quickly realised that if I learnt some basic rules of how English sounds are converted into Japanese sounds, such as adding a vowel to English words ending in a consonant, like in the loanword *salada* borrowed from English 'salad', I was gradually able to recognise English loanwords in their spoken form as well.

[1] Words borrowed from Chinese are typically categorised as Sino-Japanese and make up their own lexical stratum (see Section 2.2.4 for a discussion of Japanese lexical strata).

My research interest in English loanwords in Japanese began with investigating the number and type of words which were borrowed from English (Barrs, 2012a, 2012b, 2013b), but took a new direction after a conversation in Japanese with another lecturer at the university in which I work. During the conversation, we talked about what we had been doing over the previous weekend. I told him I had gone strawberry picking and I said the phrase たくさんのストロベリーを食べた *takusan no sutoroberii o tabeta* 'I ate a lot of strawberries', using the English loanword *sutoroberii* 'strawberry' which I knew is a common English loanword. He politely corrected me and said in this situation it was better to say 苺を食べた *ichigo o tabeta*, with the English loanword replaced by the native language equivalent. When I asked why, he explained that "people do not eat *sutoroberii*, they eat *ichigo*. *Sutoroberii* is a flavour rather than the fruit". I found this intriguing as it suggested that *sutoroberii* had a restricted grammatical usage in Japanese. I had already learnt from studying the language that some English words took on additional meanings in Japanese, such as クレーム *kureemu* meaning not only 'claim' but also the additional meaning of 'complaint', such as *kureemu wo tsukeru* 'to make a complaint', but this example of *sutoroberii* got me interested in what kind of phrases were common, and more interestingly, uncommon when using the loanword.

From some of my other research projects I was already familiar with the Sketch Engine corpus analysis software, and in particular its 'word sketch' function which produces a one page summary of the grammatical and collocational behaviour of a word (Kilgarriff, Rychly, Smrz, & Tugwell, 2004). I produced a word sketch for *sutoroberii* (Figure 1.1) to see if the restricted usage explained by my colleague was transparent in its grammatical and collocational behaviour in a corpus of written Japanese.

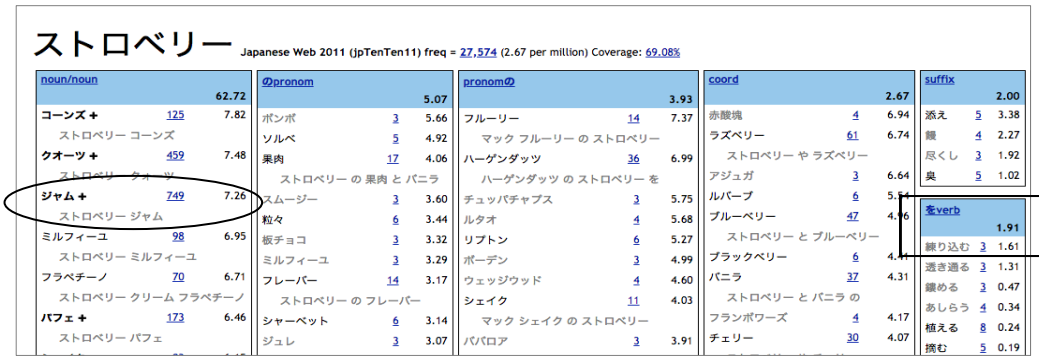


Figure 1.1 A word sketch for the English loanword *sutoroberii* produced in the Sketch Engine from the 8-billion-word jpTenTen11 corpus.

The word sketch for *sutoroberii*, generated from all of its 27,574 instances in the 8-billion-word jpTenTen11 Japanese web corpus (see further details of this corpus in Section 4.3), showed that it occurs most frequently in Japanese in a compound noun structure, shown by the first column in Figure 1.1 labelled [noun/noun]. An example collocation from this grammatical relation, or 'gramrel' as it is more commonly called within the Sketch Engine, is *sutoroberii jyamu* 'strawberry jam', shown in Figure 1.1 in the oval shape. This compound noun relationship of *sutoroberii* is shown to occur in 62.7% of the occurrences of the loanword in the corpus, 12 times more than the next most frequent grammatical relationship of [*no* pronom] which expresses a possessive grammatical relation (shown as [の pronom] in Figure 1.1). Supporting my colleague's comment, the phrase *sutoroberii o tabeta* 'I ate strawberries' does not appear at all in the word sketch, and furthermore, the [*o* verb] gramrel (shown in Figure 1.1 as [を verb]), which expresses the loanword being acted upon by a verb (i.e. as the noun object of a verb), contains only a small number of verbs and all with a very low frequency. For example, the first collocate in this gramrel, *nerikomu* 'to knead/blend into' (shown in the box shape), is ranked as the collocate most strongly associated with *sutoroberii*, but only occurs 3 times. When I investigated further and carried out a simple query in the Sketch Engine for the phrase *sutoroberii o taberu* to bring up all instances of this phrase in all grammatical inflections, such as conditional and past, only 20 hits were returned, the majority from personal blog sites. This is within a corpus of over 8 billion words.

The grammatical and collocational distribution of the loanword *sutoroberii* in the corpus from which the word sketch was produced indeed suggests that the phrase *sutoroberii o tabeta* is exceedingly rare in Japanese, and further that the grammatical environment of *sutoroberii* in general appears very much restricted to occurring in compound noun structures. From this small exploratory study, I became interested in the grammatical behaviour of other English loanwords in Japanese, particularly of those which are common, everyday words in the Japanese language, and wanted to expand the investigation to see if there were other loanwords with similar grammatical restrictions. I wanted to know what kind of grammatical behaviour of the loanwords would be uncovered if I collected and compared a large set of English loanword word sketches. This would then be interesting to compare with the grammatical behaviour of other words in the Japanese lexicon. This thesis presents this expanded investigation of the grammatical behaviour of English loanwords in Japanese. More precisely, it investigates the patterns of grammatical behaviour of frequently-used English loanwords in contemporary Japanese through an analysis of the grammatical relations in

which they participate in natural Japanese language, and explores factors which appear to account for observed patterns in these relations.

The study is further motivated by the fact that the grammatical behaviour of English loanwords in Japanese has so far gone mostly undescribed, in contrast to the extensive body of knowledge on their phonological and orthographical integration (Mogi, 2012). The general lack of attention given to the grammatical behaviour of loanwords is argued in this thesis to have been caused by several entrenched practices in the field of loanword research. The first of these is that there has been an overwhelming predominance of studies on loanword phonology, leaving other areas of study, such as loanword morphology and their syntactic integration into the borrowing language, much less explored. Secondly, where investigations of the grammar of loanwords has been undertaken, it is predominantly limited to analysing the lexical categories, or part(s)-of-speech, of the loanwords. Thirdly, loanword research can be seen to have long displayed a dependence on dictionaries as the source of loanword data. This is despite rapid advancements in the methodology of corpus linguistics over the last few decades which focuses on the analysis of the natural-language behaviour of loanwords, rather than simply their dictionary entries.

1.2 Research Problem

Lexical borrowing research addresses a wide variety of questions concerning the transfer of words out of one language, the source language, and into another, the recipient language. Major questions asked include: Which words are and are not borrowed out of a particular source language? What motivates the borrowing? How are the loanwords integrated into the recipient language? How do they develop in structural form and semantic/pragmatic meaning over time? and What functions do they perform throughout society? (Kang, 2013; Winford, 2010; Zenner & Kristiansen, 2014). With the lexical borrowing from English into Japanese, the research conducted to answer these questions has fostered the growth of a considerable body of knowledge both on the functions of loanwords in Japanese society, such as in advertising (Seargeant, 2005; Takashi, 1990), entertainment (Kuang, 2009; Stanlaw, 2004), and education (Daulton, 2008; Martin, 2004); and their linguistic form in the Japanese language, including descriptions of their phonological (Crawford, 2009; Irwin, 2011), orthographical (Igarashi, 2007), and morphological structure (Irwin, 2011; Stanlaw, 2004).

In the studies of the linguistic form of English loanwords in Japanese, by far the largest amount of research has been conducted on their phonological structure (Irwin, 2011). This is due to the extensive number of adaptations regularly needed for the phonological form of English words to be adapted to the more restricted Japanese phonemic inventory. Consequently, a substantial amount of research in this area over the last forty years has contributed to a thorough and detailed understanding of many aspects of English loanword phonology in Japanese. This body of work includes theoretical analyses of what underlies the adaptations (Crawford, 2009; Katayama, 1998; Lovins, 1975), systematic descriptions of the major adaptation processes (Irwin, 2011), explanations of computational methods of predicting adaptations (Blair & Ingram, 2003), and corpus-based examinations of naturally-occurring adaptations (Kawahara & Sano, 2013). Moreover, because of the one-to-one grapheme-to-phoneme relationship with the use of the *katakana* syllabary to write loanwords in Japanese, much attention has also been given to the description and analysis of orthographical adaptations of English words (Igarashi, 2007; Irwin, 2011; Loveday, 1996; Preston & Yamagata, 2004; Stanlaw, 2004; Tamaoka & Miyaoka, 2003).

In contrast, there are considerably fewer works focused on the grammatical structure of English loanwords in Japanese. Indeed it has been stated that "the grammatical behaviour of loanwords - how loanwords are used within sentences - has hardly been studied at all" (Mogi, 2012, p. 22). To some extent this can be explained by the fact that English words can be integrated into Japanese with relatively few morphological and syntactic adaptations (Irwin, 2011). The majority of words borrowed into Japanese from English are nouns and function syntactically as the same part of speech in Japanese. Moreover, morphologically, nouns are non-inflecting in Japanese meaning that they are not marked for gender, number, or case (Kageyama & Saito, 2016). This has inevitably caused academic attention to be focused elsewhere, such as on the aforementioned extensive phonological and orthographical adaptations. However, this absence of morphological and syntactic adaptations, whilst certainly a contributing factor, is not enough to explain the general lack of empirical treatment of their grammatical behaviour.

A more instrumental cause is rooted in how the grammar of words has traditionally been conceptualised in linguistic research. Words have typically been seen to fill open slots in the grammatical structure of a language, with the available slots defined by a small number of lexical categories, or parts of speech, such as nouns, verbs, adjectives, and determiners (Sinclair, 1991). The positing of a small number of lexical categories is

advantageous in linguistic research in that it allows rules to be defined for each category and these can be used to efficiently describe the grammatical structure of words within that category (Smith, 2015a). Modern lexical research, however, particularly that which uses a corpus linguistics methodology to investigate the naturally-occurring behaviour of words in large samples of language, has shown that lexical categories are not only inadequate for accounting for the complexity of the grammatical behaviour of words, but are also internally problematic in that words placed into each category very frequently display different properties from one another, making the category classification for the most part arbitrary (Smith, 2015b). This means that if the analysis of the grammar of words stops at the point of classifying them into lexical categories, as has been traditional in loanword research, the complexity of their grammatical behaviour in naturally-occurring discourse goes undescribed.

A further major cause of why the grammatical behaviour of English loanwords remains largely unexplored is rooted in the methodological practice of using dictionaries as the primary source of evidence in loanword research. Dictionaries have long been an invaluable resource for specifying which loanwords have been codified into a language, when they were first attested, which language they were borrowed from, what phonological and orthographical form they take, which lexical category they function as, and what they typically mean (Kilgarriff, 2005). However, they rarely provide evidence of frequency or recurring patterns of usage. As such, whilst a dictionary entry for a loanword may list it as a noun, it is unlikely to specify how probable it is that the loanword appears as the modifying element in a compound noun and whether the loanword occurs more frequently as the subject or the object of a sentence. For such detailed grammatical description, corpora can provide the necessary large collections of natural language, and corpus-analysis software can provide the necessary tools with which to analyse the language. The adoption of a corpus linguistics methodology, however, is infrequent in loanword research (Zenner, Dirk, & Geeraerts, 2014) and has been almost completely absent in the analysis of English loanwords in Japanese (Inagawa, 2010).

Because of these issues, the understanding of the grammatical behaviour of English loanwords in Japanese has been largely limited to theoretical sketches such as "loanwords that have been incorporated into the Japanese language system generally follow the morphological and syntactic rules of Japanese grammar" (Stanlaw, 2004, p. 77) and "loan words [*sic*] fit into the Japanese syntactical structure as if they were native words, being ascribed particles such as subject and object markers where

necessary" (Kay, 1995, p. 72). Not only are such sketches broad generalisations, but they also lack empirical evidence to support their validity. Consequently, there is a wide gap between the comprehensive and systematic understanding of the phonological and orthographical behaviour of the loanwords on the one hand, and the limited and mostly theoretical understanding of their grammatical behaviour on the other.

The fact that the grammatical behaviour of English loanwords in Japanese has rarely been studied is a problem for two main reasons. Not only have English loanwords been integrated into the core of the Japanese lexicon, meaning that they are found in general, everyday Japanese discourse rather than being restricted to specialised domains, but also the number of English loanwords which are found in this core is very high. Taken together, these mean that English loanwords are frequently encountered in general Japanese discourse, which warrants an understanding of their grammatical behaviour for the applied value this knowledge can have to areas such as Japanese lexicography and Japanese-language education.

This gap in understanding presents an opportunity to contribute to a richer overall description of English loanwords in Japanese, through a systematic and large-scale analysis of their grammatical behaviour. Conducting such research would address the issues presented above, by going deeper in the analysis than the limited notion of lexical categories and seeking usage-based rather than dictionary-based evidence. Taking this opportunity, the present study examines the grammatical usage-patterns of English loanwords in Japanese. It uses corpus methods to analyse patterns in the grammatical distribution of loanwords in a large body of naturally-occurring Japanese language.

The analysis is conducted on a sample of 587 frequently-used English loanwords in Japanese, which are extracted from word-frequency lists of three large written Japanese corpora containing texts from newspapers, novels, non-fiction works, magazines, academic articles, and general web data. A 'word sketch' is produced for each of the loanwords using the Sketch Engine corpus analysis software, showing the most common grammatical relationships, and for each of these, the most strongly-associated lexical collocations. These word sketches are then combined to create a database of over 5000 grammatical relationships, and this is the basis for exploring patterns in the grammatical behaviour of the loanwords. To elucidate such distributional behaviour, it is compared to that of a sample of words from the native and Sino-Japanese strata of the Japanese lexicon. The aim is thus to provide the first comprehensive, systematic, and comparative account to be given of the grammatical behaviour of English loanwords in Japanese.

1.3 Research Contributions

In conducting a corpus analysis of the grammatical behaviour of English loanwords in Japanese, the present study aims to make a conceptual contribution to this area of loanword research. The traditional perspective taken on the grammatical behaviour of words integrated into another language is narrowly focused on the open slots into which the loanwords are placed. This perspective can be compared to a child's 'shape-sorter' toy where the objective is to slot pieces of different shapes into the correspondingly-shaped holes. In the toy, the pieces are independent shapes which slot into specified holes, and their relationships with other pieces are not particularly relevant in achieving the toy's objective. In the same way, the tradition in lexical borrowing research of treating loanwords as single-word units removed from their naturally-occurring linguistic contexts (Zenner & Kristiansen, 2014) means their grammatical behaviour has generally been viewed within this open slot perspective.

Arguing for a wider focus on how loanwords are organised with other words into grammatical relationships, made possible with corpus methods of analysis, the present study adopts a different perspective on the analysis of the grammatical behaviour of loanwords. This perspective is one which views lexis and grammar as two interdependent ends of a continuum, as lexico-grammar (Halliday, 1985; Sinclair, 1991). Framed in this perspective, which has become known as Neo-Firthian linguistics following the influential work by J.R. Firth on how language is a pattern-oriented system of communication (Thomas, 2017), loanwords are seen as interconnecting with other words around them. This shifts the perspective to that of a 'jigsaw puzzle'. For a piece to be fitted into the puzzle, attention needs to be given to the other pieces to see if and how they fit together. In the same way, the present study gives attention to how loanwords interconnect with other words around them. This perspective allows a fuller and more detailed description of their grammatical behaviour than is possible by only analysing the gaps into which the loanwords slot.

In their methodologies, most previous studies on English loanwords in Japanese have used manual methods of data selection and analysis. This has meant that they have only been able to analyse a very small percentage of the overall number of English loanwords. Moreover, the extent to which each loanword can be analysed has also been heavily restricted by the manual analysis. This has been inevitable considering the fact that large, easily-accessible Japanese language corpora have been slow to emerge (Kilgarriff et al., 2014; Srdanovic, Ida, Shigemori-Bucar, Kilgarriff, & Kovar, 2011).

Furthermore, whilst corpus methods have gradually become common in other areas of lexical research, such as lexical semantics and lexicology (Zenner et al., 2014, p. 44), lexical borrowing research still maintains its focus on loanwords as single-word units (Backus, 2014; Zenner & Kristiansen, 2014). Now, however, there are Japanese corpora and sophisticated corpus-linguistic tools available for a methodological shift to take place in this area of lexical research as well (Inagawa, 2010; Mogi, 2012). For the present study, an 8-billion-word corpus of Japanese language data and specialised corpus-analysis tools are used to analyse the naturally-occurring contexts of a large sample of frequently-used English loanwords.

1.4 Thesis Outline

Following this introductory chapter, Chapter 2 gives a historical account of Japan's contact with English, focusing on how the importation of a large number of English loanwords has been facilitated by a long history of borrowing both language and culture from a number of foreign nations. Chapter Three reviews the literature on lexical borrowing both in the field of study in general and in the context of the lexical borrowing from English into Japanese. It explains how the term loanword is defined in this study, discusses three practices in loanword research which have been seen to be problematic for the study of loanword grammar, and gives the overall aim and research questions of the thesis. Chapter Four presents the data and methodology used in the analysis of the grammatical behaviour of the loanwords, giving details of the methodological approach underpinning the research and the steps taken in constructing and analysing two databases of grammatical relationships. Chapter Five compares the two databases in order to uncover patterns of distribution, and Chapter Six further explores the patterns by examining the collocational behaviour of a sub-sample of 30 English loanwords. This behaviour is then discussed in light of several theories of the functions of loanwords in a language. Chapter Seven summarises the findings of the research by restating the conceptual, methodological, and empirical contributions, discussing several limitations of the data and methodology adopted in the research, and then finally considering some possible future avenues of investigation of loanwords as the field of corpus linguistics continues to advance.

2 Background

2.1 Chapter Overview

This chapter gives a historical account of Japan's contact with English. It describes how the intense contact with China and Korea from the 5th century A.D. laid the foundations of a practice of borrowing and integrating aspects of foreign languages and cultures which remains strong in present-day Japan. It discusses how the importance of contact with Asian nations was slowly replaced by the need for contact with European nations, who were steadily growing in global power around the 15th century. Japan's contact with the Portuguese and Spanish in particular had a huge impact on the country, not only because of the contact with the languages of these nations, but also because it brought Japan into its first contact with the English language. Over the last four hundred years, it is this contact with English which has had the most profound impact on all aspects of life in Japan, with its importance seen in particular throughout the Japanese education system, the advertising industry, and entertainment media.

A typology of this English in Japan is presented which details a broad division between *inter*-national English, used primarily for international communications conducted in English, and *intra*-national English used in domestic communications conducted in Japanese. This division is represented by two main types of *eigo*, which is the Japanese word for 'the English language': *gaikokugo eigo* which is the English used as a foreign language in international communications, and *gairaigo eigo* which is the English loanwords integrated into the Japanese language and used in intranational communications. The current study is focused on the latter *gairaigo* variety of English, and this chapter discusses its features in detail, especially regarding the modifications that turn English words into English loanwords in Japanese. To help in the overall illustration of the different types of English in Japan, examples are given of English used in public spaces in Japan, known as an area's 'linguistic landscape'. The chapter ends with a discussion of present-day attitudes towards English in Japan, showing how the overall long-term favourable reception of English loanwords fosters continued borrowing and integration into the language.

2.2 A History of Language Contact in Japan

Very little is known about the pre-history of the Japanese language, due to the absence of a writing system for the language until around the 5th century. Even the genetic relationship of Japanese to other languages, and to a larger language family, has yet to be firmly established. Whilst there has been much academic discussion on the issue, the literature over the last sixty years agrees that a precise classification remains difficult (Frellesvig, 2010; Hattori, 1959; Miller, 1967; Shibatani, 1990). Shibatani, for example, discusses in the preface to his survey of the languages of Japan that scholars are still unable to agree on whether Ryukyuan, a language used in the islands immediately south of Japan, is separate from Japanese or in fact a dialect of Japanese (1990, p. xiii).

The influence from China around the 5th century A.D. brought not only the start of the written record of Japanese, but also the start of centuries of contact with foreign languages from which a great many aspects of the Japanese language have been borrowed. Indeed, the history of the Japanese language since the 5th century A.D. is a history of language contact and linguistic borrowing, resulting in the dynamic mix of orthographies and lexical strata seen in the modern-day Japanese language. This has been called a contact tradition, a term used to summarise the centuries of influence from contacts with nations such as China, Korea, Portugal, Spain, Holland, Germany, England and America which have shaped Japan's history (Loveday, 1996). Of these influences, the impacts on the Japanese language from Chinese and English have been the most influential. Taking the Japanese lexicon alone, around 60% is made up of borrowed items, with most being derived from Chinese, then English, then other European languages (Irwin, 2011). Table 2.1 shows the main periods and countries of contact through Japan's history. This table is a revised version of that given by Koscielecki (2006, p. 26), which is itself adapted from Vos (1963). Whilst Koscielecki splits the period of contact with English into three phases, it is more accurately represented by four phases; dividing the period of 1854-1941 into a pro-English period (1854-1937) and an anti-English period (1937-1945).

Table 2.1 Historical phases of Japan's foreign contact (adapted from Koscielecki, 2006).

Phase (Revised Version)	**Date**	**Phase** (Koscielecki, 2006)	**Date**
Chinese	5th century onward	*Chinese*	5th century onward
Portuguese	1543-1639	*Portuguese*	1543-1639
Spanish	1592-1624	*Spanish*	1592-1624
English 1	1613-1623	*English 1*	1613-1623
Dutch	1609-1854	*Dutch*	1609-1854
English 2	1854-1937	*English 2*	1854-1941
English 3	1937-1945		
English 4	1945 onward	*English 3*	1945 onward

These phases cover the main periods of language contact over the last 1,500 years of Japanese history. Each of these is described below, with the first phase involving the immense influence from the Asian languages of Chinese and Korean, the next four phases comprising the early European contacts with Portuguese, Spanish, and Dutch, and the last three phases being where English makes its profound impact on Japanese language and society. The history of language contact in Japan is fundamental to the understanding of how English has come to have had such a huge impact on all aspects of society, particularly in why the country has been so receptive to the incorporation of such a massive number of foreign borrowings.

2.2.1 *Asian influences on the Japanese language*

Despite Japan's geographical isolation as an island nation off the far east coast of the Asian mainland, its proximity to the Korean peninsula has continually brought significant Asian influences into the country. One of the most important took place at the beginning of the Yayoi period, an Iron Age era spanning the last three hundred years B.C. and the first three hundred years A.D. At this time a large migration occurred from the Asian mainland, via the Korean peninsula, which brought in farming techniques and tools to Japan (Hasegawa, 2015, p. 5). This opened a route of contact that would continue to be exploited over the following centuries.

From a linguistic viewpoint, the single most important influence which has shaped the course of the Japanese language over the last 1,600 years came in the 6th century A.D., during the Old Japanese (592-794) period, with the introduction of Buddhism from China and Korea (Shibatani, 1990, p. 120). This brought with it the importation of the Chinese writing system, and even though the Chinese script had already been introduced to Japan via Korean scholars several centuries earlier in its prehistoric period, the introduction of Buddhism gave the learning and teaching of the Chinese writing system its most focused purpose.

The importation of the Chinese writing system subsequently encouraged the borrowing of an immense number of loanwords, something which greatly expanded the expressive power of Japanese. Now that the Japanese language could be officially and systematically recorded, there followed the production, publication, and dissemination of texts such as government documents and literary novels. The most fundamental aspect of this early writing system, which is still true of the modern language, was *kanji*; a term which refers both to the script in general and to the tens of thousands of individual logographic characters. The early use of *kanji* in Japan very closely mirrored the Chinese language, particularly in terms of the readings of the characters and in the syntax within texts. However, developments over the centuries gradually produced the Japanese-derived syllabaries of *hiragana* and *katakana* to help with the overall reading of the *kanji* script, along with the simplification of the writing of the individual *kanji* characters themselves. These developments have continued through to the present day where the writing of Japanese is mainly achieved through an interaction of the three scripts of *kanji*, *hiragana*, and *katakana*.

From these early contacts, the Asian influences of China and Korea continued to shape the Japanese language through the periods of Late Old Japanese (794-1192), an era which saw the domestic development of the aforementioned *kana* syllabaries, and Middle Japanese (1192-1603), which saw the beginnings of the modernisation of Japanese in terms of pronunciation and syntax. During these centuries, Chinese was the unrivalled source of new words added to the Japanese language, and whilst the overall influence of China on Japanese language and society gradually began to lose its footing to that of European powers at the end of the Middle Japanese period, the foundation of the present-day language is still very much Chinese-influenced (Shibatani, 1990).

2.2.2 *European influences on the Japanese language*

The first significant Japanese contact with European powers began in 1543, the beginning of what has been called an Iberian phase where Japan came into contact with Portuguese and Spanish missionaries and traders (Irwin, 2011, p. 23). At this time, the Portuguese entered Japan via an island route in the southern area of Kyushu (Frellesvig, 2010, p. 299; Shibatani, 1990, p. 121). This facilitated the arrival of Jesuit missionaries towards the end of the decade, with their considerable influence on Japanese culture continuing up until 1639, when the country entered a period of seclusion from outside contact, known in Japanese as *sakoku* (closed-door policy).

The impact of the Portuguese Jesuit missionaries on the Japanese language was considerable. They were the first to introduce a printing press to Japan, using it to print and distribute their *kirishitan shiryo* (Christian literary materials), such as dictionaries and grammar books (Frellesvig, 2010, p. 299). The translation and dissemination of these materials involved working between their native language, which used the Latin alphabet; and the Japanese language, which used Sino-Japanese scripts, meaning they were also the first people to systematically introduce Japan to the Latin alphabet and Western literature.

A further example of the substantial impact of the Portuguese influence is found in the beginnings of extensive lexical borrowing from languages other than Chinese. This was initiated by the Portuguese missionaries themselves, rather than by the Japanese. The missionaries had previously opted for loan translations of Latin religious terms using Chinese characters common in Japanese, but inherent difficulties in this process made them change their methods to introduce direct loanwords instead. This novel process of introducing loanwords then led to other changes in the language, brought about by mismatches between the pronunciation of the loanwords, and the available graphemes in Japanese to record them. An example is in the introduction of the *handakuten* diacritic sign, used extensively in the present-day language. Frellesvig states that it was with the publication of the Christian document *Giya do pekadoru* in 1599 that the use of the *handakuten* diacritic to allow the spelling of /p/ became commonplace (2010, p. 302). This change was necessary because of the significant number of loanwords from Portuguese which required the /p/ spelling. The initial *katakana* character in such loanwords as プライバシー *puraibashii* ‘privacy’ and パートナー *paatonaa* ‘partner’ are two examples of frequent borrowings in present-day Japanese which exploit this historical change.

Following the Portuguese came the Spanish, continuing the period of Iberian influence in Japan. But their impact on Japanese society and language is markedly less considerable than that of the Portuguese, as a result of the short period until all Spanish were expelled from Japan in 1624 (and the Portuguese in 1636) during a time of persecution of Christians under the Tokugawa Shogunate (Irwin, 2011; Stanlaw, 2004). Japan then entered its period of national isolation (1633-1866) where the only European foreign contact which continued was with the Dutch.

The Dutch arrival into Japan came in 1609, but their influence only became significant after they were granted permission to remain in Japan during the period of national isolation, when all other European contacts were prohibited. Despite being allowed to stay, they were mainly confined to an island off the coast of Nagasaki, and were not allowed to study Japanese (Irwin, 2011, p. 36). The main influences of the Dutch during this period came in three areas: the learning of the Dutch language by Japanese interpreters; periodical reports about global trade and warfare submitted by the Dutch merchants; and knowledge of Western science and medicine via Dutch language books, many of which were translated by the Japanese interpreters who were learning the language.

The Dutch influence in Japan is considered the catalyst for Japan's rapid modernisation, after it came out of its period of national isolation in the second half of the 19th century. This modernisation was fostered by the scientific and medical knowledge brought into Japan, embodied in the hundreds of loanwords which entered the language. Many of these are still common in present-day Japanese, such as *ponpu* (English 'pump', from Dutch 'pomp'), *gomu* (English 'rubber', from Dutch 'gom'), and *konma* (English 'comma', from Dutch 'komma'). One of the most important influences of this period of contact was that Japanese scholars working with the Dutch language were laying the foundations of Western scholarship, with Stanlaw stating that "there is little doubt that the story of English in Japan would be vastly different if not for the presence of Dutch speakers, both native and foreign, in the country" (Stanlaw, 2004, p.47).

2.2.3 *The influence of English on the Japanese language*

The origin of Japanese contact with English lies in the arrival of the English sailor William Adams into Southern Japan at the very start of the 17th century (Kay, 1995; Stanlaw, 2004). This brought the shogun Tokugawa Ieyasu into regular contact with English, and whilst their interactions had minimal influence in introducing the English language to Japan in general (Irwin, 2011; Stanlaw, 2004), they did lay important foundations of cooperation with English speakers which would later be built upon and exploited. The first of these exploitations, and the start of the first main phase of contact with English, was the arrival of John Saris and the opening of a merchant's office in Hirado, Nagasaki which lasted until 1623. The closing of this office, and the expulsion of the English at the beginning of Japan's national isolation, ended this initial phase of English influence in the country.

In 1808 during the height of Japan's closed period, the British ship H.M.S. Phaeton, flying Dutch flags, entered the port at Nagasaki and caused a security alert. Whilst the incident was short-lived, it made the Japanese authorities aware of the growing influence of English in the world, and initiated the beginning of formal English study in the country (Irwin, 2011, p. 54). Stanlaw states that "this incident revealed the growing weakness of the Tokugawa shogunate which had ruled Japan for two centuries, and foreshadowed the coming of the Americans forty years later" (Stanlaw, 2004, p. 49). Prior to this large-scale arrival of Americans into Japan in 1853, there were two smaller-scale incidents which continued to lay important foundations for the propagation of English influence through Japan. In 1848 an American sailor named Ranald MacDonald drifted towards the north coast of Hokkaido, seeking adventure in unknown lands, and was captured and quickly imprisoned. He was then sent to Nagasaki and ordered to teach English to a group of Dutch interpreters. He became Japan's first native-speaker teacher of English (Loveday, 1996, p. 61). The second incident involved Nakahama Manjiro, a Japanese resident who was rescued from a shipwreck and taken to America, where he stayed for ten years. He brought back to Japan both a proficiency in the spoken language, and even more importantly several English language texts such as Webster's English Dictionary. These became essential materials for the formal study of English (Stanlaw, 2004).

The start of the second main phase of the influence of English in Japan was in 1853/1854, with the arrival of Commodore Perry, a small fleet of ships, and a proposal of a trade treaty brought from the U.S. government. This was the kind of event which

the Japanese had feared, particularly after the 1808 incident with the British ship H.M.S. Phaeton, but one which they had failed to properly prepare for, having no significant navy with which to resist the foreign powers. A trade treaty (the Kanagawa Treaty) was signed in 1854, opening up two ports primarily for American trade. More substantial foreign influence arrived in the country in 1858 following the signing of the Ansei Commercial Treaties with the U.S., Great Britain, France, Russia, and Holland, bringing not only advances in trade but also immigration into port cities.

The opening of Japan to foreign influences brought various struggles with language and leadership. The terms of the treaties differed between the various primary translations of Japanese, Dutch, English, leading to confusion between the nations as to what had been agreed upon. The many concessions given over to foreign powers via the treaties created domestic unrest and eventually weakened the power of the governing shogunate. During this period, it became clear that in order to try and preserve the stability and security of Japan, in the face of foreign influence, it would need to modernise itself. Ironically, this involved embracing and exploiting the language and knowledge of the Western powers which Japan proposed to defend itself against. The biggest change necessary was a transition from the learning of Dutch to the learning of English, something very clearly seen by Fukuzawa Yukichi, one of the founders of modern Japan, when he observed that "I had inklings that English was the most widely used language. A treaty with the two English-speaking countries had just been concluded. As certain as day, English was to be the most useful language of the future" (Fukuzawa, 1899, as cited in Stanlaw, 2004, p. 55).

With the end of the Tokugawa shogunate in 1868, brought about by continued unrest with the country's acceptance of foreign influences, Japan entered the Meiji period (1868-1912), a time when the country moved from being a feudal nation to a world power. Loveday states that this period of enlightenment brought "sweeping social, economic, and political transformation, such as the adoption of the Gregorian calendar, the development of state education, and the abolition of the feudal class-system" (Loveday, 1996, p. 65). This saw an all-out determination to modernise and industrialise the country, with many economic and scientific advisors brought over from Great Britain, the U.S., and various European countries, and English language teachers invited to work in the 82 newly-established schools (Loveday, 1996, p. 54). This firmly established the importance of English throughout all aspects of Japanese culture whilst also striving to follow the "principle of modernizing society while maintaining Japanese values" (Kay, 1995, p. 67); the same principle as when Chinese culture and language

was borrowed from the mainland over one thousand years before. As most of the scientists, advisors, and scholars were British or American, both of these varieties of English took hold in the country and can be seen in the present-day language with various spellings and pronunciations of English-derived words.

In 1872, English became a compulsory subject in the newly reorganised elementary and junior high schools (Irwin, 2011, p. 54; Loveday, 1996, p. 66). However, Loveday states that its implementation throughout the education system was troubled, due to a lack of staff and teaching materials. As a result, the learning of English past the age of 11 was mainly an academic practice of the wealthy and higher classes. For the average Japanese citizen who was interacting with the sailors and merchants around the port cities, local dialects which mixed Japanese and English became a way of maintaining necessary communication. The Yokohama Dialect is a representative example of these pidgin languages which developed around the major port areas. It survives due to the publication of Exercises in the Yokohama Dialect, a pamphlet published in 1874 which documents many of the established words and phrases used in the pidgin. Some of the more regular English words entering Japanese from this time are *hoteru*, from 'hotel', and *hausu*, from 'house'; whilst many Japanese words took on English-pronunciation alterations, such as Japanese *samui* 'cold' becoming the common name 'Sammy', and *kokonotsu* 'nine' becoming 'coconuts' (Stanlaw, 2004).

Irwin (2011) documents that Japanese victory in the Sino-Japanese war of 1894-95 most likely indirectly strengthened the position of English as the main foreign language in the country. This came as a result of negative sentiment towards Russia, Germany and France who forced Japan into returning land to China which it had previously claimed in the Treaty of Shimonoseki (Irwin, 2011, p. 55). The generally positive acceptance of English speakers, teachers, textbooks, and literature, along with Western cultural artefacts and scientific processes, continued into the Taisho period (1912-1926), particularly due to the Anglo-Japanese alliance (1902-1923), signed to try to prevent Russian expansion (Loveday, 1996, p. 71). This period is described as "the heyday of Japanese-English language contact and borrowing" and a period which "established patterns of taking, modifying, and creating English vocabulary items and English-language concepts and cognitive schemas which continue to this day" (Stanlaw, 2004, p. 68). It was particularly a time of Americanisation in Japan, with the introduction of radio and cinema "which spread popular American culture and increased the speed and range of transmission" (Loveday, 1996, p. 72).

Gradually, as Japan strengthened its economy and military and began moving into further territorial disputes around Asia, anti-Western sentiment (particularly anti-American) grew in parallel to a strengthening of nationalist feelings. This sees the beginning of a third phase in the history of the influence of English in Japan, with the start of the Second Sino-Japanese war (1937) when the U.S. and Great Britain backed China, and particularly during World War Two when the two countries became direct enemies of Japan. This caused a perspective change, with English becoming the language of the enemy, and efforts were thereby made to remove English loanwords from Japanese. This even included removing Latin alphabet characters such as WC from toilets (Umegaki, 1963, as cited in Irwin, 2011). However, the considerable presence of English throughout Japanese society at this time meant that these efforts were often not particularly successful. This was in large part because replacements for the loanwords came in the form of Sino-Japanese characters which because of their logographic nature were often too formal and confusing. In terms of education, Loveday states that in the run up to World War Two, English classes in the general school curriculum began to be restricted in number, and "almost every university department of English was closed, and scholars were required to direct their knowledge of the 'enemy language' to the war effort" (Loveday, 1996, p. 75).

Defeat in World War Two in 1945 started the fourth phase in language contact between English and Japanese, a phase which continues into the present day. Defeat in the war "shattered faith in the power of Japanism and paved the way for a further intensification of Westernisation" (Loveday, 1996, p. 75). One of the most important driving forces of this renewed Westernisation was the development of another pidgin variety of English called Bamboo English during the American Occupation of Japan from 1945 to 1952. It centred around communications between American military personnel and Japanese shop workers, labourers, and hostesses. This Japanese-English form of communication helped to bring in many English loanwords into the language in this period, along with the American-based cultural artefacts and concepts which they were used to describe. Since the occupation, America has continued to have a strong influence in Japan, with active military bases spread across the country from Okinawa in the south (e.g. Camp Shields), to Yamaguchi in the centre (e.g. Marine Corps Air Station), and Aomori in the north (e.g. Misawa Air Base). This presence, along with global processes of the spread of English-based business, education and entertainment around the world, has meant that Japan continues to embrace and exploit the English language.

Despite English having no official status or major institutionalised role in society (McKenzie, 2008), present-day Japan is both fascinated with and highly dependent on English. From a global perspective, it is the principal communication medium of international trade, travel and academic research (McKenzie, 2008, p. 269) and domestically the government views proficiency in English as an essential skill for modern-day local interaction in a global community (Butler, 2007; Hongo, 2013; Mie, 2013). It continues to be embraced and exploited as the vehicle to carry forward Japan's development, now being for globalisation rather than the modernisation needed in Meiji Japan. Particularly, it continues to be comprehensively integrated into the compulsory and further education systems, with English lessons becoming a standard part of the curriculum from 2011 for students in the final two grades of elementary school, and the government-led proposal that the Test of English as a Foreign Language (TOEFL) becomes a mandatory requirement for entry to university (Mie, 2013).

As the contact with English continues, English-derived loanwords are becoming more thoroughly integrated into the internal workings of Japanese language and society; with clothes, product packaging, and TV adverts including decorative and emblematic English borrowings (Dougill, 2008; Hyde, 2002); and everyday conversations, public signs and news articles incorporating nativised English loanwords (Backhaus, 2007; Stanlaw, 2004). In such an environment, "English in Japan is like air: it is everywhere" (Stanlaw, 2004, p. 1). These borrowings flow into Japan through various channels, with the overwhelming majority entering through distant communications such as domestic translations of books and articles, internationally-produced language learning materials, and Internet-based news websites, social networking services, and digital entertainment platforms (Hoffer, 2002, p. 13). Only a very limited number have arrived from direct people-to-people contact (Irwin, 2011). It is this combination of a huge number of loanwords borrowed over a relatively short period of time with very limited direct contact that has made the Japanese-English language contact situation particularly appealing for academic research. Indeed, the impact of English on the Japanese language cannot be underestimated: "the scale of Japan's borrowing of English is virtually unparalleled in the world" (Daulton, 2008).

2.2.4 *The contemporary Japanese language*

The dynamic mix of native and borrowed elements which make up the present-day Japanese language help to make it "one of the most intensely studied languages in the world" (Shibatani & Kageyama, 2016, p. vii). Whilst historically speaking, Chinese has "continuously contributed to Japanese in an immeasurable way affecting all aspects of grammar, most notably the lexicon and the phonological structure" (Shibatani & Kageyama, 2016, p. xi), in modern times "a distant and exotic language – English – has extensively and fundamentally transformed the Japanese lexicon" (Daulton, 2008). This is clearly evident in Japanese society in the 'linguistic landscape', meaning the "linguistic objects that mark the public space" (Ben-rafael, Shohamy, Amara, & Trumper-Hecht, 2006, p. 7). In this landscape, the various scripts used to inscribe different types of words reflect a history of over nearly 2,000 years of language contact and lexical borrowing.

One of the most distinguishing features of the modern-day Japanese language is the regular and systematic mixing of words from different lexical strata, or as Kageyama and Saito describe it in their comprehensive analysis of Japanese word formation, "a number of different vocabulary strata stacked on one another" (2016, p. 12). As discussed in the previous section, these strata have developed out of the long history of contact with, and borrowing from, other cultures and their languages. Onto the foundational base of native Japanese words (*wago*) an immense number of Chinese-derived words have been introduced, many of which are direct borrowings from Chinese, whilst others are domestically-produced words based on Chinese models. Over centuries of development, these Chinese borrowings have developed into the lexical stratum of Sino-Japanese words (*kango*). Added to this already complex vocabulary mix are the massive number of non-Chinese loanwords (*gairaigo*), borrowed from European languages such as Portuguese, Spanish, Dutch, German, French, and most notably in modern times, English. Within the Japanese language these strata dynamically interact, not only at the level of the sentence, where words from different strata syntactically combine, but also at the level of the word, so that hybrid words called *konseigo* are produced from a combination of lexical elements from any of the strata. The word *keshigomu*, for example, meaning 'eraser' in Engish, is made up of a combination of the native word *kesu* meaning 'extinguish' and the Dutch loanword 'gomu' meaning 'rubber'; and the word *meiwakumeeru* meaning 'junk mail' blends the Sino-Japanese word *meiwaku* 'nuisance' and the English loanword *meeru* 'mail'.

Taking a broad overview of word types, native Japanese vocabulary contains many of the words relating to practices of traditional importance in Japanese life, such as those connected with agricultural farming and fishing industries. These include words for staple foods (e.g. *kome* 'rice' and *sakana* 'fish'), weather phenomena (e.g. *harussame* 'spring rain' and *kumo* 'cloud'), and areas of farming (e.g. *hama* 'beach' and *ura* 'bay') (Shibatani, 1990, pp. 140–141). *Wago* are also used as the grammatical or function words which help structure a Japanese sentence, such as particles and honorific prefixes. Sino-Japanese words, or *kango*, are the words borrowed from China, when it was seen as a culturally advanced nation from which Japan could learn and prosper. Traditionally, *kango* words were related primarily to scholarly writings and official court documents, whilst over time they have become standard parts of the general Japanese language. This has been particularly the case since the Meiji restoration (1868-1912) where Chinese readings of *kanji* characters were used to translate English terms relating to modernisation processes (Shibatani, 1990, p. 145). In this way, the foreign vocabulary strata, *gairaigo*, has been an outcome of the fact that "technological superiority and economic dominance have come to determine the flow of words across linguistic boundaries" (Shibatani, 1990, p. 140). Over time, the traditional cultural influence from China has come to be replaced by a modern globalisation influence, particularly related to America, meaning that "contemporary Japan is thus inundated with loan words from English and other European languages whose native speakers reside on the far side of the globe" (Shibatani, 1990, p. 140). Examples can be drawn from every aspect of Japanese society, from food terms such as *chiizu* 'cheese' and *beekon* 'bacon', to technology terms such as *onrain* 'online' and *chaajyaa* 'charger', to fashion terms such as *ankurupantsu* 'ankle pants' and *kajuaru shatsu* 'casual shirt'.

In terms of distribution of lexical strata in the Japanese language, the most cited data in the academic literature is that produced by the National Institute for Japanese Language and Linguistics (NINJAL), in their comparative analysis of words in Japanese magazines in 1956 (NINJAL, 1962, 1964) and 1994 (NINJAL, 2005). Table 2.2 below summarises these findings, and is adapted from Irwin (2011, pp. 16–18). A particularly revealing finding is the large increase in the number of both *gairaigo* tokens (i.e. number of times the word appears) and types (i.e. number of different words) over the forty-year period; although importantly Irwin also states that most *gairaigo* have very low frequency rates, with the majority appearing between only one and four times throughout the total token count (2011, p. 17). In his meta-analysis of the proportion of *gairaigo* tokens across various Japanese media types spanning the years 1906 to 2006

(clustered into categories such as magazines, newspapers, school textbooks, pop lyrics, and newspaper letter columns), Irwin further observes "the inexorable rise in the proportion of *gairaigo* vocabulary across most media since the first survey providing token data was carried out in 1906" (2011, p. 20).

Table 2.2 Lexical strata distribution in magazines (adapted from Irwin, 2011, pp. 16–18).

	Token % 1956	**Token %** 1994	+/-	**Type %** 1956	**Type %** 1994	+/-
Native	54	37	- 17	37	25	- 12
Sino-Japanese	41	49	+ 8	47	34	- 13
Foreign	3	12	+ 9	10	35	+ 25
Hybrid	2	2	+/- 0	6	6	+/- 0
(Total)	100	100		100	100	

Strongly associated with these strata are the three standard Japanese orthographies of *kanji*, *hiragana*, and *katakana*. At a fundamental level, *kanji* are for writing *kango*, *hiragana* for *wago*, and *katakana* for *gairaigo*; but in practice the writing of Japanese displays the same creativity as seen in the dynamic interactions between different lexical strata (Barrs, 2011). For example, *katakana* is frequently used to add emphasis to words from any of the strata, the same way in which italics is used in English writing, and *hiragana* is used to simplify Japanese writing, particularly in children's books and Japanese as a foreign language (JFL) textbooks. Furthermore, words formed by compounding elements from more than one lexical stratum, the aforementioned *konseigo*, can employ more than one script. An example is the word *oshipin*, meaning 'push pin' or 'thumb tack' in English, formed from the combination of the native word *osu* 'push' and the English loanword *pin* 'pin' and often written as 押しピン. This single word employs all three main orthographies: the *kanji* character 押 (*o*), the *hiragana* character し (*shi*), and the *katakana* characters ピン (*pin*).

2.3 Types of English in the Japanese Language

Over the past thirty years there have been numerous models proposed of how to categorise the different varieties of English around the world, known in linguistic research as World Englishes. These include 'A Map and Branch Model' (Strevens, 1980), 'A Circle of World English' (McArthur, 1987), 'A Circle of International English' (Gorlach, 1990), and 'Three Concentric Circles of English' (B. B. Kachru, 1988). The principal classification underlying all the models is a distinction made between English as a Native Language (ENL), English as a Second Language (ESL) and English as a Foreign Language (EFL). Kachru (1990) explains his development of the idea of the Inner, Outer, and Expanding circles of global English usage as an important step in moving away from the traditional and ethnocentrically-generated idea of a single monolithic world English to the more inclusive, culturally appropriate and pluri-centric idea of various world Englishes.

In the expanding-circle of English usage, within which Japan is included, English is used primarily as a foreign language, and as a performance variety with functions being largely restricted to those of an interpersonal and instrumental nature (Berns, 2005). Because English typically has very limited or no institutionalised functions in Expanding Circle contexts, it is viewed as being used for its creative power alongside or often inside the native language, in areas such as advertising (e.g. product packaging and billboard adverts), the mass media (e.g. magazine front covers and newspaper articles) and entertainment (e.g. song lyrics and TV subtitles) (McArthur, 1987).

There is debate about whether there exists an identifiable Japanese variety of English (McKenzie, 2008). Honna and Takeshita (1998), for example, argue for recognising Japanese English as a variety in its own right, and state that "Japanese English can be an important utilitarian language. It is part of various Non-Native Speaker Englishes used in many functions in many parts of Asia and the rest of the world" (1998, p. 126). However, it is the fact that unlike in countries where English has an institutionalised role in the country and has developed into distinct local varieties (i.e. with the Outer Circle varieties of Singapore English and Indian English), in Japan, English "has no internal reason for its promulgation; the medium of instruction in all public institutions and of all government business is Japanese" (Y. Kachru & Nelson, 2006).

In general, research into English in Japan is not commonly approached from a World Englishes perspective, but rather from a language contact perspective, with

attention given to how the two languages have influenced each other. This has meant that instead of research focusing on identifying and describing specific features which would characterise Japanese English as a distinct variety, it has focused more on areas that are typical of the fields of contact linguistics and language contact, such as Japanese-English code-mixing and switching (Nishimura, 1995), ideologies of English in Japan (Seargeant, 2005, 2008), the history of Japanese-English language contact (Miller, 1967), phonological and morphological adaptations to English words in the Japanese context (Irwin, 2011), and English's impact on issues of speaker-identity (Morrow, 1987). Because of this, there is no commonly agreed-on typology of English in Japan and instead different studies categorise and name the forms and functions of English in different ways.

The various studies agree, however, that the English found in modern-day Japan, in its myriad forms and functions on shop signs and road markings, in newspaper articles and school textbooks, and in daily conversations and government speeches, can broadly be divided into two types: English used inter-nationally and English used intra-nationally. The former type is used separately from Japanese as a foreign language. Typically, it is written in the Latin alphabet, spoken with a pronunciation that mirrors inner-circle varieties such as British and American English, and used in largely comparable ways to these standard varieties. The latter type, in contrast, is integrated into the Japanese language. As such, it typically undergoes script changes, phonological alterations, and morphological re-modellings which mask the historical English language origins not only to non-Japanese speakers but also Japanese speakers themselves (Daulton, 2004, 2008). Mixed in with this latter type of English are the words and phrases known in Japanese as *wasei eigo*, ‘made-in-Japan English’. These are English-like expressions which are created and used domestically from novel combinations or morphological alterations of English vocabulary.

The aim of the following sections is to discuss in more detail this broad division of intra- and inter-national English, and the types of English which constitute each side of the division, so as to operationally define for the present study the term English loanwords in Japanese. It will be shown that there are different forms of English loanwords in Japanese, making it necessary to explain the loanword forms which will and will not be included in the later analysis.

2.3.1 *International and intranational English in Japan*

Looking at previous research on the influence of the English language in Japan, several different terms can be found for the division discussed above between English used inter-nationally and intra-nationally. These are presented in Table 2.3 with Caesar (2005) offering arguably the most transparent terms of English of the classroom and English of the streets. As is the norm of global communications, Japan uses English as the principal language of international communication (Backhaus, 2005; McKenzie, 2008). In this form, known in Japanese as *eigo* 'English language' or more colloquially as *eikaiwa* 'English conversation', Japan looks almost exclusively to the inner circle models provided by the US and UK, with a dominance of British English during the Meiji restoration followed by an ascendency of American English after the end of World War Two and the country's occupation by American forces. This form of English is used separately from native and Sino-Japanese words, and appears in contexts such as language learning textbooks, classroom-based and online language learning lessons, university entrance exams, travel guides, road signs, and English language newspapers. As such, it is orthographically, phonologically, morphologically, and semantically the same as the English used throughout the world for the purposes of cross-cultural communication, often termed English as an International Language (McKay, 2002).

Table 2.3 A conceptualisation of two principle varieties of English in Japan.

	International English	**Intranational English**
Caesar (2005)	English of the classroom	English of the streets
Dougill (2008)	Functional English	Decorative English
Kirkpatrick (2007)	American native speaker-model English	Japanese English
McArthur (1998, 2002)	English	Decorative English/ Domesticated English (Gairaigo)
Morrow (1987)	English in the education system	Japanese variety of English (Japalish)
Nuttall (2000)	English as an examination subject	Janglish
Stanlaw (1987, 2004)	Japanese variety of English	Japanized English

For most of the Japanese population, international English is studied and used as a foreign language, with individual proficiency varying significantly depending on age, education, gender, employment, and motivation to learn. Overall, the general level of English ability of the average Japanese person is considered comparatively low in relation to the economic status of the country and the amount of time and money expended on English education (Gottlieb, 2005; Irwin, 2011; Loveday, 1996; Stanlaw, 2004). Figure 2.1 shows an example of International English, used on a sign in a clothing store.

Figure 2.1 The use of *eigo* (English as an International Language) on an information board in a Japanese branch of an international clothing store.

Intranational English in Japan is the English which has been integrated into Japanese-language communicative practices. This involves the borrowing of English vocabulary and its adaptation to the rules of the Japanese linguistic system. This is what the term *gairaigo*, or '(English) loanword in Japanese', is most commonly taken to mean: the words and phrases derived from English and used in the Japanese language. Because it is an integral part of the Japanese language, this intranational English has permeated all aspects of Japanese society, and is seen on signs, product packaging, road markings, and building names, and heard in public announcements, TV commercials, music lyrics, and radio broadcasts (Daulton, 2008; Stanlaw, 2004). These loanwords, which include tens of thousands of words, acronyms, and initialisms, represent by far the most extensive and prominent outcome of English-Japanese language contact.

Academic interest has particularly been attracted to this lexical borrowing from English because of the fact that it does not sit comfortably within traditional theoretical models of contact-induced language change (Irwin, 2011; Stanlaw, 2004). These models, such as Thomason's (2001) borrowing scale, suggest that because the overwhelming majority of Japan's contact with English has occurred at a substantial distance, primarily via an orthographic medium and without colonisation by English-speaking nations or extensive bilingualism in the native population (Irwin, 2011, p. 3), there should only be a limited amount of borrowed vocabulary. The reality however, is that Japan has in fact borrowed a remarkable number of English nouns, along with a more limited number of verbs, adjectives, function words, and affixes, and many of these have been integrated into the core of the modern Japanese lexicon (Daulton, 2008; Inagawa, 2010; Irwin, 2011; Kay, 1995).

Indeed, in terms of the rate at which these words have been integrated into the Japanese language, and the general levels of receptivity towards them throughout society, Japan's lexical borrowing from English is considered unique, possibly unmatched in the modern world (Daulton, 2008). The borrowing has been so prolific that many specialised resources have been published which catalogue and categorise the forms and functions of the loanwords. These include dedicated loanword dictionaries with as many as 54,000 entries, such as Sanseido's Concise Dictionary of Katakana Words (2010), government-funded research reports into intelligibility rates of the loanwords in daily society (NINJAL, 2004), and loanword lists for use in educational contexts (Daulton, 2008).

With both of these types, English exists in relation to the Japanese language both extra-lingually and intra-lingually. These terms are used here to make a distinction between whether English is used separately from Japanese words (i.e. extra-lingual) in utterances and sentences, or whether it is mixed in (i.e. intra-lingual). These uses of English can be related to the linguistic concepts of code-switching and code-mixing, whereby languages are separated out from one another and used interchangeably (code-switching) or mixed into one another and used interconnectedly (code-mixing) (Myers-Scotton, 2006). When a distinction is also made between the spoken and written registers then eight combinations become possible, as shown in Table 2.4. With international English, extra-lingual uses include translations of Japanese words on signs (see again Figure 2.1 above) and in textbook-based conversations in English language classes. Intra-lingual uses include English words un-adapted to Japanese (i.e. code-switches rather than loanwords) which are mixed into Japanese utterances and sentences

by Japanese-English bilingual speakers (i.e. code-mixing). With intranational English, extra-lingual instances include its decorative uses in lyrical messages printed on T-shirts, and in spoken advertising slogans in Japanese TV commercials. Intra-lingual uses include the use of English-based words mixed in with Japanese vocabulary on billboard signs and in Japanese conversations.

Despite these seemingly well-defined distinctions, usages of English in Japan, like language usage in general, are open to different interpretations and whilst it may be easy to categorise some aspects of English usage, others may be more blurred and lie somewhere on the borders between the categories shown in Table 2.4. This is especially the case because of the way in which that all of the different combinations, when written, can appear in any of the three main and two additional scripts used in Japanese (Barrs, 2011; Stanlaw, 2004). This can make the categorisation of each combination complicated, especially as to whether an English word written in the Latin alphabet is a borrowed word which has been integrated into Japanese (i.e. loanword) or a foreign word that remains un-adapted to Japanese linguistic rules (i.e. code-switch) (Irwin, 2011; Loveday, 1996).

Table 2.4 Extralingual and intralingual uses of international and intranational English in Japan, in the spoken and written registers.

	Extra-lingual Uses		**Intra-lingual Uses**	
	Spoken	Written	Spoken	Written
International English	Conversations in English-only classes	Translations on information boards	English code-switches in bilingual conversations	English code-switches used on Japanese advertising posters
Intranational English	Advertising slogans in TV commercials	Poetic messages printed on T-shirts	English loanwords in monolingual conversations	English loanwords in Japanese newspapers

The broad categorisation seen in the literature of two distinct varieties of English in Japan is not only made at the functional level. There is evidence in the literature of an attitudinal distinction of the uses of English in Japan, where intranational uses are considered creative, and the international variety is seen as standard. For example, Stanlaw (2004) describes Japanized English as a “creative force in Japanese

sociolinguistics and various forms of artistic expression" (p. 299) while he sees the Japanese variety of English primarily as an academic subject modelled on British and American varieties (p. 287). The creative uses of English in Japan are sometimes viewed negatively, as a form of deviance from a supposed Standard English and a sign of linguistic deterioration or pollution of the Japanese language (Crystal, 1997). This point will be returned to again in Section 2.4.2.

2.3.2 *Varieties of international and intranational English in Japan*

Characterising the broad divisions made above between international and intranational, and extra-lingual and intra-lingual uses of English in Japan, several specific and regular uses of the language can be identified and categorised as typical varieties of English usage in Japan. For the discussion below, rather than using the division between international and intranational English just discussed, a division between Latin alphabet English and *katakana* English is used. This division focuses on the written language and allows for a clearer identification of the different forms and functions of English in Japan.

One of the varieties of English written in the Latin alphabet in Japan can be described as Decorative English, an application of the English language where the function is more stylistic than linguistic. Decorative English can be found on clothing, product packaging, sign boards, posters, menus, and magazine covers. This practice is not exclusive to Japan, but rather seen globally because of English's desirable associations with modernity, globalisation, fashion, and Western-ness (McArthur, 1987; Stanlaw, 2004). Figure 2.2 shows an example of decorative English on a sign board in a clothes store, where the accuracy of the message appears less important than the stylistic effect created by the use of English words.

Figure 2.2 A sign in a Japanese clothes store showing the use of English as decoration.

Advertising English is a variety similar in function to decorative English, but with a particular focus on drawing the reader's attention towards what is being advertised. It can be found throughout the linguistic landscape, used in the advertising of products as diverse as cans of beer, sports equipment, movies, cosmetics, holidays, and snacks. This English may or may not be understandable to the reader, depending on various factors such as the linguistic accuracy of the message and the language ability of the intended receiver of the message. Figure 2.3 shows an example of advertising English in a Japanese sports store where English has been used in place of Japanese to demarcate a particular area of the store, but attention has not been paid to achieving full linguistic accuracy in the use of the English words.

Figure 2.3 A sign in a Japanese sports store showing the use of English for advertising.

Translation English is a variety intended to assist people unable to read Japanese, in order to communicate a specific linguistic message. The accuracy of the English varies greatly depending on how it was produced (e.g. through the use of translation software in lieu of consultation with an English speaker), and this subsequently impacts on whether or not this usage of English achieves its purpose of communicating a translation of the original Japanese message. Translation English is found wherever it has been thought necessary to provide a translation for a non-Japanese audience, such as train station platforms, restaurant menus, local government office pamphlets, and ATM screens. Figure 2.4 shows the control panel of a bidet toilet in Japan (a toilet with built-in washing functions) with the main controls translated from Japanese into English.

Figure 2.4 A control panel of a Japanese bidet toilet with the main functions translated into English.

One other major variety of Latin alphabet English in Japanese society is Academic English. This is the variety studied in the curriculums of elementary schools through to universities, and in private language school classrooms and online courses. This a norm-dependent variety of English borrowed directly from inner circle countries such as America and Britain. Traditionally, it has been studied as an academic subject through a grammar-translation methodology, and this remains the dominant practice throughout the school system (Fukada, 2010). However, in recent years the Japanese government has been promoting a focus on developing communicative competence in English, in order to create what it calls a global workforce of the future (Mie, 2013). As an academic subject, English in Japan differs little from many other countries where

English language learning is seen as a tool of opportunity, progression, and development in a globalised world.

As was discussed in in Section 2.2.4, English very frequently appears in Japanese in the *katakana* script, most typically used to write the loanwords which have been integrated into the Japanese language. One variety of *katakana* English is *katakana* advertising English. Very often this use of English has an equivalent expression available in the native or Sino-Japanese strata, meaning that *katakana* advertising English is often used for its stylistic effect more than its semantic denotation. Because of this, the literal meanings of the words used in this variety of English may not always be well understood by Japanese people. Figure 2.5 shows the header of a website of a popular Japanese clothing store using the English words 'bra-feel' and underneath the same two words written out as ブラフィール in *katakana*. This term, referring to a women's shirt shaped to act and feel as a bra, is not found at all in the jpTenTen11 8-billion-word web corpus (see Section 4.3.1) and as such seems to be a word domestically created by the clothing company from English elements (i.e. *wasei eigo*, see Section 2.3).

Figure 2.5 The header of a Japanese company's website using the word ブラフィール.

Another variety is *katakana* core English, where the loanwords have become frequent, indispensable parts of the Japanese lexicon and often have no alternative expressions in the other lexical strata. An example is the loanword ページ *peeji* 'page', a highly frequent word in Japanese with no viable semantic equivalent.[2] The use of *peeji* is shown in Figure 2.6 below inside the box in the phrase 詳しくは 02 ページへ *kuwashiku wa 02 peeji e* 'for more details, see page 2'. This type of English usage can be found throughout Japan in every aspect of society, and is so common that regular

[2] Whilst there is no viable semantic equivalent in the native or Sino-Japanese lexical strata for the loanword *peeji*, the *kanji* character 頁 is often used in place of the *katakana* graphemes.

daily Japanese communication would become extremely difficult, if not impossible, if all of these loanwords were to suddenly disappear from the language (Hoffer, 2002; Stanlaw, 2004).

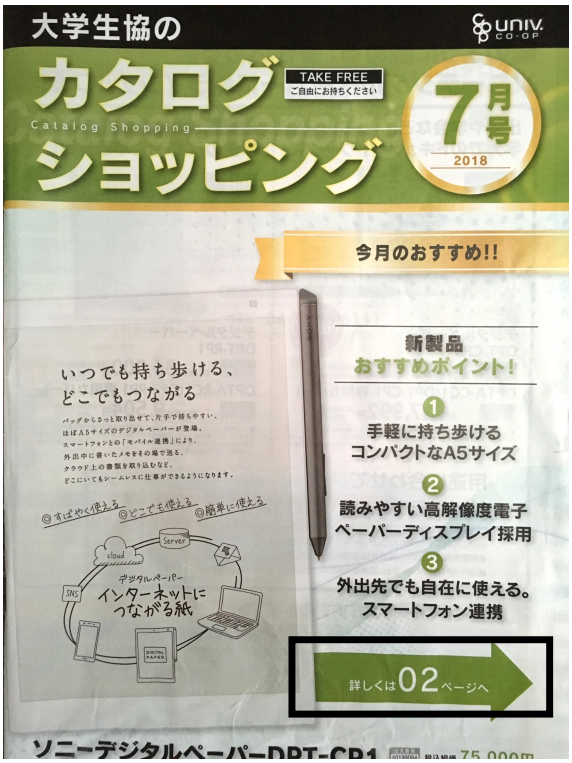


Figure 2.6 The front cover of a shopping catalogue showing various uses of *katakana* core English.

2.4 English Loanwords in Japanese

As discussed in Section 2.2.3, most of the imported English vocabulary in Japan has been borrowed post-World War Two and has flowed into society via mass media communications, such as newspaper and magazine publishing, the music and movie industries, and internet-based content. Thus, the borrowing from English has not been owing to a large community of bilingual speakers, but instead to the country's contact with English-speaking nations during historical periods of industrialisation and modernisation, and the modern-day global influence of English as an international language (Hoffer, 1990; Shibatani, 1990). English loanwords are by far the largest contributing factor to the expansion of the modern Japanese lexicon, not only in specialised fields such as medical science and computer technology, but also general

domains including daily conversations and product advertising (Daulton, 2008; Irwin, 2011; Stanlaw, 2004). They also make up the overwhelming majority share of the *gairaigo* strata (non-Chinese loanwords) of the Japanese lexicon. Their vast number in the Japanese lexicon and considerable extent of integration into Japanese syntax have helped facilitate their academic study, and have contributed to investigations such as the history of contact and lexical borrowing between Japanese and English (Loveday, 1996); the linguistic adaptations made to English words as they are borrowed into Japanese (Irwin, 2011); the impact on English education in Japan (Daulton, 2008); and the effect on the Japanese linguistic landscape (Backhaus, 2007).

The omnipresence and pervasiveness of English loanwords throughout modern Japanese society is fostered by highly productive processes of morphological, phonological, orthographical, and semantic adaptations, allowing the borrowings to be deeply integrated into the workings of the Japanese language (Irwin, 2011; Shibatani, 1990). This has resulted in a sizeable sub-set of English loanwords becoming highly-frequent lexical items in the everyday communicative practices of Japanese speakers.[3] As an example of this deep integration, Figure 2.7 shows the home screen of a copy machine installed in Seven Eleven convenience stores throughout the country.[4] Alongside several words written in the English alphabet ('English', 'print', and 'big'), seven of the nine main selection options contain an English borrowing encoded in the Japanese script of *katakana* (コピー *kopii* 'copy', プリント *purinto* 'print', スキャン *sukyan* 'scan', ファクス *fakukusu* 'fax', チケット *chiketto* 'ticket', プリペイドサービス *puripeidosaabisu* 'pre-paid service', スポーツ *supootsu* 'sports', and サービス *saabisu* 'service'). The fact that these loanwords appear on the home screen of a machine installed in tens of thousands of stores throughout the country is a readily observable example of the common-place occurrence of English loanwords in Japanese society.

[3] The term Japanese speakers is used broadly to include speakers of Japanese as a native language, second language, and foreign language.

[4] As of the end of July 2018 there were 20,437 Seven Eleven stores in Japan. Copy machines are installed in all but a small number of these stores, according to the Seven-Eleven website.

Figure 2.7 The home screen of a ticket machine installed in Seven-Eleven convenience stores in Japan.

2.4.1 *Linguistic adaptations to English loanwords*

This section overviews the main adaptations which English words undergo when borrowed into Japanese. The purpose is to lay the foundations for the discussions of the integration of English words into Japanese which take place in Chapter Three and Chapter Six. The modifications have previously been well-described (see for example, Irwin, 2011; Quackenbush, 1977), and what follows is a brief summary of the main adaptation processes discussed in Irwin (2011).

For an English word to be integrated as a loanword in the Japanese language, it needs to be structurally conformed to the norms of the Japanese linguistic system. In lexical borrowing research, four principle types of structural adaptation have been recognised: phonological, orthographical, morphological, and syntactic (Matras, 2009; Winford, 2003; Winter-Froemel, 2008; Zenner & Kristiansen, 2014). Phonological and orthographical adaptations are concerned with how loanwords are represented in speech and writing in the recipient language (RL). Phonological adaptations include segmental changes, such as the replacement of source language (SL) phonemes that are disallowed in the RL phonological system with the closest RL equivalents; and suprasegmental changes, such as adaptations made to syllable structure, tone and stress. These changes may also need to be reflected in orthographical adaptations; and if the SL and RL use different orthographies, complete graphemic adaptation will be needed for full orthographic integration. Morphological and syntactic adaptations are concerned with

the grammatical behaviour of loanwords in the RL, in how they are typically organised in grammatical relationships with other words. Adaptations which an SL lexical item may undergo for its grammatical integration into an RL include the assignment of gender and number for it to conform to the RL morphological system, and a possible change in its part of speech.

Taking the common English loanword アナウンス *anaunsu* 'announce' in Japanese as an example, the differences between the two languages' inventories of phonemes, graphemes, and morphemes have resulted in multiple adaptations being necessary to make the original English form fit into Japanese. Phonological adaptations have been made with the word-initial schwa sound of /ə/, for example, disallowed in Japanese phonology, being replaced with its closest equivalent of /a/. Also, the consonant-ending /s/, also disallowed in Japanese, is appended with an epenthetic vowel to make the ending phoneme /*su*/. In its orthography it has been completely adapted to be regularly written in *katakana* script instead of the Latin Alphabet. Morphologically speaking, the adaptations are more complicated to describe because, as Irwin discusses (2011, pp. 137-138), it is unclear whether the Japanese form of *anaunsu* was originally borrowed directly from the English verb 'announce', or via the English noun 'announcement'. If the former is the case, then the fact that *anaunsu* is a noun in Japanese, and then made into a verb with the addition of the Japanese verbaliser *suru*, means that its word class was changed from a verb in English to a noun in Japanese. This would then be more of an aspect of syntactic integration than a morphological one. If, however, *anaunsu* was borrowed from the English noun 'announcement', it instead shows morphological reduction in that the latter morpheme was clipped.

Looking more specifically at the individual types of modifications, the main adaptations made to English words for their integration into Japanese are in the area of their phonological structure. Irwin (2011) divides these into substitution (the replacement of sounds), epenthesis (the addition of sounds), and deletion (the removal of sounds). Substitution occurs with both consonants and vowels and happens because of the limited phonemic inventory of Japanese compared with that of English. A major type of consonant substitution is with both /l/ and /r/ in English becoming /r/ in Japanese, turning English words like 'lock' into *rokku* in its Japanese form (Irwin, 2011, p. 91). This can then cause potential confusion in that the two English words 'lock' and 'rock' take the single form of *rokku* in Japanese. Epenthesis typically involves the addition of vowels in order to open closed syllables and to break up consonant clusters. An example of the former is the addition of 'u' to the end of English words, such as

guruppu ‘group’, *chiimu* ‘team’, and *piggu* ‘pig’; and an example of the latter is ‘u’ breaking up the consonant clusters in words like *sukuuru* ‘school’, *furasshu* ‘flash’, and *gureeto* ‘great’. Deletion of English sounds most typically occurs at the end of a word, where the consonant is deleted. This can be seen with words such as *roosu* ‘roast’, *daasu* ‘dozen’, and *ramune* ‘lemonade’ where the word-ending consonant has been deleted.

Because the *katakana* script used to write English loanwords is a ‘shallow orthography’, with a very strong correspondence between the phonemes of the spoken language and the graphemes of the written script, the phonological adaptations described above are mirrored in the transcription of English words in *katakana*. For example, the *katakana* grapheme ト, represented in English as ‹to›, is used for the phoneme /to/ which appears on the end of the loanword グレート *gureeto*, but is then deleted from the end of the English word ‘roast’ to make ロース *roosu* rather than ロースト *roosuto*. [5]

In terms of morphology, adaptations can render some loanwords very different from their source words, such as インフレ *infure* truncated from インフレーション *infureeshon* ‘inflation’. Truncation is common in English loanwords because of the previously described vowel epenthesis which often renders words much longer in syllabic length than the English source word and is particularly common with jargon, youth speech, slang and dialect (Irwin, 2011). An example of truncation is with how リストラクチャリング *risutorakucharingu* ‘restructuring’ is rendered with 8 syllables (or 9 morae) in Japanese but only 4 syllables in English. Therefore it is commonly truncated to リストラ *risutora*. As with this example, sometimes loanwords are truncated to an impermissible form in English which can be deceptive for learners. In production they may transfer over the truncated form, and in comprehension they may fail to recognise any similarity (Daulton, 1999; Ringbom, 2007b). Another observed morphological adaptation is the reduction of native morphology, where grammatical elements such as plural and gerund are removed. An example of the former is the loanword パジャマ *pajama* with the plural ‘s’ removed, and an example of the latter is スキー *sukii* for the English word ‘skiing’.

[5] Interestingly, ト is often retained when the English word ‘roast’ is used in a compound noun in Japanese, such as ローストビーフ *roosutobiifu* ‘roast beef’.

Semantic modifications can also take place and can restrict or extend the meaning of the original English word. The loanword カンニング *kaningu* 'cunning', for example, restricts the meaning of the English word to mean cheating on an exam in Japanese, whilst the range of meanings of the loanword ハンドル handoru 'handle' is extended to include a car steering wheel. It is the semantic modifications made to some English words when they become loanwords in Japanese that are considered to be especially problematic to Japanese learners of English (Daulton, 2008; Stanlaw, 2010).

2.4.2 *Social attitudes to English loanwords in Japan*

As English has flowed in, it has for the most part been welcomed and exploited. It has long been seen as a tool of opportunity with which the Japanese nation can strengthen and modernise, and in modern times, globalise. It is the principal second language studied within the country and is a major element of compulsory school education, university admissions, higher education classes, company recruitments, career promotions and overseas job-postings (Mie, 2013; Tanikawa, 2013).

Despite this, there have been some concerns about the negative effects of English loanwords on Japanese society. Such concerns are particularly strong in relation to English language education, where Japanese students begin studying English already with a pre-installed sub-lexicon of English loanwords in their native language (Barrs, 2012a; Brown, 1995; Daulton, 2008; Ringbom, 2007b). In relation to the importance of cross-linguistic similarities in language learning (Ellis, 1994; Nation, 1990; Odlin, 1989; Ringbom, 2007a), this English sub-lexicon is regarded as having a pedagogical influence on English language education (Barrs, 2013b, 2013a, Daulton, 2008, 2011; Ringbom, 2007a; Stubbe & Yokomitsu, 2012). However, judgements on the nature of this influence, on whether it is facilitative or disruptive, differ greatly. Many maintain that English loanwords are a hindrance, causing confusion in relation to phonological and semantic differences between the loanword in Japanese and the source word in English (Martin, 2004; Shepherd, 1995; Simon-Maeda, 1995; Walker, 2009). Many others take a very different view and see the borrowings as a helping learners, with the similarities in form and meaning between the words particularly helping to lower the effort, or learning burden, associated with the acquisition of new vocabulary (Barrs, 2012b; Daulton, 1999, 2003, 2009; Kistler, 1995; Ringbom, 2007a).

In society in general, although English loanwords are now ingrained into Japanese culture and embraced especially by the media, advertising industry and younger generation of citizens, some of these loanwords are felt by some to have a damaging effect on Japanese society and language (Chavez, 2009; Nuttall, 2000; Osaki, 2013). This is not a new concern but rather can be seen at different periods in the history of Japanese contact with English. Writing in post-war Japan, Miller, for example, seems less than favourable about the number and form of English loanwords in Japanese, and sees a "growing contamination of older Japanese behaviour patterns with foreign models" (1967, p. 250). This view could be explained by the fact that at the time of his writing, in the early to mid-1960's, the Japanese language was undergoing one of the most intensive historical periods of lexical borrowing from English. He was witnessing a rapid change in the language in the number of loanwords, and saw this as "the principle of 'total availability'" where "virtually any English word in the book is fair game in writing or public speaking" (1967, p. 249).

In more recent times, concerns over the number of English loanwords in Japanese has motivated the Japanese government to conduct official surveys on the public's comprehension of loanwords (Irwin, 2011). It then took steps to offer possible native replacements for the English loanwords using words already present in the native and Sino-Japanese lexical strata (Irwin, 2011). The replacements themselves, however, have often been criticised for being harder to understand than the original loanword. Irwin (2011) for example, discusses how the government's suggestion that the loanword *infoomudo consento* 'informed consent' should be replaced with the Sino-Japanese word 納得診察 *nattokushinsatsu*, which translates literally as 'consented medical examination', was considered inappropriate because it does not contain any meaning of being 'informed' (Irwin, 2011, p. 206).

2.5 Chapter Summary

This chapter has given a historical account of Japan's contact with, and borrowing from, the English language; a situation arising out of its rich history of contact with a number of Asian and European countries. It has been shown that Japan's extensive borrowing of English words since the end of World War Two has been the single most significant change to the Japanese language in its modern history. The modern-day presence of English in Japan was categorised into inter-national English, used primarily for

international relations, and intra-national English, used as an established part of the Japanese language. Furthermore, it was explained how English in Japan can appear in its written form in both the Latin alphabet and the *katakana* script, allowing a vast amount of creativity in how English can be displayed in the Japanese linguistic landscape. This discussion was used to define the type of English in Japan which is to be researched in the present study: the English words which have been integrated into the Japanese language and used frequently within the general language. An explanation was then given of the main features of English loanwords in Japanese, in particular the adaptations made to English words for them to become loanwords in Japanese, as well as an overview of society's generally favourable acceptance of such a large number of foreign borrowings in the native language.

The next chapter looks in detail at how English loanwords in Japanese have been researched, and in particular at the reasons why their grammatical behaviour has so far been for the most part unexplored. It discusses how several practices in the general field of loanword research can be seen to have contributed to this lack of attention on the grammar of loanwords. The chapter then presents the research questions which guide the subsequent corpus analysis of the grammatical behaviour of English loanwords in Japanese.

3 Literature Review

3.1 Chapter Overview

In reviewing previous studies in the field of loanword research, this chapter aims to account for why the grammatical behaviour of English loanwords in Japanese remains a largely unexplored area of research. The chapter begins by defining the terms linguistic borrowing, lexical borrowing, and loanwords for how they are used in the present study. A general overview is then given of the main areas of loanword research over the last century, focusing on three practices in particular which are seen to have contributed to a lack of attention given to the grammatical behaviour of loanwords: a predominance of studies on loanword phonology, a preoccupation with determining the lexical category of loanwords, and a reliance on dictionaries as the main lexical resource for evidence in loanword research. Following this discussion, previous treatments of the grammatical behaviour of English loanwords in Japanese are reviewed, showing how they remain largely theoretical and intuition based. The chapter ends with an exploration of how this grammatical behaviour could be explored empirically, leading to the presentation of the overall aim and research questions of the study.

3.2 Defining Linguistic Borrowing, Lexical Borrowing, and Loanwords

Whether close or distant, language contact always results in change (Hickey, 2010, p. 7). This phenomenon of contact-induced language change is studied within the field of contact linguistics, and aims to study "the varied situations of contact between languages, the phenomena that result, and the interaction of linguistic and external ecological factors in shaping these outcomes" (Winford, 2003, p. 5). The outcomes of language contact include a move from communication in a minority language to that of a dominant language (e.g. language shift and language death), the mixing of two or more languages in the same communicative act (e.g. code-mixing and code-switching), the use of a common language of communication (e.g. pidgins and creoles), and the transfer of linguistic elements from one language into another (e.g. linguistic borrowing) (Zenner & Kristiansen, 2014).

3.2.1 *A definition of linguistic borrowing*

Linguistic borrowing is the contact-induced importation of linguistic features from one language into another (Haspelmath, 2009; Hoffer, 2002; Thomason & Kaufman, 1988). In an archetypal borrowing situation, a recipient language (RL; also called receiving, receptor, replica, or borrowing language), receives linguistic features from a source language (SL; also called donor or model language), and these features contribute to expanding the range of expression of the recipient language. Stated more simply, "items/structures are copied from language X to language Y, but without speakers of Y shifting to X" (Hickey, 2010, p. 13). Whilst the use of the metaphor 'borrowing' is somewhat inaccurate in that nothing is actually borrowed (i.e. removed) from the source language, it is now a well-established term in the field of language contact studies (Matras, 2009; Winford, 2010).

Borrowing is often contrasted with interference, where borrowing is the introduction of foreign features into a language by speakers of that language (Thomason & Kaufman, 1988, p. 37), and interference is the influence of features of a first language on the usage of a subsequent language (Winford, 2010, p. 170). An example of borrowing is if a Japanese speaker, speaking in Japanese, says *kinoo supootsu sentaa ni ikimashita* ('yesterday I went to a sports centre'). This utterance uses the compound noun *supootsu sentaa*, which has been derived from English 'sports centre' and incorporated into Japanese. In interference, a Japanese speaker speaking in English and having trouble with the phonological distinction between /r /and /l/, which is not explicitly made in Japanese, is showing interference in their production of English due to the influence of (absent) features in Japanese.

Borrowing can also refer to different types of transfer between languages. Broadly defined, it can be used as a hypernym referring collectively to specific types of transfer of linguistic material from one language into another, including sounds (phonological/phonetic borrowing), sentence patterns (syntactic/grammatical borrowing), meanings (semantic borrowing) and words (lexical borrowing). Viewed more narrowly, it can be used as a hyponym referring to just one of these specific types of transfer. It can also refer to both a process and a product of transfer. As a process, borrowing can describe the activity of transferring linguistic material from one language into another, as with the borrowing of linguistic material from English into Japanese. As a product, borrowing can describe the outcome of transferring linguistic material from

one language into another, as with English borrowings (i.e. loanwords) which are used in the Japanese language.

In light of the ambiguity of the term borrowing, attempts have been made to offer a more precise definition. Van Coetsem (1988) defines borrowing as "the transfer of linguistic materials from a SL into an RL via the agency of speakers for whom the latter is the linguistically dominant language, in other words, via RL agentivity" (p. 172). This definition utilises two important distinctions discussed by Van Coetsem: (1) a distinction between recipient-language and source-language agentivity, and (2) between a linguistically dominant and linguistically non-dominant language (Van Coetsem, 1988). Explaining the first distinction, Van Coetsem proposes that there is always a transfer of linguistic material in the direction from a source language (SL, i.e. the language from which the linguistic material departs) into a recipient language (RL, i.e. the language into which the linguistic material arrives). Within this framework, borrowing is when the agent of transfer is an RL speaker (Van Coetsem, 1988). An example situation is when a Japanese speaker uses English loanwords when speaking Japanese, such as with the phrase given earlier where the English loanword *supootsu sentaa* is used in the Japanese utterance of *kinoo supootsu sentaa ni ikimashita* ('yesterday I went to a sports centre'). When the agent is a SL speaker, the process is called imposition (Van Coetsem, 1988). An example is the likely substitution by a Japanese speaker of the unstressed short-vowel schwa pronunciation in English words with the equivalent stressed-vowel sound in Japanese. This would change the /ət/ pronunciation in the second syllable of the English word 'pilot' (/ˈpīlət/) into /ɒt/, making a distinctive Japanese-English pronunciation of /ˈpīlɒt/.

In the second distinction, the issue of which language is linguistically-dominant for the speaker is of major importance. A linguistically dominant language is the one in which the speaker is most proficient, which is not always the native language of the speaker. This issue of language dominance is often referred to by the use of the terms L1 (Language 1) to refer to a linguistically dominant language, and L2 (Language 2 or 3 etc.) to refer to a linguistically non-dominant or foreign language. Table 3.1 gives an overview of Van Coetsem's framework.

Table 3.1 The distinction between contact-induced borrowing and imposition (adapted from Van Coetsem, 1988).

	Contact-Induced Language Change	
	Borrowing	**Imposition**
Who transfers the material?	A speaker whose dominant language is the Recipient Language	A speaker whose dominant language is the Source Language
Technical name	Recipient Language Agentivity	Source Language Agentivity
Example Situation	A speaker of Japanese using English loanwords when speaking Japanese	A speaker of Japanese using Japanese pronunciation when speaking English

Linguistic borrowing can therefore be understood as referring to situations where the agent of transfer is an RL speaker and the direction of transfer is from a non-dominant SL (L2) into a dominant RL (L1). In imposition the transfer is still from SL to RL, but the agent is now a SL speaker and so features of the dominant SL (L1) are transferred into the non-dominant RL (L2). Adopting this framework for the present study, which examines English vocabulary used in the Japanese language, borrowing refers to Japanese agents (RL-agentivity) using English words (from an SL/L2) within their dominant language of Japanese (an RL/L1).

3.2.2 *A definition of lexical borrowing*

Within linguistic borrowing, theoretically any linguistic feature can be transferred from an SL into an RL, including the borrowing of phonology, morphology, orthography, syntax, and meaning. Such borrowed features have been broadly categorised into two types: material borrowing and structural borrowing (Haspelmath, 2009, p. 38). Material borrowing involves the transfer of pairs of sound and meaning, most typically associated with the borrowing of lexemes (Hoffer, 2002; Matras, 2009; Winford, 2010). An example is the English word ‘salad’, borrowed into Japanese and rendered as *sarada*, showing a similar (although not identical) phonological structure in its RL loanword form as in its original SL form. Structural borrowing involves the transfer of

grammatical patterns, such as word order and case-marking patterns, and/or semantic patterns such as kinship term systems (Haspelmath, 2009, p. 39).

The vehicle of transfer of this linguistic material between languages, both material and structural, is primarily lexis. Indeed, it is said that of all the influences on languages arising from language contact, the borrowing of words is most common (Thomason, 2001, p. 10). This is because the lexicon is less stable and more open to change than other aspects of a language such as phonology and semantics (Winford, 2010, p. 172). The borrowing of words, or more technically lexical borrowing, has been defined as "the transfer of lexical material from one language (the donor, source or model language) to another language (the receptor or replica language)" (Zenner & Kristiansen, 2014, p. 1). In this way, lexical borrowing can be seen as both a process and a product, that is, either the act of transfer of words and phrases, or the actual transferred words and phrases themselves. When referring to the product of lexical transfer, the term loanwords is typically preferred in the literature, although the term borrowings is also widely used.

3.2.3 *A definition of loanword*

Haugen (1950) put forward a taxonomy of three main categories of lexical borrowings: loanwords, loanblends and loanshifts (1950, pp. 213-215). This categorisation remains popular in present-day lexical borrowing research (Durkin, 2014; Greavu, 2014; Winford, 2010). Haugen's taxonomy centres on the use of the terms importation, where a foreign language pattern is imported into the language, and substitution, where a foreign language element is substituted with something already in the native language (1950, p. 213). These two processes are then further analysed into morphemic importation/substitution and phonemic importation/substitution, the various combinations of which are used to define the concepts of loanwords, loanblends and loanshifts.

Haugen explains that the term loanword refers to words which have had their meaning as well as phonemic shape imported and gives the example of American English 'shivaree' borrowed from [French] *charivari,* meaning an uninvited serenade of newlyweds (1950, p. 13). Loanwords do not involve the substitution of SL morphemes with RL material and this means that the shape of the word in the recipient language is very similar, but not always identical to, its shape in the donor language. There is,

however, often a certain degree of phonemic substitution to allow the words to fit into the RL phonology, which Haugen classifies into none, partial, or complete (1950, p. 214). If there is no phonemic substitution, then this is a loanword in its purest form, where it is used phonologically in the recipient language in virtually the same way as in the donor language, as with the use of the Spanish loanword 'burrito' in English which very closely mirrors the SL phonology (Winford, 2010, p. 172).

Many loanwords, however, undergo at least some extent of phonemic substitution to help them integrate into the phonological system of the recipient language. With English loanwords in Japanese, the vast majority involve at least partial phonemic substitution to allow them to fit into Japanese phonology (Irwin, 2011). The example discussed earlier of the English loanword *sarada* is a morphemic importation of the English word 'salad', with partial phonemic substitution. The word-ending consonant /d/ in its English form is substituted with the consonant-vowel combination of /da/ in its loanword form, and the English phoneme /l/ is replaced with the Japanese liquid phoneme /r/. In many cases English loanwords in Japanese undergo extensive phonemic substitution. This happens in large part because of the need for multiple epenthetic vowels to break up consonant clusters to make them fit into the consonant-vowel pattern of Japanese phonology. A well-known example of an English word which has undergone extensive vowel epenthesis to be integrated into Japanese is マクドナルド *makudonarudo* which is the rendering of the American burger chain McDonald's. Extra vowels have been inserted to break up consonant clusters and conform the word to regular Japanese consonant-vowel (CV) word-endings. This changes its syllabic structure from three syllables in English to six in its Japanese form (more technically, six morae rather than syllables, as Japanese phonology typically uses the phonological unit of mora rather than syllable).

Loanblends involve the morphemic importation of SL material as well as some morphemic substitution with RL material, as in Haugen's (1950) example from Pennsylvania German of 'bassig' meaning 'bossy' which combines a derived form of English 'boss' with German '-ig' (Haugen, 1950). Greavu writes about how this category can be sub-divided into blended derivatives and blended compounds (2014, p. 99). The former type is when native-language derivational suffixes are substituted for ones in the donor language, like in the example of 'bassig' just given. In the lexical borrowing from English into Japanese such blended derivatives are rare but do exist. An example which shows both the creativeness and complexity of English borrowings in Japanese is the loanblend *daburu* 'to be doubled'. This English borrowing has two

forms in the Japanese language: a loanword form and a loanblend form. As a loanword, *daburu* is typically written all in *katakana* and shows morphemic importation without morphemic substitution, as well as a certain degree of phonemic substitution to assimilate it into Japanese. It is used like the English adjective 'double', especially in compound nouns, like *daburu kurikku* 'double click' and *daburu sutandaado* 'double standard'. In its loanblend form, daburu is written in a combination of *katakana* script for the imported morpheme of ダブ *dabu* and *hiragana* script for the native verbalising morpheme of る *ru*, making ダブる *daburu* 'to double up/be doubled'. Greavu's second category of blended compound loanblends, which are words containing SL and RL language stems (2014, p. 101) are represented by the Japanese-English compound of *meiwakumeeru* 'junk mail' discussed previously in Section 2.2.4.

Loanshifts involve morphemic substitution but no importation. These are also called loan translations, calques, or loan meanings, and involve the use of RL material to mirror a construction in the SL, where existing words in the RL either have their meaning extended to include the meaning of a word in the SL (i.e. loan meaning) or are used to replicate formation models in the SL (i.e. loan translation). Winford gives examples of the way the English language has influenced American Portuguese, to extend the meaning of the word 'frio' from 'cold temperature' to also include 'cold infection'; and has influenced German to create the loan translation of 'Wolkenkratzer' which is modelled on the English form 'skyscraper' (2010, pp. 172-173).

3.2.4 *A definition of English loanwords in Japanese*

A variety of Japanese terms has been used to refer to loanwords in the language, such as *gairaigo* 'words coming in from outside', *gaikokugo* 'foreign words', *katakanago* 'words written in *katakana*', *yoogo* 'Western words', *shakuyoogo* 'borrowed Western words' and *hakuraigo* 'words coming in from the West' (Tomoda, 2002). None of these are particularly precise, with the one most favoured being *gairaigo* 'words coming in from outside' (Irwin, 2011). Viewed broadly, *gairaigo* can refer to a number of different types of loanwords including ones directly borrowed, in that their RL and SL structural form and semantic meaning are similar, such as Japanese *bukku* from English 'book', and ones which are indirectly borrowed, in that they are domestically produced from borrowed elements, such as *gattsupoozu* (guts pose) borrowed from English 'guts' and 'pose' and meaning a 'celebration pose' after a win. When indirectly borrowed

from English vocabulary, these domestically-produced lexical items are known as the sub-category of *wasei eigo* 'made-in Japan English'. The same made-in-Japan type of word is found in the Sino-Japanese lexical stratum, where they are called *wasei kango*, such as with 電話 *denwa* 'electric talk' which is a modern creation for a modern object using historically borrowed Chinese *kanji*.

A narrower perspective typically excludes these made-in-Japan borrowings and limits the definition of loanword to the words which have had their sound and meaning borrowed from a word in another language (see Section 3.2.3 above). The current study is focused specifically on English loanwords in Japanese and uses Haugen's definition (1950) to include only those words which have had at least part of their sound and meaning borrowed from English. This includes words such as *hoteru* 'hotel', *dezain* 'design', *peeji* 'page', *akushon* 'action', and *furasshu* 'flash'. The small number of loanshifts and loanblends in Japanese derived from English language words are outside the scope of the current research.

3.3 General Areas of Loanword Research

Lexical borrowing as a field of research has received significant attention since the late 19th century, and particularly from the mid-20th century with the works of two authors in particular, Haugen (1950, 1953), and Weinrich (1953). Their detailed analyses of the processes of borrowing elements of one language into another laid the foundation for the subsequent development of lexical borrowing research during the latter half of the 20th century into what has been called "a rich and longstanding tradition in linguistics" (Zenner & Kristiansen, 2014, p. 2). Kang (2013) gives a bibliographic overview of over eighty previous studies on loanwords, collected from book chapters, journal articles, and PhD monographs. She identifies six principal areas into which the studies can be categorised and represents each area with a research question that guides the overall direction of the studies: (1) definitions (what are loanwords?), (2) borrowability (why are words borrowed?), (3) emergence and diachronic evolution (how are loanwords introduced and spread?), (4) integration and adaptation (in what ways are loanwords adapted to the receiving language?), (5) lexical stratification (to what extent do loanwords follow the same restrictions as native words?), and (6) extralinguistic factors (how are loanwords accepted and used in society?) (Kang, 2013).

Kang's overview is neutral in approach and is not intended to offer comment on the nature of the research conducted in the field. She instead simply recommends to the interested reader some of the most important studies from amongst what is available. A more critical approach to giving an overview of the previous literature on lexical borrowing is taken by Zenner and Kristiansen (2014). They have the aim of showing that the majority of research has so far been conducted within a structuralist paradigm, with analyses largely limited to inventories of loanwords in different languages and descriptions of the formal aspects of the loanwords such as their phonological and morphological features (2014, p. 1). They argue that this is because of a predominant focus on examining loanwords in their position within the RL, at the expense of a comparative analysis between their position in both the RL and SL. Giving examples of the 'new perspectives' which they say need to be taken on lexical borrowing research, they call for analyses which look beyond single borrowed words and consider multi-word borrowed phrases, such as idioms and metaphorical phrases, and research which examines loanwords in combination with near-equivalent expressions which may already be present in the recipient language (2014, pp. 1-5).

In their overview, Zenner and Kristiansen (2014) categorise the current literature on lexical borrowing into four categories, which they call prime topics in current lexical borrowing research, and which they argue have a strong structuralist perspective: (1) the construction of borrowing taxonomies and the definition of transfer types, (2) the analysis of morphophonological and orthographic rules which affect the integration of loanwords into the RL, (3) the demarcation of borrowing from the related field of code-switching, and (4) the analysis of the borrowability of linguistic items in terms of identifying universal constraints on which items are more likely than others to be borrowed.

From a comparison of the summaries of lexical borrowing research given by Kang (2013) and Zenner and Kristiansen (2014), presented in Table 3.2, it can be seen that whilst they do not reference each other, they almost completely overlap in the areas into which they categorise previous studies. Five out of six of Kang's major questions can be matched directly with Zenner and Kristiansen's prime topics (with Zenner & Kristiansen's single topic of adaptation rulesmatching with Kang's two questions of how the loanwords are adapted and to what extent loanwords have the same restrictions as native words). Further, the third column in the table highlights that it is only the last major area of study put forward by Kang (2013), concerning extralinguistic factors related to the loanwords (including social issues such as the age, gender, and class of

bilingual speakers), which is not overtly structuralist in nature. Taken together, these two overviews offer a detailed and critical summary of the main areas of research where attention has been focused in the study of lexical borrowing and loanwords.

Table 3.2 A comparison of the summaries of lexical borrowing and loanword research given by Kang (2013) and Zenner and Kristiansen (2014).

	Major Questions in Previous Loanword Research (Kang, 2013)	**Prime Topics in Current Lexical Borrowing Research (Zenner and Kristiansen, 2014)**	**Structuralist Perspective?**
1	What are loanwords? *(Definitions)*	Defining transfer types. Demarcating borrowing from codeswitching.	✓
2	Why are words borrowed? *(Borrowability)*	Universal scale of receptivity to foreign material.	✓
3	How are loanwords introduced? (Emergence & Diachronic Evolution)	Universal scale of receptivity to foreign material.	✓
4	Why and how are loanwords adapted to the receiving language? *(Adaptation)*	Adaptation rules.	✓
5	To what extent do loanwords follow the same restrictions as native words? *(Lexical Stratification)*	Adaptation rules.	✓
6	How do extralinguistic factors affect loanwords? *(Extralinguistic Factors)*	-	X

3.4 Traditional Practices in Loanword Research

This section discusses how at the same time that many of the previous studies carried out in the areas in Table 3.2 above have contributed invaluable data to various areas of linguistic research, they have also led to the entrenchment of several traditional research practices which can be seen to be unconducive to the analysis of the loanwords' grammatical behaviour. Each practice will be discussed from three viewpoints: How has the practice arisen in loanword research? Why is the practice particularly unconducive for the analysis of the grammatical behaviour of loanwords? and How is the practice specifically seen in the context of research on English loanwords in Japanese?

3.4.1 *Research on loanword phonology*

The most intensively studied area of lexical borrowing research is the examination of loanword phonology. Issues in loanword phonology are involved in several of the areas of loanword research in Table 3.2 above, particularly loanword adaptations and lexical stratification, meaning that attention to phonological features of loanwords has come from numerous different directions. Much of the attention on the phonological features of loanwords centres on the phonological adaptations made to words borrowed out of one language for their integration into a recipient language. This research has been influential not only in understanding the process of integration itself, but also in better understanding the overall phonological structure of the recipient language. The investigation of loanword phonology, however, involves perspectives of analysis and methodological practices which are markedly different from those necessary for an analysis of loanword grammar.

As words are borrowed out of a source language and integrated as loanwords into a recipient language, they commonly undergo various adaptations to make them conform to the recipient language's linguistic system. These adaptations are studied within the question of why and how loanwords are adapted, the fourth major area of research within Kang's (2013) summary shown in Table 3.2 above. The various adaptations made to loanwords can be broadly categorised into structural adaptations and meaning adaptations (Winford, 2010, p. 175). Structural adaptations concern the modifications which allow the loanwords to fit into the structure of the recipient language. These have been described as dimensions according to which loanwords can be grouped, or more precisely "the degree of morphophonological, orthographic or syntactic integration to the receptor language" (Zenner et al., 2014, p. 43). Meaning adaptations concern both semantic modifications, such as semantic narrowing and broadening to restrict or widen their denotations (see Section 2.4.1), and pragmatic modifications, such as their adoption as euphemistic terms for taboo language.

The most frequent type of adaptation made to loanwords, across virtually every language contact situation, is phonological adaptations. This is because they are the most fundamental type of adaptation needed for the usage of source language vocabulary in a recipient language. Breiter (1997) describes a framework of loanword adaptations, shown in Table 3.3, which describes phonological adaptations to loanwords being necessary for their initial penetration into the recipient language, along with graphemic adaptations that are dependent on the orthographies of the languages in

contact. Phonemic and graphemic transformation allows a loanword's integration into the recipient language's grammatical system, which can include morphological adaptations such as the assignment of gender, case, and number. The next stage of derivational integration involves the formation of derivational nests, whereby loanwords can be adapted to follow the derivational rules of the RL grammatical system. The final stage brings a possible extension or restriction of meaning in the SL word to fit its function in the recipient language and can also involve the loanword replacing equivalent expressions already available in the language.

Table 3.3 Stages of a borrowed word's integration into a recipient language (adapted from Breiter, 1997).

Stage	Integration Type	Process	Explanation
Stage 1	Phonological and Graphemic	Penetration	The SL word undergoes a formal transformation
Stage 2	Morphological	Grammaticalisation	The word is then assimilated into the RL grammar system
Stage 3	Derivational		
Stage 4	Semantic	Lexico-semantic assimilation	Integration is completed when the meaning of the SL word is adapted for RL purposes

As a loanword goes through each of the integration stages, the adaptations become more idiosyncratic, meaning it is easier to generalise the adaptations that a loanword will undergo at stage one of their integration than it is at stage four. Stage one adaptations are for the most part dependent on a contrastive analysis of the phonological and orthographical systems of the languages in contact. Where differences are found, adaptations will be needed to conform the source language word to recipient language norms. In the Japanese language, for example, the absence of the schwa sound means that English words containing schwa are typically adapted with the substitution of the closest available sound in Japanese phonology (Irwin, 2011). Furthermore, Latin alphabet characters are commonly substituted for *katakana* syllabary characters for the full integration of an English word into Japanese.

At this level of integration, adaptations can normally be predicted by analysing the word in isolation and identifying which phonemes and graphemes will need to be modified. At stage four of integration, however, the linguistic context of the loanword needs to be considered for an analysis of its meaning. Such contextual data has long been considered essential in properly understanding how words and phrases function in

a language, summarised in the often-quoted observation that "you shall know a word by the company it keeps" (Firth, 1957, p. 11). In other words, when the context of a word changes, it is possible that its meaning also changes. Because of this, establishing a set of rules by which a loanword will undergo certain modifications is manageable at stages 1 and 2, possible but more complicated at stage 3, and highly complex at stage 4. This can be seen as one of the main reasons why lexical borrowing research on loanword adaptations is more strongly represented with studies of structural adaptations than semantic adaptations, and that phonological adaptations are by far the most common type of structural adaptation studied.

Much of the research on loanword phonology attempts to identify patterns of phonological adaptations which can inform the description of a set of rules in the recipient language's phonological system applied to borrowed words. The description of such rules has been of major influence in two specific areas of loanword research. One of these is the ongoing debate as to what underlies the adaptations: abstract phonological representation, detailed phonetic representation, or a combination of both. Uffmann (2015) for example, discusses the fact that segmental adaptations are not only made when a phoneme in the source language word is absent in the recipient language, but at any time when a speaker makes a decision as to which recipient language phonemes are the best-equivalent of ones in the source language. This leads him to conclude that it is not a single strategy which determines the adaptation process, but rather a combination of phonetic similarity and phonological equivalence and is particularly dependent on the level of bilingualism of the speaker (p. 23).

The second area where the description of rules of phonological adaptation have had a major influence is the issue of a language's lexical stratification. In this research, constraints of the native phonology which do and do not apply to foreign vocabulary are used to try to identify different lexical strata in the language's lexicon. More specifically, it is investigated what causes some loanwords to be fully integrated into the lexicon so that they conform to norms of the recipient language's phonology, and others to retain features of their source language form which do not conform to patterns displayed in other lexical strata in the lexicon. Holden (1976), for example, found that, different loanwords assimilated at different rates depending on the type of constraint involved in the adaptation of different phonological features. Some of the most influential work in this area has been that of Ito and Mester (2008) and their proposal of a core-periphery structural model of a language's lexicon, where different phonological constraints on different words are used to model various lexical strata that are layered

outwards from a central core of native words. Not only has the issue of lexical stratification been important for studies within the area of historical contact linguistics, but also, when viewed synchronically, varying constraints on words within a language's different lexical strata have implications particularly in the learning of the language (Uffmann, 2015).

The significant contribution to linguistic research of studies on loanword phonology cannot be understated, but not only has the wealth of attention on this specific feature of loanwords caused other areas of investigation to be overshadowed, the methodological approach to researching loanword phonology is largely unconducive to researching their grammatical behaviour. As was discussed in the previous section, a major goal of loanword phonology research is to posit a set of rules which can account for what is possible regarding phonological adaptations and constraints, and these rules are generally constructed from a contrastive analysis of the phonologies of the languages in contact. In this way, the rules are theoretical, stating that if language X has phonological feature A, but language Z does not have this feature, then a word borrowed from language X will need to be adapted with phonological feature B for its integration into language Z (see below for an example). This approach, however, is not conducive to the analysis of the loanwords' grammatical behaviour because the grammatical relationships in which they participate cannot be accounted for by a set of rules. Just as with all words in a language, the grammatical behaviour of loanwords is context-dependent, with a single loanword participating in a large number of different grammatical relationships with other words in its local context. As such, the examination of a loanword's grammatical behaviour needs a methodology which accounts for what is natural in its usage rather than just what is theoretically possible.[6]

Taking an example from the analysis of phonological adaptations made to English loanwords in Japanese, a contrastive analysis of the phonemic inventories of English and Japanese would show that words in Japanese have to end in a vowel sound, apart from the nasal consonant *n* used in words such as *gakumon* 'school gate' and *shinbun* 'newspaper'. This constraint means that English words ending with a vowel sound, such as 'ticket', 'kick', 'pump', and 'bat', are appended with an extra vowel when borrowed into Japanese in order to conform to Japanese phonological norms, thereby making *chiketto*, *kikku*, *panpu*, and *batto*. The vowel which is appended is dictated by a further

[6] The distinction made here between 'possible' and 'natural' is borrowed from Hoey (2005) who uses these terms in his theory of lexical priming. He argues that the majority of both traditional and modern accounts of the relationship between lexis and grammar are just theories of what is possible because they make little or no examination of what is natural in attested language.

rule based on the preceding vowel (Irwin, 2011). To explain this rule, evidence is only needed of several examples of loanwords where this adaptation is made, and this evidence can be gained from dictionaries or from intuition. Reference to the natural behaviour of loanwords, such as in a corpus of the Japanese language, is mostly superfluous. However, if the grammatical behaviour of the two loanwords was to be investigated, rules would not be able to account for all the ways in which the loanwords are structured with other words in the language. For this, examination of the loanwords in naturally-occurring language would be needed.

In Zenner and Kristiansen's (2014) review of the field of lexical borrowing research discussed in Section 3.3 above, their overall main criticism of the previous research is centred on this issue of loanwords being removed from their naturally-occurring contexts, or what they term the handling of loanwords as single-word units. They argue that whilst the practice is advantageous for certain areas of loanword analysis, it has led to other areas being neglected, such as the analysis of loanwords in fixed phrases like 'kick the bucket' and the understanding of relationships between the loanwords and existing alternative expressions in the borrowing language. For the analysis of phonological adaptations made to loanwords, handling the loanwords as single-word units facilitates their systematic deconstruction into individual phonemes and the analysis of where modifications have and have not been necessary. For the analysis of the grammatical behaviour of the loanwords, however, an examination is needed of the relationships which connect a loanword to other words around it, such as if and when it acts as the object of a verb, how often it is the modifying part of a compound noun phrase, and whether or not it can perform the role of multiple parts of speech. This requires a large number of naturally-occurring instances of these contextual units, in a sample of the language in which the loanwords have been integrated, something which is not possible if the loanwords are examined as free-standing individual words. The traditional practice of treating loanwords as single-word units is therefore unconducive to the analysis of their context-dependent grammatical behaviour.

The type and degree of adaptations which a borrowed word undergoes are for the most part dependent on the linguistic differences between the languages in contact (Poplack, Sankoff, & Miller, 1988; Winford, 2003). The considerable phonological differences between English and Japanese would therefore predict extensive adaptations to English words for their integration into Japanese, and this is indeed the case in the spoken representation of English loanwords in Japanese. With only five regular vowel

phonemes in the contemporary Japanese phonemic system, numerous English vowel sounds such as the unstressed schwa therefore need to be substituted for the closest available Japanese equivalent. Moreover, the insertion of epenthetic (additional) vowels to break up consonant clusters in English words disallowed in the Japanese consonant-vowel (CV) open-syllable structure often entirely alter the phonological structure of their loanword form. The mono-syllabic English word 'text', for example, gets modified into a four-syllable structure in its loanword form テキスト *tekisuto*. Significant differences in orthography as well, with loanwords in Japanese most typically being written in the *katakana* syllabary, mean that English loanwords need to undergo complete graphemic adaptation for their orthographic integration.

Morphological and syntactic adaptations to English words for their integration into Japanese, however, are much less common; in spite of the fact that significant differences between the morphology and syntax of the two languages would predict extensive adaptations being necessary. They differ, for example, in their fundamental composition with Japanese being an agglutinative language that takes multiple morphemic conjugations appended to verbs (Kageyama & Saito, 2016), and modern English largely a fusional language with morphemes merged into the root of the word (Meyer, 2009). In their syntactic structure also, the two linguistic systems vary considerably in that Japanese displays a subject-object-verb word order, in contrast with a subject-verb-object order in English. Even with these differences, it is because the vast majority of English loanwords in Japanese are nouns that morphological and syntactic adaptations of English words are for the most part unnecessary.

The nouns of a language typically have the richest lexical content of all parts-of-speech and are much less structurally tied to other words around them compared with other word-classes such as prepositions and adverbs. This makes it relatively easy to borrow them out of a syntactic noun-slot in one language and integrate them into a noun-slot in another language, so much so that the assignment of loanwords to syntactic slots in the borrowing language has been described as "virtually categorical" (Poplack et al., 1988, p. 52). Indeed, it has been stated that of the four main ways in which a word is structurally integrated into another language, "loanwords pose little problem for syntactic adaptation, simply behaving like their counterparts of different syntactic categories in the recipient language" (Winford, 2003, p. 48).

With the majority of words borrowed from English into Japanese being nouns and also functioning as nouns in Japanese, their syntactic integration can therefore be considered for the most part unproblematic. Moreover, nouns in Japanese are non-

inflecting, showing no marking of gender, number, case, or article. This means that morphological adaptations are also unnecessary for the majority of English words borrowed into Japanese (Irwin, 2011, p. 137). Using the loanword *tekisuto* ('text') again as an example, this is a single, root-morpheme noun in both languages.[7] This is in contrast to other language contact situations, such as lexical borrowing from English into Dutch, French, and German, where complex morphological adaptations are often needed due to the languages' assignment of grammatical gender to borrowed words (Winford, 2003, p. 49).

Syntactic and morphological adaptations to English words in Japanese are not completely absent, however. Some English loanwords function as verbs and adjectives in Japanese and do so with the addition of grammatical markers such as the verbaliser *suru* 'to do' for verbs and *na* (a form of the copula) for adjectival nouns. In such cases, the syntactic role of the words in English is often ignored and they are instead borrowed as a noun and then appended with grammatical markers for the desired syntactic role in Japanese. The loanword *romanchikku*, for example, borrowed from the English adjective 'romantic' is borrowed initially as a noun and then appended with *-na* to turn it into an adjectival noun in Japanese. Overall, however, with the vast majority of English loanwords being nouns, morphological and syntactic adaptations are minimal.

For the reasons above, it is unsurprising that the vast majority of research on English loanwords in Japanese has been conducted on their phonological features. Indicative of this is Irwin (2011), which stands as the most comprehensive English-language general treatment available on English loanwords in Japanese. In his examination of lexical borrowing into Japanese over the last 450 years, English loanwords make up the majority of the examples in the book. His main focus is on the phonological behaviour of loanwords, shown by the extended chapters on phonology, morpho-phonology, and orthography of loanwords in Japanese, and the much briefer treatments of loanword semantics and societal attitudes towards the loanwords. Indeed, concerning any treatment of the grammatical behaviour of loanwords, it is interesting to observe that within this comprehensive work the words 'syntax' and 'syntactical' do not appear at all and the word 'syntactic' is used only once in a passing reference to phonological and syntactic constraints. Moreover, the word 'grammar' appears only 5 times and with most of these referring to the grammars of other languages.

[7] The verb form 'text' (i.e. 'to text') is not used in Japanese (the verbalized form of *tekisuto suru* only appears 13 times in the 8-billion-word jpTenTen11 corpus).

Whilst Irwin's (2011) approach is mainly to document and describe the extensive phonological adaptations, other studies have taken the approach of using the adaptations as evidence in discussions of specific aspects of the Japanese language. Two aspects in particular have attracted a large amount of discussion: the structure of the phonological lexicon, and the influence of loanwords in the teaching of English in Japan. The first of these, which is primarily a linguistic topic, is largely attributed to the work of Ito and Mester (1999) who examined the constraints which apply to the phonological behaviour of loanwords in order to reveal stratification patterns in the Japanese lexicon. They then used these patterns to theorise on the overall structure of the Japanese phonological system. Their work has been highly influential in the field of phonological research, mostly because of the debate which has arisen, discussed briefly in Section 3.4.2 above, as to whether adaptations are based on abstract phonological representations, detailed phonetic representations, or a combination of both (Uffmann, 2015).

In regard to the importance of the phonological adaptations made to loanwords in the area of English language education in Japan, a primarily applied linguistic topic, much attention has been given to the *katakana* English pronunciation of Japanese speakers of English. It is so-called because of how English words are often transliterated into *katakana* script, especially in textbooks for English language learning. Writing English in the *katakana* script causes considerable changes to the pronunciation and prosodic features of English, such as the merging of certain distinct sounds in English into a single sound in Japanese. Examples include English /b/ and /v/ being pronounced as Japanese-English /b/, and /l/ and /r/ being pronounced as /r/, and the open ending of syllables (i.e. a CV syllable structure) caused by epenthetic vowels attached to consonant-ending syllables in English. Changes such as these have been called a *katakana* effect, whereby a supposed block happens in communication due to the limited Japanese phonemic inventory (Martin, 2004). The issue of *katakana* English has been taken up in various areas of pedagogical research, such as whether or not native Japanese speakers activate their lexical knowledge of English during the cognitive processing of English words written in *katakana* (Tamaoka & Miyaoka, 2003), and the strategies which learners use in order to employ *katakana* for the transcription of newly encountered loanwords (Preston & Yamagata, 2004).

Overall, the considerable number and highly predictable nature of the phonological adaptations made to English words in Japanese has focused a great deal of academic attention on this area of loanword research. In contrast, because morphological and syntactic adaptations are much less common, these can be seen as one of the major

causes of only limited research having been carried out on their grammatical features. The next section explores one more of these causes: that traditionally the grammar of loanwords has been conceptually limited to categorising their parts-of-speech.

3.4.2 *Research on the lexical categories of loanwords*

When analyses have been conducted on the grammar of loanwords, they have invariably been limited to determining their part(s)-of-speech. This is due to the long history of studies on the borrowability of linguistic elements across languages, investigating the extent to which a linguistic item or group can be borrowed from one language into another (Van Hout & Muysken, 1994). For over a century, the analysis of loanwords has been integral in the construction of hierarchies of borrowability of different parts of the recipient language's lexicon. These hierarchies are used in the fields of contact linguistics and historical linguistics for the evidence they can provide in helping to establish genetic relationships between languages. They are also important for helping to understand why words are borrowed in the first place, because all languages are able to create new words from their own linguistic resources (Haspelmath, 2009, p. 35).

In such research, analysis is made of which parts of a language's lexicon are most receptive to borrowing, in order to try to isolate a core non-borrowed part. This non-borrowed part of the lexicon can then be compared with similar non-borrowed parts of other languages to try to establish genetically-shared features across languages, rather than ones which have simply been borrowed. Borrowability studies, however, are only concerned with categorising loanwords into parts-of-speech, such as nouns, verbs, and adjectives, and this has been a major factor in why the investigation of the grammar of loanwords has rarely moved beyond this basic and somewhat arbitrary categorisation.

Research on borrowability is the second major question in Kang's (2013) overview of previous research on loanwords in Table 3.2 above, and concerns what Zenner and Kristiansen (2014) term the universal scale of receptivity to foreign material. The contribution of this area of research has been invaluable in the field of historical linguistics, where it has helped in understanding the genealogical relatedness between languages (by separating out inherited and borrowed linguistic material), and the types of situation whereby languages come into contact (by analysing the type of words borrowed, e.g. marine or political terms) (Haspelmath & Tadmor, 2009, pp. 1–2).

Historically, the issue of borrowability was most concerned with whether it is possible to have truly mixed languages (Winford, 2003). In seeking an answer, the presence of loanwords across languages was considered problematic because they clouded the issue of whether languages could be seen to have genetically-shared vocabularies (Kang, 2013). As such, much of the early research was focused on producing lists of loanwords, and then excluding them from further analysis to uncover the non-borrowed vocabulary which, if shared, could be seen as evidence of genetic relationships between the languages. Whitney (1881), in one of the earliest major treatments of the issue, calls for a comprehensive scale of borrowability, and this was taken up by Haugen (1950) whose work on the issue introduced some of the important terms discussed in Section 3.2. Later, Muysken (1981) developed a more sophisticated hierarchy of nouns as the most borrowable lexical category, then adjectives, verbs, prepositions, down to subordinating conjunctions as the least borrowable.

Interest in the topic of borrowability has continued into the present day, with the methods employed in the analyses developing significantly, but the central aspect of the analyses remaining the establishment of the loanwords' part(s)-of-speech. Van Hout & Muysken (1994), for example, in their investigation of Spanish borrowings in Bolivian Quechua, developed analytical techniques which could be used to determine a new hierarchy of how easily lexical items could be borrowed. They had identified that there was a general lack of empirical evidence on the subject, especially in relation to comparing the distribution of the words in the recipient language with that in the source language, arguing that this was most likely caused by an insufficiency in the tools typically used in structural linguistic analyses. In light of this, their study adopted a new approach of using a bilingual corpus to investigate general constraints on the borrowing of loanwords. The focus of the study, however, remained centred on categorising parts-of-speech and explored the issue of whether content words (such as nouns, verbs, and adverbs) were more or less likely to be borrowed than function words (such as pronouns and quantifiers), eventually ascertaining a strong effect of content words being more borrowable than function words (Van Hout & Muysken, 1994, p. 51).

One of the most recent studies on borrowability, the large-scale cross-linguistic Loanword Typology Project (2009), has continued the practice of categorising loanwords by their part(s)-of-speech. The project aimed to widen the perspective on borrowability hierarchies from its narrow focus on individual language contact situations, such as in Haugen (1950), Muysken (1981), and Van Hout & Muysken (1994), to a cross-linguistic consideration of how borrowing in one language compares

to other languages. The project looked at 41 languages, aiming to identify a core part of each language's lexicon which was non-borrowed and produce systematic empirical evidence of the statistical likelihood that each of about 1000-2000 words in a language's core lexicon were borrowed. Their research produced a ranking of languages according to the overall percentage of the lexicon which could be seen to be constituted of loanwords and placed the languages along a continuum from highly receptive to borrowing to highly resistant to borrowing. For each language, the words which were analysed were also categorised into their major parts of speech, such as nouns, verbs, adjectives, and adverbs. The results were published online as the World Loanword Database (WOLD), and also in print (Haspelmath & Tadmor, 2009). Table 3.4 shows the top and bottom ten languages in the WOLD ranking.

Table 3.4 The top and bottom ten languages ranked by the percentage of loanwords which constitute their core lexicon.

Top/Bottom 10	Language Name	Percentage of loanwords
1	Selice Romani	63%
2	Tarifiyt Berber	53%
3	Gurindji	48%
4	Romanian	43%
5	Saramaccan	42%
6	English	42%
7	Ceq Wong	38%
8	Japanese	36% (6% English loanwords)
9	Indonesian	35%
10	Takia	32%
32	Q'eqchi'	15%
33	Iraqw	15%
34	Seychelles Creole	13%
35	Hup	12%
36	Oroqen	11%
37	Otomi	11%
38	Ket	10%
39	Manange	10%
40	Old High German	6%
41	Mandarin Chinese	2%

The methodologies of most of the studies on borrowability begin with the theoretical understanding that words in a language are parts of a structured system, and that this system imposes constraints on where the words can and cannot occur in the language (Haugen, 1950; Van Hout & Muysken, 1994). It is these constraints which are of interest in studies of borrowability because they dictate which categories of language can be borrowed more easily than others (Whitney, 1881). The analysis typically involves examining word-class distributions of the lexical items in a sample of language, with the outcome being lists of borrowed words categorised into the different word-classes. These lists are typically rankings of the frequencies of individual loanwords, and their respective word classes, which are removed from their original contexts. Illustrative of this is the database of loanwords in the Loanword Typology Project introduced above, where the pages for each of the 41 languages in the study list individual, context-free loanwords that were classified as borrowings from another language.

In many aspects of linguistic research, the lists of individual loanwords produced in such borrowability studies are of great value. The single most important finding from these studies has been that across all language contact situations, nouns are by far the most borrowable (Backus, 2014; Haspelmath, 2009; Hock & Joseph, 2009; Poplack et al., 1988; Thomason & Kaufman, 1988; Whitney, 1881; Winford, 2003). Besides this principal aim to establish the extent to which different word-classes can be borrowed, the lists can also be used as evidence of exactly what words have been borrowed between languages, how frequent the words are, and what kinds of words they are in terms of general, academic, and specialised vocabulary. Each word can also be deconstructed into its phonemes, graphemes, and morphemes and analysed for where adaptations have been made for the word to conform to the norms of the recipient language. The information that cannot be provided by these lists, however, is context-dependent features of loanwords, such as how they combine with other words in the formation of specific grammatical relationships. This is an example of what Zenner and Kristiansen (2014) mean in their argument that much of the previous research in the field of lexical borrowing has analysed loanwords narrowly as single-word units: that loanwords in borrowability studies have been typically studied in the form of single words removed from the linguistic contexts in which they naturally occur.

The most extensive study so far undertaken on the issue of borrowability in relation to the Japanese language was conducted as part of the Loanword Typology Project, described above, which culminated in a page of Japanese vocabulary on the online

World Loanword Database and a chapter in the print volume accompanying the database (Schmidt, 2009). For the study, Schmidt undertook an extensive analysis of loanwords in the Japanese lexicon from all sources, including European and Chinese loanwords, but excluding words created domestically from foreign material (i.e. *wasei eigo* 'Japan-made English', and *wasei kango* 'Japan-made Chinese-derived words')

In order to provide statistical data on the percentage of the core Japanese lexicon which could be established as having been borrowed, a 'meaning list' of 1,460 entries categorised into 24 semantic fields common to all studies in the project was used. As shown in Table 3.5, it was found that 28% of the words on this list are Chinese loanwords, a high percentage but one that is explained by Japan's long history of contact and borrowing from its Asian neighbour for over a thousand years (see Section 2.2.1 for a discussion of this history). The second largest group of borrowings are English loanwords, at 6% (Schmidt, 2009).

Table 3.5 The percentage of loanwords from different languages in the core of the Japanese lexicon (adapted from Schmidt, 2009, p. 562).

Source Language	Contribution of borrowed vocabulary (%)
Chinese	27.9
English	6.0
Dutch	0.3
Portuguese	0.2
French	0.2
Korean	0.1
Ainu	0.1
Spanish	0.1
Ryukyuan	0.1
German	0.1
All languages	34.9

In the part-of-speech analysis, it was found that of all the words on the list, 9.3% were English loanword nouns, 0.9% were verbs, 0.7% function words, and 0.6% adjectives. This shows an overwhelming tendency for English nouns to be borrowed into Japanese rather than other word classes such as verbs and adjectives, a finding which is consistent with other research studies (Irwin, 2011; Matras, 2009) The data for the English non-nouns contrasts sharply with the data of Chinese loanwords where the

overall percentage of verbs, function words, and adjectives was 67.9%. In terms of the semantic field of English loanwords, most were found to be words related to areas of material culture, such as clothing and grooming (26.1%), food and drink (19%), and the modern world (19.4%), and in comparison with loanwords from Chinese, were less evenly distributed through the 24 semantic fields.

This study has been of great value in establishing with empirical data that the Japanese lexicon has integrated an extensive number of loanwords, to the extent that it ranks 8th out of the forty-one languages studied in the Loanword Typology Project for the percentage of the core lexicon which can be identified as borrowed. Whilst the largest share of these loanwords is from Chinese, which have such a long history in the language that they are generally not treated as part of the loanword lexical stratum in the language, it is interesting that a further 6% of the core lexicon is made up of English loanwords. This shows the impact which Japanese-English has had on the language, in that a large number of English loanwords have been fully integrated into the most fundamental part of the Japanese lexicon. This study, however, is also an example of how the analysis of the grammatical features of the loanwords usually remains limited to categorising them by part-of-speech. Whilst such part-of-speech data is useful, particularly as data included in the dictionary entries of loanwords, which will be discussed in more detail in the following section, it offers only a very basic sketch of the overall grammatical behaviour of the English loanwords within Japanese.

3.4.3 *Research using dictionaries for loanword data*

The discussions above have shown that in much of the lexical borrowing literature, loanwords have traditionally been examined as free-standing, context-independent words, or short-word units (Zenner & Kristiansen, 2014). The outcome of this approach is various categorised lists of loanwords, such as lists of which loanwords appear in a language, which ones are most frequent, and which part(s)-of-speech they can be categorised into. These lists can then be utilised in the construction of dictionaries, where data such as the pronunciation, orthography, and part-of-speech of the loanwords' can be added to their individual entries. Subsequently, dictionaries have traditionally been considered the primary authoritative, quick-access reference resource for lexicographic information on the established words of a language, with many studies in loanword research relying on dictionary-based evidence in their analyses.

Dictionaries have not, however, traditionally included substantial evidence of how loanwords are integrated into the grammatical structure of the borrowing language, beyond their basic categorisation into part(s)-of-speech. They do not in general, for example, give details of the common grammatical relationships in which the loanwords participate. Such data can be gained through a corpus-based methodology which examines patterns of behaviour in natural language, but the traditional dependence on dictionaries has meant that a corpus-based approach to investigating loanwords has rarely been adopted (Zenner et al., 2014; Zenner & Kristiansen, 2014).

Dictionaries have for centuries been the principal resource of lexicographic data (Kilgarriff et al., 2014). Every major language of the world has several major publishers continually producing large, up-to-date, scholarly dictionaries of the language. For the English language, for example, publishers such as Cambridge, Oxford, and Longman all have well-established dictionary series, and for the Japanese language, Kenkyusha, Sanseido, and Iwanami Shoten offer similar resources. Dictionaries offer a quick-access means of getting authoritative lexicographic information for many areas of research, such as the first attestation of a word, and the compositional features of words such as their phonological, orthographical, and morphological form. It is therefore unsurprising that linguistic research often turns to dictionaries for lexicographic data. This is indeed the case with lexical borrowing research, where dictionaries have long been a primary source of loanword data (Zenner & Kristiansen, 2014).

An area of lexical borrowing research where the use of dictionaries has been especially common is in studies of the functions of loanwords. In this research, loanwords are typically categorised into two types of function: a lexical-gap-filling function and a special-effect-giving function (Takashi, 1990). The categorisation has traditionally been based on whether or not the loanword has an existing equivalent expression already in the borrowing language's native lexicon. When an equivalent expression is found, the loanword is considered to being adding a stylistic and pragmatic effect to the language (Onysko & Winter-Froemel, 2011), such as providing an alternative expression for taboo words in the native language, increasing the range of expression of the language, and providing a means of obscuring the direct reference of a word, a technique often employed in the speech of politicians (Stanlaw, 2004). The English loanword *un job* in French, for example, which exists alongside the native expression 'emploi', implies a more casual type of employment than the native expression. When an equivalent expression cannot be found, the loanword is considered to be filling a gap in the lexicon, providing the only viable means of expressing the

word. As an example, the word ‘sushi’, borrowed into English from Japanese, has no existing viable native-English expression, and therefore acts to fill a gap in the English lexicon opened up by the introduction of a new cultural concept.

Crucially, this categorisation into a special-effect-giving function or a lexical-gap-filling function has traditionally been made by searching for an equivalent expression in a dictionary. As an example, in one of the more recent studies on the topic, Onysko and Winter-Froemel (2011) explain how in their categorisation of the functions of English loanwords in German they “consulted different dictionaries and lexical resources to check whether a German semantic (near-) equivalent exists for each of the anglicisms [English loanwords]” (2011, p. 1556). The authors concede, however, that the categorisation based on the presence or absence of an equivalent expression is only a basic one, and “taking decisions on the general pragmatic function of an anglicism is frequently a complex task” and subsequently that “a reliable categorization crucially depends on usage-based evidence as gathered from corpora and on diachronic information” (2011, p. 1563). This is because dictionary data rarely provides details of a loanword in its context, and it is this context which is necessary for more accurately determining how a word is functioning in the language (Gries & Divjak, 2009).

In a study along similar lines, but in a different lexical borrowing context, Doi (2014) examined the naturalisation process of Japanese loanwords into the English language. As his main source of data, he used the Oxford English Dictionary (2nd ed.) and followed a methodology of examining whether a loanword could be judged as “totally foreign”, at one end of the naturalisation scale, down to “fully incorporated” at the other, based on whether the loanwords were seen in the dictionary with paraphrases and/or other explanatory devices (i.e. foreign) or by themselves (i.e. incorporated) (Doi, 2014, p. 677). In the examination, he inspected the illustrative phrases given in the dictionary as examples of the loanword in use. Similar to Onysko and Winter-Froemel’s (2011) study discussed above, his choice to use a dictionary as the data for the phrases rather than corpus data, however, led to several complications due to where the phrases had been selected from in the initial construction of the dictionary, for example only coming from specialised periodicals or encyclopaedic reference books (Doi, 2014, p. 689). Corpus data would have allowed him to more carefully analyse the contexts of the loanwords in use and would thereby have led to more reliable findings.

The lexicographic value of dictionaries in linguistic research is undeniable, but they come with many limitations. One of the largest limitations is whether or not the dictionary is corpus-based, meaning whether the details in the headword entries are

based on natural language usage (i.e. from large bodies of text known as language corpora) or the intuition and introspection of the dictionary compilers. If the entries are derived from intuition and introspection, the authoritative nature of the dictionary can be called into question because the data collection methods cannot be checked for accuracy. Adopting a corpus-based methodology in dictionary compilation, however, is something that has only become widely viable in the last few decades, in terms of cost, corpora availability, and the sophistication of the software (Kilgarriff et al., 2014).

This limitation is recognised by the author of A Dictionary of European Anglicisms (Gorlach, 2001), a comprehensive and cross-linguistic analysis of the structural and semantic integration of English loanwords into a range of European languages. The work was envisaged as an extensive lexicographic resource of over one-thousand English loanwords integrated into 16 European languages, commonly called Anglicisms in the context of English loanwords in European languages, which could be used to compare both the structural and the semantic adaptations that English words have undergone as they have been integrated into a large number of European languages (Gorlach, 2001). For each of the 16 languages into which the words have been borrowed, the work contains detailed information on the loanwords' pronunciation (including main and secondary stress), morphology (for example, inflection and pluralisation), orthography (such as variant spellings), and semantic meaning compared to the definitions given in the Concise Oxford Dictionary of Current English (9th ed.). In the introductory chapter, Gorlach (2001) states that at the time of its construction, in the 1990's, there were first of all not enough corpora for all the 16 languages from which to derive the explanations of the loanwords' style and frequency in the language, and second of all the technology of the time was not able to handle the cross-linguistic nature of the dictionary (Gorlach, 2001, p. xvi). As a result, the majority of the entries were derived from the various intuitions of the large number of compilers.

Such limitations restrict in general the value of dictionaries as evidence of word behaviour. These restrictions are compounded when analysing the grammatical behaviour of words, because grammatical behaviour can only be properly ascertained from an analysis of the word embedded in its various linguistic contexts. This can be explained as an analysis of grammatical behaviour requiring the examination of contextual units rather than single-word units, in that a word needs to be analysed in natural-language usage in a large number of the grammatical relationships in which it partakes with other words around it, and then these relationships need to be summarised in order to build up a profile of its grammatical behaviour. This type of data is not

contained in the limited entry of each headword in a dictionary. Such data can be retrieved, however, through the application of the methodology of corpus linguistics, which uses large samples of natural language for linguistic investigations, but the corpus linguistics approach is a methodology which is still rare in loanword research. Indeed, because loanwords are typically analysed as single-word units, with evidence coming from their dictionary entries, there has been no great demand for a corpus linguistics methodology in lexical borrowing research (Zenner et al., 2014).

As with loanword research in general, dictionaries have been a popular data source for various aspects of the research into English loanwords in Japanese. Hoffer (1990), for example, analysed the dictionaries themselves and summarised the historical development in the quantity of English loanwords, finding that their number increased rapidly from roughly 10,000 in the 1965 *Gaikoku Kara Kita Shingo Jiten* (Dictionary of New Words from Foreign Countries) published by Shueisha, to almost 30,000 in the 1987 *Konsaisu Gairaigo Jiten* (Concise Loanword Dictionary, fourth edition) published by Sanseido. Focusing on English loanwords recorded in general Japanese dictionaries, Daulton (2008) recorded a similar increase with the *Genkai* dictionary in 1859 containing 551 loanwords, increasing to 1428 loanwords in the 1956 *Reikai Kokugo Jiten*, and 13,300 in the 1989 *Nihongo Daijiten*.

Takashi (1990) is an example of a study in the Japanese context on the functions of loanwords using dictionary data as the evidence-base upon which the functional categorisation was made. She analysed 506 Japanese TV commercials and 413 print advertisements, finding 1523 loanword tokens in the spoken sample and 4033 in the written. For the categorisation of function, she used the traditional methodology discussed above in Section 3.7.1, of the presence or absence of a native equivalent expression in general dictionaries of the language. She used four dictionaries to ensure a wide coverage of data, three of which were specialised loanword dictionaries and the other a general Japanese-English dictionary. Her analysis concluded that 15.9% of the loanword sample were lexical-gap fillers, and the rest were being used for their pragmatic effect, such as giving products an air of modernity (1990, p. 330).

At the time of her writing in 1990, Takashi most likely was not able to exploit corpus linguistics methods for her study, so was limited in her research to using dictionaries to determine the presence or absence of a native equivalent for the loanwords. However, as was explained in Section 3.7.1 above, just because a loanword does or does not have a possible native-language equivalent expression listed in a dictionary, this cannot be used as strong evidence of the function which the loanword

fulfils in the language. Instead, actual instances of the loanwords in use need to be examined for a more accurate understanding of their functions (Onysko & Winter-Froemel, 2011).

3.5 Considerations of the Grammatical Behaviour of English Loanwords in Japanese

Whilst the traditional practices in loanword research discussed above have been put forward as contributing factors as to why the grammatical behaviour of English loanwords in Japanese remains a largely unexplored area of research, the topic has not gone entirely unexplored. This section reviews previous works which have at least in some way addressed the grammatical behaviour of English loanwords in Japanese, but as will be shown, only one of them stands as an empirically-based account of the behaviour. This is the study by Bordilovskaya (2016) and will be discussed in Section 3.6. The other treatments are limited to brief, theoretical sketches of the behaviour derived from the personal observations of the different authors, but are nevertheless important for the aspects of the behaviour which they describe. The treatments by Loveday (1996), Stanlaw (2004), and Kay (1995) are the most detailed of these and will be discussed first.

Loveday (1996) gives one of the earliest considerations of the grammatical behaviour of English loanwords in Japanese, and whilst his work is only a theoretical discussion of the behaviour, it had an important influence on the study by Bordilovskaya (2016), which, as has just been mentioned, is the only empirically-based study so far carried out in the area. In his work, Loveday puts forward his theory of the "Westernization of Japanese culture" (1996, p. 81) and explains it by giving examples of pairs of words in semantic opposition, where an English loanword refers to a Western phenomenon, and an equivalent Sino-Japanese word refers to the same phenomenon in native culture. For example, he states how *futon* 'futon' refers to the traditional Japanese place of sleep, and *beddo* 'bed' refers to the same concept but in a Western conceptualisation (p. 81). Within this theory, Loveday makes an important statement about the grammatical behaviour of the loanwords, and states that "a significant linguistic feature of such synonymous borrowings is the tendency for them to be employed in compounds rather than singly" (p. 83). He then describes this as a "morphological restriction of such loans to compounds" and says the reason is because "the fundamental or 'natural' condition of the phenomenon is made with (Sino-)

Japanese resources" (p. 83). He offers no evidence, however, to support the statement, beyond his own personal observations of several pairs of words which he suggests are in semantic opposition and does not give any details of the usage contexts in which the synonymous pairs can be found.

Loveday (1996) offers further comment upon the grammatical behaviour of English loanwords in a later section of his book titled The Grammar of Integration. The description is only a brief comment, however, about how the majority of English loanwords in Japanese are nouns and take the necessary particles to allow them to fit into Japanese syntax, and if used as adjectives, verbs, and adverbs then they are appended with a grammatical suffix such as *-na* for adjectives, and *-ni* for adverbs (pp. 118-119). Whilst his description is only theoretical, his ideas were taken up in the work of Bordilovskaya (2012, 2016) who aimed to investigate his theories through a corpus-based methodology of research into the grammatical behaviour of synonymous pairs of a loanword plus a native word (see Section 3.5 below).

In a later general investigation of English loanwords in Japanese, Stanlaw includes a chapter on the integration of English loanwords, with the aim of providing "a concise summary of the formal aspects of contemporary linguistic contact through a discussion of a number of salient linguistic features, at the levels of phonology, morphology, and syntax, relating to the nativisation of English loans in Japanese" (p. 73). He makes the important statement that "loanwords that have been incorporated into the Japanese language system generally follow the morphological and syntactic rules of Japanese grammar" (2004, p. 77). He explains that because most English loanwords in Japanese are nouns, they can be incorporated into Japanese as the object of a sentence using the objective case marker of *o* which is also used for native and Sino-Japanese words, such as *doa o akete kudasai* 'please open the door' (p. 77). He then describes how loanwords can behave like adjectives when the particle *-na* is attached, such as *furesshu na kudamono* 'fruit which is fresh' (fresh fruit) (p. 78), and in some cases the *-na* particle is dropped to make compounds from an adjective-noun combination. When used as adverbs, Stanlaw explains how the *-na* particle is changed to *-ni*, to make words such as *naisu ni* 'nicely'. Lastly, he describes the use of loanwords as verbs which is most commonly done with the addition of the auxiliary verb *suru* 'to do', such as *gorufu suru* 'to play golf' (p. 78). Whilst an important description of the loanwords' grammatical behaviour, Stanlaw's account, like Loveday's above, is based only on a small selection of loanwords and seems to be his own casual observations of loanword behaviour.

Similar to the above two studies, Kay's (1995) article on English loanwords in Japanese includes a description of how the majority of English loanwords function as nouns and can therefore be "easily incorporated into Japanese sentence structure" (p. 72), but that they can also be changed into verbs with the addition of the verbal suffix –*suru*, into adjectives with –*na*, and into adverbs with –*ni.* She also comments in the same way as Loveday (1996) about how some English loanwords only occur in compound phrases, with "their corresponding Japanese word being used to represent the meanings of the words on their own" (1995, p. 71). As an example, she states that "the loan item *fuudo*, for example, is never used on its own to refer to food in general" (p. 71). Overall, she concludes that "loanwords fit into the Japanese syntactical structure as if they were native words, being ascribed particles such as subject and object markers where necessary" (p. 72). In the same way as Loveday (1996) and Stanlaw (2004), however, no supporting evidence is given for her description and it remains only a theoretical, but still important sketch of the loanwords' grammatical behaviour.

Apart from these three somewhat more in-depth descriptions, several other important but much briefer considerations of the grammatical behaviour of English loanwords in Japanese can be found in the literature. In Shibatani's (1990) general treatment of the Japanese language, for example, a very short mention is made that "entire phrases can be composed solely in foreign loan words except for inflectional endings, particles, and other minor function words" (1990, p. 152). Writing about the impact of loanwords in Japanese advertising and the psychological effect which the loanwords can impart on the reader or listener of the phrase, he gives the two examples of *hippu o 3-senti appu-suru* ('to raise/*appu-suru* the hips/*hippu* by 3 centimetres/*senti*') and *derakkusu na puran wa kono koonaa o* ('please use this corner/*koonaa* for deluxe planning/ *derakusu na puran*'). Through this observation, he highlights the important fact that English loanwords can appear in different grammatical relationships in a sentence, such as *hippu* 'hips' being the object of a verb and designated as such by the postposition subject marker *o*; *appu* 'up' being remodelled into a verb with the addition of the verbaliser -*suru*; and *derakusu* 'deluxe' being used as an adjectival noun with the attachment of -*na*. This observation that entire grammatically-correct phrases can be constructed just from a combination of English loanwords and native grammatical markers shows the flexibility of the loanwords within the syntactical structure of the language, and importantly suggests a complexity of grammatical behaviour not given in the three more detailed descriptions above, that English loanwords in Japanese do more than just slot into lexical gaps in Japanese syntax.

Morrow (1987) takes a similar approach to Shibatani (1990) in stressing the flexibility of how English loanwords can be structured within the Japanese language, but goes a step further in attempting to account for the behaviour within the theory of code-mixing. He states that "if we look at loanwords in terms of the types of elements which are mixed, we find that they generally follow the hierarchy of exponents of mixing which Kachru (1978, 1983) has proposed" (p. 53). He then gives some examples from his own unpublished paper of the nativization of English loanwords in Japanese, but because he does not provide specific examples, it is not exactly clear what he is attempting to do in his description.

On the one hand, he may be treating the use of English loanwords in Japanese as a form of code-mixing. However, working with the definition of code-mixing given by Kachru in the papers which Morrow references, this seems problematic. Kachru defines code-mixing as developing another linguistic code which comprises formal features of two or more codes (1978, p. 8). However, in the sports writings, fashion articles, and short stories which Morrow analysed, the loanwords are not used as part of a newly-created linguistic code, but rather an integrated part of the single code of Japanese. On the other hand, Morrow might just be using Kachru's hierarchy of exponents of code-mixing to try to explain the different grammatical relationships he observed the loanwords to be participating in in his texts. However, in this case too there is a problem in that the processes involved in code-mixing are different from processes involved in the integration of loanwords into the syntax of the borrowing language. Despite the problems in the clarity of his description, Morrow's attempt to at least briefly describe and account for the grammatical behaviour of the loanwords shows the academic interest in this area of English loanword analysis.

One other consideration of the grammatical behaviour of English loanwords in Japanese to be discussed here is Hoffer (1990) who gives several instances of what he sees as "the integration of English loanwords into the Japanese grammatical system" (p. 8). He gives examples of loanwords which have been suffixed with the Japanese verbal ending of *–ru*, such as *demoru* 'to demonstrate' and *memoru* 'to take notes', and ones which have taken the adjectival ending *–i*, such as *naui* 'to be happening now' (*nau* being the loanword derived from 'now'). He also gives examples of innovative compound nouns which can be made up of two English elements, or a mix of English and a word from a different Japanese lexical stratum. However, just as with all the other considerations of how English loanwords are integrated into Japanese given in this section, he does not extend beyond these brief comments.

In defence of the lack of empirical analysis in the studies reviewed above, the researchers were limited in their analyses by the tools available to them at the time of the research. As will be discussed in the next section, it has primarily been with the development of the corpus linguistics methodology over the last few decades that the opportunity to investigate the grammatical behaviour of loanwords has properly opened up.

3.6 Towards the Corpus Analysis of Loanwords

In linguistic research over the last few decades, big advances in the tools and techniques of the methodology of corpus linguistics, where large samples of language can be quickly and reliably examined, have greatly deepened the understanding of the patterns that underlie language usage (Biber, Conrad, & Reppen, 1998; McEnery & Hardie, 2012; Thomas, 2017). The methodology of corpus linguistics has given birth to sub-fields of linguistic study, such as corpus semantics, which is an approach to language analysis that uses observational data from corpora as evidence for the meanings of words (Stubbs, 2002). In studies of corpus semantics, the focus is on using empirical observational methods to uncover the invisible meanings of words and build up lexical profiles, such as being able to discern the fact that the word 'cause' nearly always co-occurs with words which have a negative sense, as in 'cause a problem', 'cause damage', and 'cause disease' (Stubbs, 2002, p. 47). Such findings, Stubbs states, are only possible with the principle that "words should be studied, not in isolation, but in collocations" (2002, p. 45). In this sense, lexis and grammar are not separate but are interdependent, in the manner discussed by researchers such as Halliday (1985) and Sinclair (1991) who see the two at different ends of a continuum (see Section 1.3 above and again in Section 4.2 below).

This work by Stubbs (2002) on lexical profiles was further developed by researchers such as Gries (2012) into the concept of behavioural profiles, which considers the distributional properties of words on a more extensive scale. Firmly grounded in the same corpus analysis view of repeated occurrences of words within their linguistic contexts helping to reveal their various senses, he states that the "distributional characteristics of a linguistic expression reveal many if not most of its semantic and functional properties" (Gries, 2012, p. 57). Furthermore, patterns of behaviour of not just individual lexical items but across large samples of vocabulary

have shown that "distributional similarity reflects, or is indicative of, functional similarity" (Gries & Divjak, 2009, p. 59), meaning that lexical items which display similar linguistic behavioural profiles to one another are likely to be serving a similar function (Gries, 2012; Stubbs, 1996, 2002).

Corpus analyses have also gone far beyond the insightful but somewhat limited notion of collocation, to investigating entire grammar patterns that link collocations with colligations, or the frequent grammatical relationships which patterns language usage (Hunston & Francis, 2000). These patterns concern "phraseology frequently associated with (a sense of) a word, particularly in terms of the prepositions, groups, and clauses that follow the word" and again see lexis and grammar as interdependent, in that "each pattern occurs with a restricted set of lexical items, and each lexical item occurs with a restricted set of patterns" (Hunston & Francis, 2000, p. 3). Such work on identifying these patterns in which lexical items are commonly found, conducted most prominently within the Collins COBUILD project out of the University of Birmingham, has led to the publishing of two extensive dictionaries of grammar patterns, one for verbs (Francis, Hunston, & Manning, 1996) and one for nouns and adjectives (Francis, Hunston, & Manning, 1998). They list the regular grammar patterns in which verbs, nouns, and adjectives are found in the COBUILD corpus, resulting in entries such as 'N among pl-n' which structures the phrases 'a confusion among customers' and 'the fear among many' (Francis, Hunston, & Manning, 1996, p. 127). Taking the interdependence of lexis and grammar even further, Hoey (2005) uses corpus methods to argue for a new theory of language that gives a more prominent role to lexis, arguing that grammar is an outcome of lexical structure (p. 1). In this theory, language users are 'primed' in their linguistic choices by the recurring patterns in language which they have been previously exposed to.

Such extensive and detailed insights as these, in Hoey's (2005) work on lexical priming, Hunston and Francis' (2000) work on pattern grammar, and Stubb's (2002) and Gries' (2012) work on lexical and behavioural profiles have only been possible due to the advances in the computer-based analysis of language. Because these corpus methods are now available at low cost and many of them online, many areas of linguistic research, such as those discussed above, have undergone a methodological shift in how the research is conducted (Zenner et al., 2014). Crucially, however, Section 3.4.3 above discussed how the field of lexical borrowing research has yet to experience this shift towards an emphasis on linguistic context rather than just words in isolation (Zenner, Speelman & Geerarts, 2014, p. 44). Indeed, in Zenner and Kristiansen's (2014)

review of previous lexical borrowing research, discussed above in Section 3.3, they argue that the sustained focus in the field on examining loanwords as single-word units has led to several areas of loanword research being neglected. They discuss, for example, how research on borrowability (discussed in Section 3.4.2) has rarely analysed phraseological units, such as idioms (e.g. 'pimp my ride') and metaphorical phrases (e.g. 'as good as it gets') (Backus, 2014; Zenner, Speelman, and Geeraerts, 2014), because loanwords continue to be extracted from language resources as individual lexical items rather than as extended units of meaning.

The analysis of the grammatical behaviour of loanwords can be considered another area which has been largely neglected because of the continued analysis of single-word units. For whilst the analysis of the phonological and orthographical structure of loanwords can be conducted with them removed from their surrounding lexical and grammatical context (as was discussed in Section 3.4.1), the analysis of their grammatical behaviour is dependent on observing the loanwords in this surrounding linguistic environment. Poplack, Sankoff, and Miller's (1988) corpus analysis of English loanword usage in francophone neighbourhoods of Canada is one of the very few studies to address the grammatical behaviour of loanwords with corpus-based evidence. Their overall aim in the study was to investigate how a bilingual community receives loanwords into their linguistic repertoire, looking at the social factors that influence the extent of loanword usage. They also aimed to address several methodological problems which they identified in previous research on the integration of loanwords into a community: (1) the use of artificial methods of data elicitation, (2) the analysis of only a very small number of naturally-occurring borrowings, and (3) the creation of anecdotal lists of loanwords (Poplack, Sankoff, & Miller, 1988, p. 49). Their research was therefore positioned as a call to examine loanwords in their naturally-occurring contexts, by using corpus methods of analysis.

Their study involved the analysis of a sample of twenty-thousand English loanwords in French used in everyday language in francophone communities in Canada, for the extent to which they had been linguistically integrated into French and socially assimilated into the Francophone community. Very interestingly, however, their section on the syntactic integration of loanwords found that only 10 (0.05%) diverged from the regular patterns of French syntax. That leads them to the conclusion that "integration of loanwords into host-language syntactic structures is virtually categorical" (Poplack et al., 1988, p. 69). Very little information is given, however, on how they judged the extent to which a loanword deviated from French syntax, and considering the age of the

study, conducted in 1988, it was most likely only possible with a manual analysis of a sample of concordance lines.

In the context of English loanwords in Japanese, only a handful of projects have so far adopted corpus methods in the analysis. At the time of the studies discussed above in Section 3.5 which have considered the grammatical behaviour of English loanwords in Japanese, the corpus linguistics methodology was still in its infancy and mostly limited to analyses of the English language (Kilgarriff et al., 2014). Furthermore, large easily-accessible corpora of the Japanese language have only become available in the past decade (Kilgarriff, Kovar, Krek, Srdanovic, & Tiberius, 2010). Indeed, the first national, large-scale, balanced corpus of Japanese, the Balanced Corpus of Contemporary Written Japanese (BCCWJ) only became publicly available in 2011 (Maekawa et al., 2014). This means that researchers have had little choice but to be reliant on introspective analyses gained from their own intuitions and observations of the loanwords in use. Even with the development of corpus linguistics tools over the past few decades, the corpus methods employed in the work of Inagawa (2010, 2012, 2014), Bordilovskaya (2014, 2016), and Mogi (2012) whose work will be discussed below, have been limited to a manual analysis of concordance lines of a loanword in a corpus. It is only in much more recent times that the combination of sophisticated corpus analysis tools applied to very large corpora allows the researcher to quickly summarise the behaviour of thousands of instances of a loanword in its naturally-occurring contexts. This in turn opens up the possibility of analysing much larger samples of words than has been possible before.

Whilst the work of Inagawa (2010, 2012, 2014) is not specifically focused on the grammatical behaviour of English loanwords in Japanese, it is important for how she considers it to be the first corpus-driven analysis of English loanwords in Japanese. She grounds her work in the observation that previous descriptions of English loanwords in Japanese have lacked an empirical basis, as has also been discussed in the present study in Section 3.5 above. She explains how the previous descriptions are derived from the researcher's intuition, "resulting in an inability to describe the language change of the English lexicon in Japanese contexts in a systematic and informed way" (p. 380). The lack of empirical data in previous studies is what leads Inagawa to adopt a corpus-driven approach in her attempt at a "systematic and informed" investigation of the loanwords. Concerning this methodological approach, she observes that "not only are there very few corpora or corpus-based studies on *gairaigo* [loanwords] at present, relatively few are available on the Japanese language itself" (2010, p. 56).

Inagawa's doctoral thesis (2010) conducted a corpus-driven examination of diachronic changes in the meanings and frequency of usage of a small set of English loanwords in contemporary Japanese: *karaa* 'colour', *doraibaa* 'driver', *daun* 'down', *majikku* 'magic', *monitaa* 'monitor', *purinto* 'print', *masutaa* 'master', *furonto* 'front'. She states that "through concordance lines, each meaning of the English-derived words was analysed in the context where it appeared" (2012, p. 62). Her methodology first involved extracting hundreds of concordance lines for each instance of the eight words in her self-compiled corpus of Japanese newspapers, and then analysing their meanings through the construction of questionnaires consisting of the concordance lines and then a choice of possible meanings taken from the *Kojien* dictionary (2012, p. 61). The questionnaires were then answered by herself and two other native speakers of Japanese, by manually reading each instance of the loanword in its context, deciding the meaning of the loanword within each of these contexts, and matching the meaning to the options given on the questionnaire.

Bordilovskaya (2012, 2016) also employed corpus methods in her work, which stands as the only corpus analysis of the grammatical behaviour of a set of English loanwords in Japanese so far conducted. Building on her small exploratory corpus analysis of eight English colour loanwords with Sino-Japanese equivalent expressions (2012), she expanded the analysis into a doctoral thesis where she examined collocational tendencies of English loanwords in Japanese in comparison with their Sino-Japanese synonyms. In her thesis she used the Balanced Corpus of Contemporary Written Japanese (see Section 4.4.2 for details of this corpus) to search for bi-gram and tri-gram collocations of a group of 12 English loanword adjectives, in order to ascertain whether there were any collocational preferences of the loanwords compared to the Sino-Japanese words. She builds on Loveday's theory (1996), discussed above, of loanwords referring to Western concepts and the native Japanese words referring to the core concept of the word, and finds that the loanword adjectives tend to collocate more frequently with other loanwords than with native Japanese words.

Bordilovskaya's (2012, 2016) findings stand as a very important initial step in producing detailed empirical data on the grammatical behaviour of English loanwords in Japanese, and will be returned to several times during the present study. In her adoption of corpus methods, for example, where she uses an analysis of the loanwords in the Balanced Corpus of Contemporary Written Japanese to derive her findings, her work is one of only two studies so far undertaken to ground the analysis in an examination of the naturally-occurring linguistic contexts of the loanwords. That being

said, her findings are limited to a small sample of twelve adjective loanwords, eight of which are all colour loanwords. This represents a very specialised sub-sample of English loanwords in Japanese. Furthermore, as will be commented upon in the next chapter, several issues in her methods of corpus analysis limited the quality of the findings she was able to uncover.

Mogi (2012) carried out a significantly more thorough corpus-derived grammatical analysis than that of Bordilovskaya (2016), however his study only examined a single English loanword, *katto suru* 'to cut', and is further limited by the fact that only 252 occurrences of this loanword (in its verbal *-suru* form) were found in the corpus he used, which was the 2009 monitor version of the BCCWJ corpus. Similar to Inagawa (2012) and Bordilovskaya (2016), he conducted a manual analysis of the concordance lines of the loanword and ordered them by several different criteria such as their senses, post-positional particles and sentence-ending structures. Using this methodology he was able to build up what Gries (2012) terms a behavioural profile of the loanword, finding, for example, that the loanword was most typically found in a transitive construction across all of its four main senses of 'cutting up something, such as food', 'cutting hair', 'deleting something, such as text', and 'cutting off something, such as a road' (2012, p. 27).

Whilst his study only analyses a single loanword, Mogi (2012) was able to build up a detailed picture of its grammatical behaviour which he states has applied value in areas such as the teaching of Japanese as a foreign language and lexicography. Importantly, he sees the need to build up an extensive bank of individual grammatical analyses of loanwords, into what he calls a "dictionary of grammatical patterns for learners of Japanese" (2012, p. 22). Whilst not available to him at the time of his research, corpus methods have now developed to the point where what he carried out manually can be done automatically, in seconds rather than hours. The Sketch Engine corpus analysis software, for example, allows the automatic generation of a loanword's grammatical and collocational behaviour in the form of 'word sketches', and their production only takes only a few seconds. In the use of these word sketches as the main data source for the grammatical behaviour of English loanwords in Japanese, the research in the present study builds on this important work of Mogi (2012) and is positioned to address what he describes as a need for more work in this area:

> "Descriptive research on loanword verbs, adjectives and nouns, especially research examining their syntactic behaviour, is therefore needed both from a general linguistic point of view, in order to have a better description of this part of the Japanese lexicon, as well as from an applied point of view, to obtain basic data from which applied linguistic research such as lexicography or Japanese language teaching could greatly profit" (Mogi, 2012, p. 23)

Building on these previous corpus analyses of English loanwords in Japanese and exploiting the most recent advances in the automated analysis of the distributional characteristics of words in large corpora, the present study aims to provide extensive data on the grammatical behaviour of English loanwords in Japanese. Furthermore, rather than focusing on a small set of loanwords, as has been shown above to be the case with both the theoretical sketches and empirical analyses conducted in this area previously, the present study conducts an observation of patterns across a large sample of loanwords made possible with advancements in the size and availability of Japanese corpora and in the software tools which can be used to query them for patterns (Cheng, 2012; Kilgarriff et al., 2014; Lindquist, 2009; McEnery, Xiao, & Tono, 2006). The final section in this chapter presents the research questions which will guide this study, and the following chapter describes the corpus methods used in their investigation.

3.7 Aims and Research Questions

The present study aims to examine the grammatical behaviour of frequently-used English loanwords in naturally-occurring written Japanese texts. To investigate this grammatical behaviour of the loanwords, the following three research questions are asked:

1. What patterns of distribution can be observed in the grammatical relationships of English loanwords in Japanese?
2. In what ways are these patterns of distribution similar to and different from patterns of distribution in the grammatical relationships of non-loanwords in Japanese?
3. What factors appear to account for observed differences in patterns of distribution in the grammatical relationships of loanwords and non-loanwords in Japanese?

Frequently-used English loanwords are defined in this study as those which occur most frequently in three Japanese corpora, and naturally-occurring Japanese texts are defined as the language contained in the jpTenTen11 web corpus. The next chapter gives details of the language resources, tools, and methods of analysis used for the investigation of these research questions.

4 Data and Methodology

4.1 Chapter Overview

This chapter details the data and methodology used in the investigation of the grammatical behaviour of English loanwords in Japanese. It first considers the overarching view of language in which the research is positioned, and then describes the corpus methods approach taken in the analysis. This description addresses several critical issues in the adoption of a corpus linguistics methodology, concerning the nature of the Japanese corpus used in the research and features of the software used in its analysis. Details are then given of how word sketches were produced in the Sketch Engine corpus analysis software for the main sample of English loanwords, as well as the comparative sample of non-loanwords, and how the grammatical relationships in each word sketch were recorded and collated into two databases. The creation of the lists of 587 frequently-used English loanwords and 130 frequently-used native and Sino-Japanese words for which the word sketches were created is then described, and the chapter ends with an explanation of how the tables of gramrels presented in Chapter Five were constructed from the databases.

4.2 The Methodological Approach of the Research

This research views lexis and grammar as interdependent, as what has been termed lexico-grammar (Halliday, 1985), where language is seen to be realised by a connected series of lexico-grammatical patterns. Investigating these patterns manually is complicated. Not only are language users often unaware of the lexico-grammatical patterns in their own language usage, but more critically, manually collecting together a sufficient number of patterns in order to uncover behavioural tendencies is a labour-intensive and error-prone task (McEnery & Wilson, 2001). The development of the methodology of corpus linguistics, particularly over the last few decades, has allowed lexico-grammatical patterns to be explored on a scale and to a depth of analysis that was not previously feasible by manual methods (Biber et al., 1998; McEnery & Hardie, 2012; Meyer, 2004).

The research to be carried in this study views loanwords not as individual, free-standing lexical items, or single-word units (Zenner & Kristiansen, 2014), but rather as

contextual units made up of the loanword embedded in their naturally-occurring linguistic contexts. Importantly, these contextual units are not simply collocational pairs, which has been a major focus of many corpus studies (Cheng, 2012; Stubbs, 2002), but rather a fusing of collocations (i.e. word relationships) and colligations (i.e. grammatical relationships, cf. Hoey, 2005). One of the most widely-used corpus-based resources of such a fusion of collocations and colligations is the word sketch function in the Sketch Engine corpus analysis software (Kilgarriff et al., 2014; Srdanovic et al., 2011; Thomas, 2017). This one-page, corpus-based automatic summary of a word's grammatical and collocational behaviour, described as a "stunning overview of a word's paradigmatic and syntagmatic relations" (Thomas, 2017, p. 14), represents a much richer analysis of the behaviour of a word than what has been possible with the more standard output of corpus software: the Key Word in Context (KWIC) concordance lines (Cheng, 2012).

The present study adopts an approach which carries out a qualitative analysis of these quantitative word sketches of English loanwords in Japanese. The qualitative approach follows that described by Richards (2003) and what he terms the "aspects of analysis" (2003, p. 271). The fundamental underpinning of the analysis involves an ongoing interaction between categorisation and interpretation, something which is common to qualitative research in general (Silverman, 2006), and includes a description of the data, an analysis of features in the data, and an interpretation of what the features mean. Within these actions, Richards stresses the importance of "breaking down and recombining the data in an effort to build a picture that will respond to the aims of the research" (2003, p. 270) and that such a process involves the analysis being exploratory in nature (p. 271). He outlines these multiple and non-linear interactions in diagrammatic form, shown in Figure 4.1.

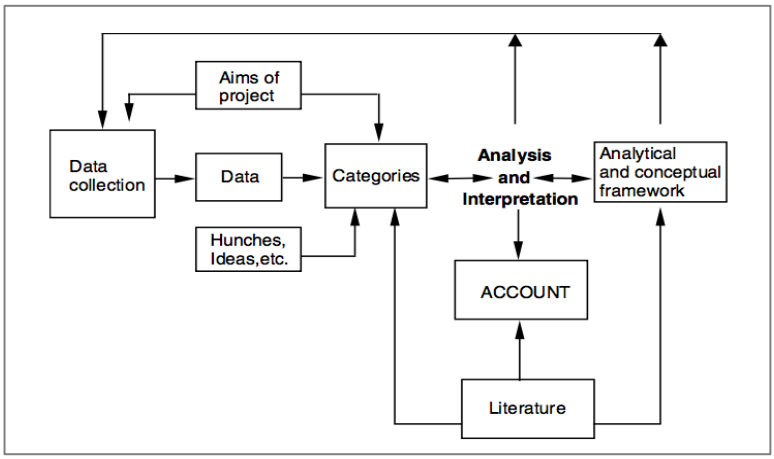


Figure 4.1 The aspects of analysis in qualitative inquiry (from Richards, 2003, p. 271).

Richards discusses how the central element of the diagram is where 'categorisation' connects 'data' and 'interpretation' together, and furthermore that accounting for the data comes out of a link between the researcher-based interpretation of the data itself and references back to the existing literature (2012, p. 271). Within this process, the role of the researcher is to collect the data, think about the data with reference to how it applies to the aims of the project, categorise the data and then reflect on the categorisations, reorganise the categories to see the data in different ways, link discoveries from the data categorisation to existing theories, and then collect further data on the basis of the insights gained (2012, p. 272).

Framed within this approach to qualitative analysis, the present study involves the following aspects of analysis, which will each be elaborated upon in the subsequent sections. The research first involves collecting together data on the grammatical relationships of a sample of English loanwords and a sample of native and Sino-Japanese words. This data will then be processed into two databases and compared in order to draw out patterns of behaviour that can be identified in the grammatical distribution of the loanwords. The findings will be further explored with an additional analysis conducted on a sub-sample of the loanwords. The findings from these analyses will then be related to previous literature describing various categorisations of loanwords, to see if the different patterns of behaviour in the loanwords in the present study can match with pre-existing categories from other studies. Before describing these different aspects of the analysis, the following sections first address important considerations that need to be borne in mind when adopting a corpus linguistics methodology and relying on the automated production of linguistic data.

4.3 Evaluating Word Sketches in the Sketch Engine Corpus Analysis Tool

In order to investigate the patterns of distribution in the grammatical relationships of English loanwords in Japanese, the present study uses a word sketch created for each of the loanwords. As introduced above, and also in Chapter One, a word sketch is defined as a corpus-based, one-page summary of a word's grammatical and collocational behaviour (Kilgarriff et al., 2004, p. 1). They are produced in the Sketch Engine corpus analysis software, an online corpus query tool which became publicly available in 2004. The word sketches in this study are created from the jpTenTen11 corpus, an 8-billion-word web corpus which was created as one member of a large family of multi-billion-word corpora to be used within the Sketch Engine software. Because all of the findings presented in Chapter Five and discussed in Chapter Six are derived from these Sketch Engine word sketches, this section discusses some of the principal issues underlying the word sketch composition. These issues concern the nature of the jpTenTen11 corpus upon which the word sketches are built, and the ability of the Sketch Engine software to produce accurate and informative word sketches.

4.3.1 *An evaluation of the jpTenTen11 corpus*

The 8-billion-word jpTenTen11 web corpus of Japanese is part of the TenTen family of web corpora within the Sketch Engine corpus query software, a group of around 10-billion-word corpora for thirty-three languages, including English (enTenTen12), Korean (koTenTen12), and Spanish (esTenTen11) (Jakubicek, Kilgarriff, Vojtech, Rychly, & Suchomel, 2013). The corpora in the family were all constructed using a similar process of first crawling the web for language using the Spiderling web crawler, then using the jusText corpus cleaning tool to remove unwanted text (called boilerplate) and non-textual material, and then de-duplicating repeated text at the level of duplicated paragraphs using the corpus de-duplicating tool called Onion (Jakubicek et al., 2013, p. 1). Language-specific tools were used for the tokenisation of each corpus to deconstruct it into tokens, lemmas, and parts of speech. For all members of the TenTen family, meta-information is available about the top-level domain, website, web domain, URL, word count, and length of each document included in the corpus.

One of the main concerns of web corpora, such as the jpTenTen11, is the issue of their representativeness, a term used in corpus linguistics to define the extent to which a

corpus represents some variety of a language, such as newspaper or academic language (Biber et al., 1998). Representativeness, along with authenticity, balance, sampling, and size are considered the fundamental aspects of what makes a corpus a corpus instead of just an unprincipled collection of texts, and because of this, if the representativeness of a corpus cannot be established, by documenting the kinds, number, length, and selection of texts it includes, the value of the corpus for linguistic study is questioned (Gatto, 2014). The appearance of web corpora constructed from automatically crawling millions of webpages and extracting their language has raised the question of whether "the web as corpus makes the notion of a representative corpus redundant" (Leech, 2007, p. 144). The answer is that the appearance of the web corpus has forced a reconceptualisation of the notion of representativeness in corpus linguistics, in that the "scope, variety and, above all, its immense size seem to legitimize the opinion that these characteristics can counterbalance the limits of representativeness" (Gatto, 2014, p. 45). In particular, the extremely large size of most web corpora has been used to argue that their representativeness is achieved not by "meticulous proportional sampling" (Brezina & Gablasova, 2015, p. 7), but by the coverage of a massive amount of online material.

As such, compared to traditional corpora of text developed along principled means of ensuring the representativeness of the texts which are included in the corpus, such as the British National Corpus, web corpora typically contain much less metadata on what texts have gone into the corpus, for example information about the text genre and demographic characteristics of the texts' author. Significant efforts were made in the Sketch Engine software, however, to not only compile each member of the family of TenTen corpora along the same general lines, but to also record exactly where on the web the texts were drawn from. Figure 4.2 presents the overall meta-data for the jpTenTen11 corpus, showing that the 15,553,141 documents in the corpus come from 76 top level domains (i.e. the end part of a web address) covering 340,505 websites.

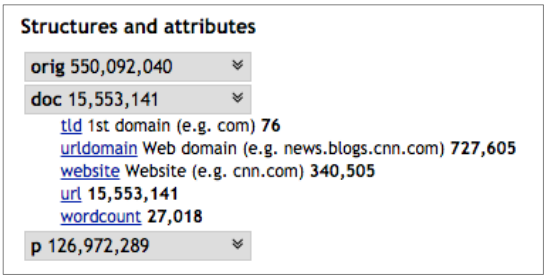
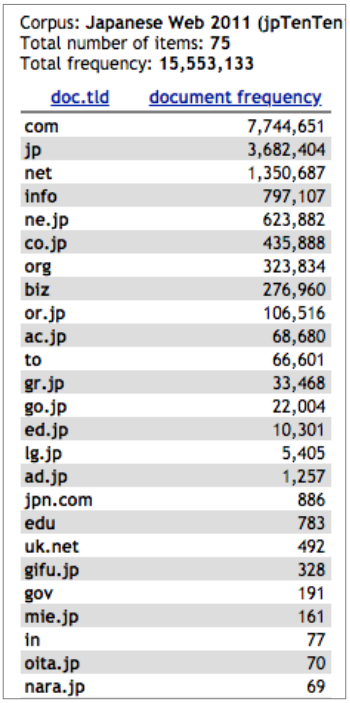


Figure 4.2 Meta-data for the contents of the jpTenTen11 corpus

Looking at this metadata more closely, it is possible to get details of all the 76 top level domains from where documents were crawled as well as links to each individual document in each domain. Figure 4.3 shows the first 25 of these 76 top level domains for the jpTenTen11 corpus. The left-hand column shows the domain name, such as '.com', '.jp', and '.net'; and the right-hand column shows the number of documents in the corpus coming from each domain. This gives a basic but very important overview of the diversity of the corpus, which, in addition to size, is a major consideration in its representativeness (Biber et al., 1998). For example, it can be seen that the domain names cover a wide variety of areas, such as '.ac.jp' and '.ed.jp' being related to academia, '.biz' and '.ord' being related to business, and '.gifu.jp', and '.gov' being related to government and politics.



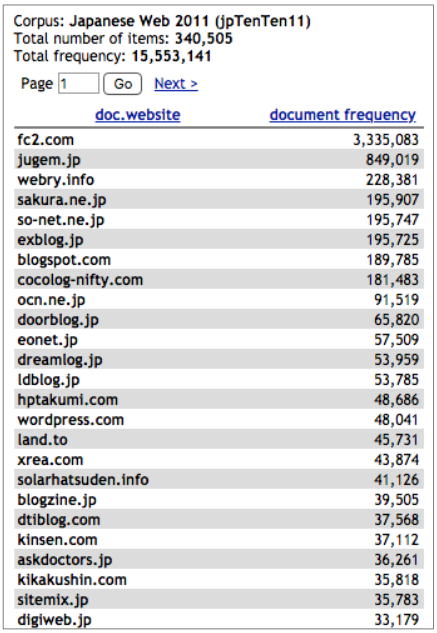
Corpus: **Japanese Web 2011 (jpTenTen**
Total number of items: **75**
Total frequency: **15,553,133**

doc.tld	document frequency
com	7,744,651
jp	3,682,404
net	1,350,687
info	797,107
ne.jp	623,882
co.jp	435,888
org	323,834
biz	276,960
or.jp	106,516
ac.jp	68,680
to	66,601
gr.jp	33,468
go.jp	22,004
ed.jp	10,301
lg.jp	5,405
ad.jp	1,257
jpn.com	886
edu	783
uk.net	492
gifu.jp	328
gov	191
mie.jp	161
in	77
oita.jp	70
nara.jp	69

Figure 4.3 The first 25 top level domains of documents in the jpTenTen11 corpus.

Looking more specifically into these domains, Figure 4.4 shows the top 25 websites from where the documents were crawled. Similar to the data on the top level domains, this meta-data on the websites can give an overview of the various text types in the corpus. The website 'fc2.com' is a personal blogging service, 'sakura.ne.jp' is a web-hosting service for websites often used in Japan by private businesses, such as 'sushi.sakura.ne.jp' which is a private business selling software that teaches touch-typing skills, and 'askdoctors.jp' is an online medical consultation service. Overall, the

meta-data available for the jpTenTen11 corpus allows it to be seen that this corpus is representative of general web language.



Corpus: **Japanese Web 2011 (jpTenTen11)**
Total number of items: **340,505**
Total frequency: **15,553,141**

Page 1 Go Next >

doc.website	document frequency
fc2.com	3,335,083
jugem.jp	849,019
webry.info	228,381
sakura.ne.jp	195,907
so-net.ne.jp	195,747
exblog.jp	195,725
blogspot.com	189,785
cocolog-nifty.com	181,483
ocn.ne.jp	91,519
doorblog.jp	65,820
eonet.jp	57,509
dreamlog.jp	53,959
ldblog.jp	53,785
hptakumi.com	48,686
wordpress.com	48,041
land.to	45,731
xrea.com	43,874
solarhatsuden.info	41,126
blogzine.jp	39,505
dtiblog.com	37,568
kinsen.com	37,112
askdoctors.jp	36,261
kikakushin.com	35,818
sitemix.jp	35,783
digiweb.jp	33,179

Figure 4.4 The first 25 websites of documents in the jpTenTen11 corpus.

Because of its ultra-large size, one challenge involved in the analysis of the corpus is the fact that searches for individual words often return tens of thousands of concordance lines. It is therefore a complex task not only to analyse these concordance lines, but also to summarise the different text types from where the concordance lines are extracted. The Sketch Engine software does, however, allow the user to create a frequency list of which websites contain the documents where a search word was found. As will be explained below, the functionality of the word sketch option in the Sketch Engine software addresses this issue by using various means of data summarisation, such as applying a word association statistic in the ranking of collocates of the search word.

4.3.2 *An evaluation of the word sketch function*

The main resource necessary for the analysis of patterns in distribution of the grammatical relationships of English loanwords in Japanese is a collection of the linguistic contexts of the loanwords. Collecting a sample of these contexts manually

was not considered appropriate for the research. The number of contexts which could be feasibly collected and recorded by hand would be severely constrained by time, and the analysis and summary of more than a small number of contexts would be mostly beyond the capabilities of a human analysis by a single researcher. Using corpus methods of analysis largely overcomes these difficulties, by routinising the collection, analysis, and summary of the contexts (McEnery & Hardie, 2012). With a large corpus and advanced tools of analysis, tens of thousands of instances of the linguistic context of a word can be searched for, recorded, and summarised in a few seconds.

This section gives details of the word sketch function in the Sketch Engine corpus query software, which allows the rapid summation and linguistic annotation of tens of thousands of instances of a word and its surrounding lexical context. At the time of the research, the Sketch Engine's word sketch function was the only tool available which could automatically identify, analyse, summarise, and categorise such a massive amount of data, and in particular, such a massive amount of Japanese language data. Other corpus software, such as the AntConc corpus analysis toolkit, can also process Japanese language data, but the user needs to upload their own corpus and the software does not include defined grammatical relationships within which the grammatical behaviour of the loanwords could be categorised. Similarly, other software such as WordSmith Tools involves the user uploading their own corpus and has a limited range of text-analysis functions compared to the Sketch Engine. At its core, Sketch Engine focuses on a combination of lexical and grammatical analyses of the search word, making it the most powerful publicly-available software for the analysis of the grammatical behaviour of loanwords at the time the research was conducted.

The word sketch function of the Sketch Engine corpus software generates an overview (i.e. a sketch) of the grammatical and collocational behaviour of a search word in all of its contexts in a specified corpus of language. Because of the detail available in these word sketches and their value in a wide-range of linguistic fields, in particular lexicography and language education, the Sketch Engine has become one of the most widely-used corpus query tools for linguistic research (Kilgarriff et al., 2014; Thomas, 2017). A major advantage of the Sketch Engine over other corpus-analysis software is the large number of corpora which can be analysed directly within the program. As was described in the previous section, a group of web corpora constructed for the world's major languages, the TenTen corpora, is contained within the software and includes the jpTenTen11 corpus of web-based Japanese.

As an illustration of the richness of the jpTenTen11 corpus, Table 4.1 shows a comparison of the number of instances in the BCCWJ and jpTenTen11 corpora of five of the most frequent English loanwords in Japan. Taking the loanword システム *shisutemu* 'system' as an example, the frequency per million words of this loanword is almost exactly the same in the two corpora, at 157. However, the raw frequency figures show that there are almost 100-times more instances of the loanword in the jpTenTen11 than in the BCCWJ.[8] The larger the number of instances of a word in a given corpus, the more information there is available to analyse, and the more information there is available, the richer the sketch will be of the behaviour of that word.

Table 4.1 A comparison of the frequencies of five common English loanwords in the BCCWJ and jpTenTen11 corpora.

	BCCWJ			jpTenTen11	
Loanword	**English**	**Freq.**	**Freq./mill**	**Freq.**	**Freq./mill**
ページ	page	24,642	235.6	1,871,432	181.3
システム	system	16,458	157.3	1,627,507	157.7
サービス	services	16,630	159.0	1,805,679	174.9
ブログ	blog	10,205	97.6	1,805,454	174.9
テレビ	TV	15,644	149.5	1,692,378	164.0

Word sketches are produced in the Sketch Engine by analysing and grouping large amounts of tagged corpus data (e.g. part-of-speech, lemma information etc.) and are intended to show a "full and complete account of a word's grammatical and collocational behaviour" (Kilgarriff et al., 2010, p. 373). This data is shown in a word sketch in two ways: (1) by grammatical relationships, or what the Sketch Engine calls gramrels, which are shown horizontally across the word sketch and explain how a word interacts with other words in its local context, and (2) by a ranking within each gramrel of a word's most strongly-associated collocates.

The gramrel columns in the Japanese word sketches are ordered by the number appearing in the header of each gramrel. In Figure 4.5 below, which shows the word sketch of the English loanword ビジネス *bijinesu* 'business', the number appearing in the header of the first gramrel column (which has been highlighted in a yellow box) is

[8] A reason for this large discrepancy is that it is most likely an outcome of the types of corpora involved. The jpTenTen11 corpus is a corpus of web language and is therefore likely to have a high number of technological words, such as 'system', 'page', and 'blog'.

62.14. This expresses the percentage of all the 761,770 occurrences of this loanword which appear in the [noun/noun] grammatical relationship. In other words, the loanword *bijinesu* appears in this corpus in a [noun/noun] grammatical relationship 62.14% of the time. Continuing along the columns, *bijinesu* occurs in a [particle] grammatical relationship 31.91% of the time, a [を(*o*) verb] gramrel 9.04% of the time, and so on. The [noun/noun] gramrel can therefore be considered the most preferred gramrel for the loanword *bijinesu* from all the different grammatical relationships it participates in throughout the corpus. For the present study, the first gramrel column in the word sketch will be called the 'first most preferred gramrel', the second gramrel column will be called the 'second most preferred gramrel', and so on. Using this terminology, the first most preferred gramrel of *bijinesu* is [noun/noun], and the second most preferred gramrel is [particle]. This terminology is very important for explaining the results in Chapter Five and will be returned to in Section 5.2 in the next chapter.

Figure 4.5 A word sketch of the loanword *bijinesu* 'business' showing its grammatical relationships and collocations.

The other coloured boxes in the first gramrel column in Figure 4.5 show the other principal parts of a word sketch. The red box highlights the gramrel name, in this case [noun/noun]. The green box highlights the ranked collocates of *bijinesu*, with an example of the collocation given under each collocate. The orange box shows the frequency of each collocate in that specific grammatical relation, and the purple box shows the logDice statistic for the strength of association between *bijinesu* and the respective collocates. By default the collocates in each gramrel column are ranked by this logDice association score, which measures the amount of statistical association between two words (Rychly, 2008, p. 6). This logDice score will be discussed in more

detail below, after introducing one other important aspect of the word sketches which are utilised in the loanword analyses in Chapter Six, the concept of the Longest-Commonest Match.

The Longest Commonest Matches (LCM) can be seen in Figure 4.5 in the light-grey text under each of the collocates. The LCM is a concept developed by one of the creators of the Sketch Engine software to show the longest, commonest realisation of the pairing of the search term with its collocate in the corpus (Kilgarriff et al., 2014; Thomas, 2017). They are generated by searching for strings longer than two words, which include a lemma + a grammatical relationship + another lemma, which are used multiple times in the corpus. If a certain string of words accounts for a high proportion of the instances of the search term in the corpus, then it becomes an LCM candidate (Kilgarriff et al., 2014). The software continues to search for other triples and then returns the one which is longest and commonest in the corpus. As will be seen in Chapter Six, sometimes the LCM is only a bi-word collocate, and this occurs very frequently in the [noun/noun] gramrel which, as will be explained in Chapter Five, expresses a noun compound. These bi-word collocates usually do not express anything more than what is already seen in the ranked list of collocates. However, the LCM becomes very revealing in other gramrels, such as particles, where it often takes the form of 3 or 4-word phrases. These help in giving a more explicit view of how the word combines with its collocates in the grammatical relationships.

Looking more specifically at the collocates listed in each of the grammatical relationships, it can be seen that their ordering is according to their logDice score rather than their raw frequency. This is the default setting in the word sketch function. The concept of collocate ranking is another critical issue in corpus linguistics, in that the method used to rank the collocates greatly influences the collocates which will be returned in the analysis. This issue is addressed by the developers of the Sketch Engine who provide in the help pages of the software a detailed explanation of the various statistics used. Whilst other functions of the Sketch Engine such as the collocations option within the standard search function allow the user to choose the statistic used to rank the collocates, such as T-score, MI, log likelihood, and logDice, the word sketch function only allows an ordering of collocates by raw frequency or logDice. Even without the user needing to choose from amongst a range of statistics, it is important to understand what kind of statistic this default selection of logDice is, and how it compares with other widely-used statistics.

A statistical association measure is defined as "a formula of an association score which indicates the amount of statistical association between two words" (Rychly, 2008, p. 6) and a range of association measures are available to rank collocates. Ranking by raw frequency alone does not express the strength and/or statistical significance of association between the node (i.e. search term) and collocate, whilst an association measure, on the other hand, compares what has been observed about the co-occurrence of the node and collocate with what would be expected under the null hypothesis, which is the assumption that the node has no statistically significant influence over the words that surround it (McEnery & Hardie, 2012). The most commonly-used measures of association in modern studies which adopt a corpus linguistics methodology are the MI, T-score and log-likelihood measures (Hunston, 2002; Lindquist, 2009; McEnery & Hardie, 2012). Each measure, however, has a weakness that gets amplified as the corpus gets bigger, and is something that the researcher needs to be aware of when utilising one or other of the measures (Lindquist, 2009; McEnery et al., 2006). The weakness of MI is that it tends to highlight collocates which are rare in the corpus, whilst the weakness of T-score and log-likelihood is that they tend to highlight function words and punctuation (McEnery & Hardie, 2012). Because of this, many corpus analysis software programs such as AntConc allow the user to exclude function words and punctuation from the analysis.

With web corpora, these weaknesses are amplified because the massive size of the corpora increases the amount of rare words, which affects the MI measure, and increases the frequency of function words and punctuation, which affects the T-score and log-likelihood measures. An example of this is given in Table 4.2 showing the top-25 collocation candidates for the highly-frequent search term 'zero', produced in the Sketch Engine from the enTenTen12 English-language web corpus.[9] A span of five words to the left and the right of the search word was used to define the range from within which the collocates could be extracted. The table shows the different collocates which were returned when different ranking methods were chosen. As can be expected, the raw frequency list gives little to work with when attempting to build up a semantic profile of the word 'zero', as it highlights only function words and punctuation. Furthermore, because the T-score and log-likelihood measures are biased towards frequent words in the corpus (and a ten-billion-word corpus gives a lot of data), these

[9] Whilst the present study is focused on the Japanese language, an English corpus was used for this example to avoid needing to translate all of the collocates.

scores produce lists dominated by function words and punctuation items, and which differ little from the raw frequency list. However, log-likelihood is the better of the two, as it includes three words which start to give a sense of how ‘zero’ is used in the corpus (Ground, cost, tolerance). For the MI score, which is biased towards low-frequency words, of which there are many in web corpora, the list of collocates produced by this measure seems to be dominated by company names, web addresses and spelling mistakes. Indeed, the average frequency of these collocates is 24.6, which in a ten-billion-word corpus is extremely low.

When the collocates are ranked by logDice, which is a variant form of the Dice score that fixes the issue of the scores being very low numbers (Rychly, 2008, p. 6), the list is markedly different. The logDice measure brings out collocates which give a clear overview of the variety of ways in which ‘zero’ is being used. A check of a sample of the concordance lines of each collocate showed that ‘Ground’ very often refers to ‘Ground Zero’ (which also explains the frequent occurrence of ‘Zero’ with a capitalized ‘Z’), the collocates ‘emissions’ and ‘waste’ show that zero is being used to talk about the reduction and control of something, and ‘cost’, ‘interest’, ‘tolerance’ and ‘gravity’ reveal that zero relates generally to an absence of something. This brief analysis highlights the problem of applying the most common association measures to large-scale web-crawled corpora. The massive scale of word tokens in large-scale web corpora, such as the enTenTen12, most of which will be function words, renders the T-score and log-likelihood measures largely ineffective at doing anything more than ranking function words and punctuation high up on the collocate lists. And the fact that web corpora contain many slogans, company names, web addresses and spelling errors means that the MI score is similarly problematic.

Table 4.2 Top 20 collocates of the word 'zero' ranked by their raw frequency and then several association measures.

Rank	Raw Frequency		MI		t-score		log-likelihood		logDice	
1	.	189438	xpeople	15.311	.	395.043	.	641005.57	Ground	9.176
2	,	144808	dougpositive	15.044	,	334.836	,	402171.22	zero	8.782
3	the	133602	Paymydownpayment	14.851	the	314.867	to	345614.6	tolerance	8.713
4	to	111724	mask-charge	14.851	to	301.849	the	332816.48	Zero	7.841
5	and	85901	pointsaffiliations	14.851	and	257.555	is	235870.91	Double	7.653
6	a	77624	UnionsAbsolutely	14.851	a	249.234	and	225711.1	cost	7.251
7	of	75102	Kiyona	14.851	is	239.744	a	222246.29	emissions	7.235
8	is	66836	tazzari	14.851	of	237.233	of	181910.05	Waste	7.175
9	in	57581	budgetaug	14.851	in	214.61	in	162256.11	gravity	7.087
10	with	40860	degreesbelow	14.851	with	188.077	with	145225.34	ground	7.061
11	for	37856	Wait-State	14.851	for	174.854	Ground	131636.04	degrees	6.846
12	that	34553	predecession	14.851	on	169.934	cost	129623.48	reset	6.808
13	on	33961	Fome	14.819	that	163.719	zero	121395.94	carbon	6.783
14	have	27222	Tsukaima	14.805	have	153.514	on	113066.01	near	6.614
15	or	26143	lygerzero	14.787	or	150.576	for	108046.88	balance	6.612
16	you	25362	Craigslitst	14.752	(	147.93	have	95975.606	interest	6.601
17	be	25200	childrenreadingbookswithparents	14.714	at	147.68	tolerance	93109.086	below	6.552
18	(	24940	superif	14.699	)	144.059	or	92703.602	sum	6.531
19	at	24784	distributelab	14.681	be	143.394	at	92661.343	percent	6.527
20	The	24461	MeltDown	14.681	The	141.809	(	92233.513	mosque	6.493

The above discussion has shown the logDice score to be an effective association measure at bringing high quality collocates to the top of collocation candidate lists. This explains why it is used as the default measure to rank collocates in the Sketch Engine's word sketch function. It has other strengths including the fact that it is not corpus-specific because it does not depend on corpus size, so a logDice score from one corpus can be compared to the score in another corpus of different size (Rychly, 2008). Also, the theoretical maximum score is 14, which means it is much easier for the user to comprehend than some of the very large and very small numbers given by other measures (Rychly, 2008, p. 9). Another key feature of logDice in the Sketch Engine software, which is to be discussed in more depth below, is that word sketches return collocates based on grammatical relations whilst the collocation function of a simple search query uses a range of +/- *n*-words. The logDice measure is particularly suited to the former method of extracting collocates, further explaining its use as the default measure used to rank collocates in the word sketches.

4.3.3 *An evaluation of the quality of the word sketches*

A word sketch in the Sketch Engine software is built upon a set of grammatical relations, called gramrels, which are defined in a language-specific sketch grammar. A sketch grammar is defined as a "mini-grammar of syntactic patterns" allowing the various functions of the Sketch Engine, including word sketch, thesaurus, and sketch

diff, to identify relationships between a search word and other words (Erjavec, Erjavec, & Kilgarriff, 2008, p. 13). Different corpora of the same language can use different sketch grammars. For example, the Japanese sketch grammar for the jpWaC web corpus contains 40 pre-defined grammatical relations while the one for the jpTenTen11 corpus contains an updated and expanded set of 157.

The grammatical relations in a sketch grammar are regular expressions built upon part-of-speech tags, meaning that a tagger is needed for each language to establish the set of tags. For the part-of-speech tags, the jpTenTen11 corpus was processed with the MeCab tagger and morphological analyser combined with the UniDic Japanese dictionary tool (The tagset is given in Appendix 1). This software was used in the construction of the Balanced Corpus of Contemporary Written Japanese, a government-funded corpus of general Japanese modelled on the British National Corpus, and achieved high rates of accuracy (Maekawa et al., 2014, p. 357). Because of its speed and precision, it has also been used by the technology company Apple in the Japanese input for the Mac OS and iOS operating systems. Figure 4.6 shows all of the 157 possible gramrels contained in the Japanese sketch grammar for the jpTenTen11[10], and Figure 4.7 shows the interface of the word sketch tool with its default settings.

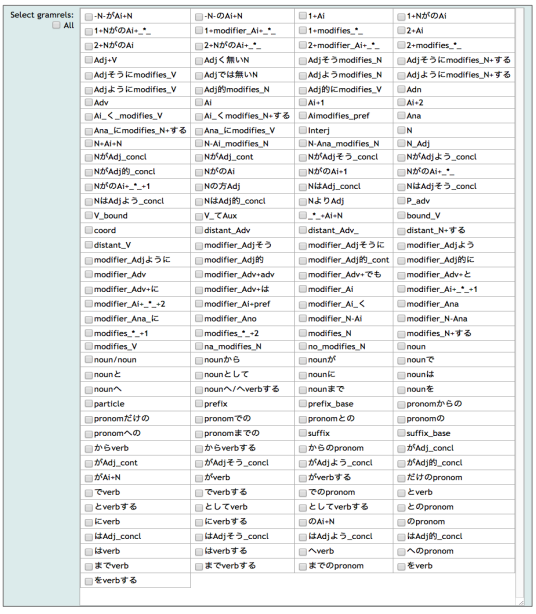


Figure 4.6 The full set of 157 gramrels used in the production of word sketches from the jpTenTen11 corpus.

[10] In the Japanese sketch grammar, if a particle is used in the gramrel name the software maintains the Japanese name and script of the particle. In Chapters Five and Six, explanations of the meanings of these gramrels are given where necessary.

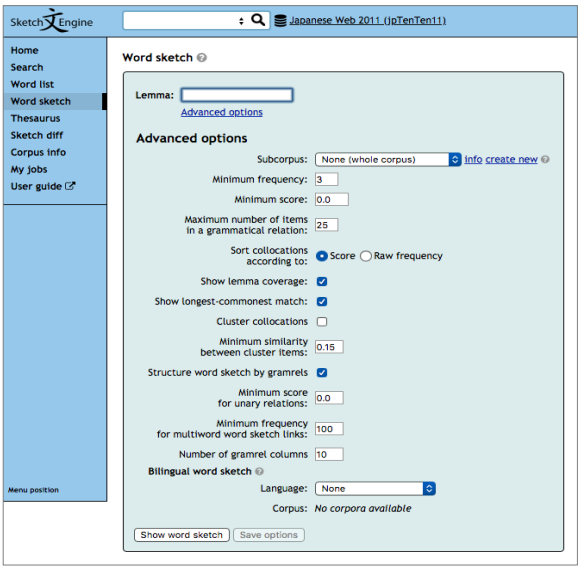


Figure 4.7 The default interface of the word sketch tool in the Sketch Engine.

As was introduced above, the gramrels defined in a sketch grammar use regular expressions over part-of-speech tags. Using the example given in Erjavec et al. (2008), a basic adjective modifier relation is defined as follows:

=modifies
2:[tag="Ai"] 1:[tag="N.*"]

In this example, '1:' indicates the search term (a noun in this example) and '2:' indicates the collocation candidate which will be captured in the specific grammatical relation (an adjective). This means that when an adjective (Ai) is followed by a noun (N), the adjective modifies the noun. The jpTenTen11 sketch grammar includes more complex definitions where duals are recognised. Using another example from Erjavec (2008), the [modifier_Ai] and [modifies_N] relations are duals of each other and are combined into the complex regular expression of:

*DUAL =modifier_Ai/modifies_N
2:[tag="Ai.*" & word!="ない|無い"] [tag="Pref.*"]? 1:[tag="N.*" & tag!="N.Suff.*" & tag!="N.bnd.*"]

When the word sketch function is employed for a specific word, these patterns are run across the corpus and when matches are found, they are stored and summarised into a word sketch. Searching for a noun, for example, would bring up a list of adjectives (Ai) in a [modifier_Ai] relation with the noun (N). On the other hand, searching for an adjective would bring up a list of nouns that are in a [modifies_N] with the adjective. The salience between the adjective and noun, and vice versa, would be represented by the logDice statistic described earlier, and the nouns/adjectives would be ranked accordingly within the [modifier_Ai]/[modifies_N] gramrels of the relevant word sketch.

In this way, Sketch Engine gramrels are algorithms which return triples of a head word + grammatical relation + collocate. This is markedly different from the more standard corpus queries which return a search term within a pre-defined range of language, organised into concordance lines, as well as two-word collocates typically returned from a pre-defined range, such as four or five words to the left and right of the search term (Kilgarriff & Tugwell, 2001; McEnery & Hardie, 2012). Figure 4.8 shows part of the word sketch for the Sino-Japanese noun 抱擁 *houyou* ‘embrace’. The oval shape highlights the [modifier_Ai] gramrel column and the box shape highlights the first listed collocate 熱い *atsui* ‘hot/warm’, making the collocation of ‘warm embrace’. In this collocation, the adjective collocate ‘warm’ is in a [modifier_Ai] relation with the noun ‘embrace’. The same collocation can be found in the word sketch for the adjective 熱い, shown in Figure 4.9. Here, the oval shape highlights the [modifies_N] gramrel and the second listed collocate, highlighted in the box shape, is 抱擁 *houyou* ‘embrace’. The collocation is again ‘warm embrace’, but this time the word sketch shows the noun collocate ‘embrace’ in a [modifies_N] relation with the adjective ‘warm’.

Figure 4.8 A word sketch for the Japanese noun 抱擁 *houyou* 'embrace'.

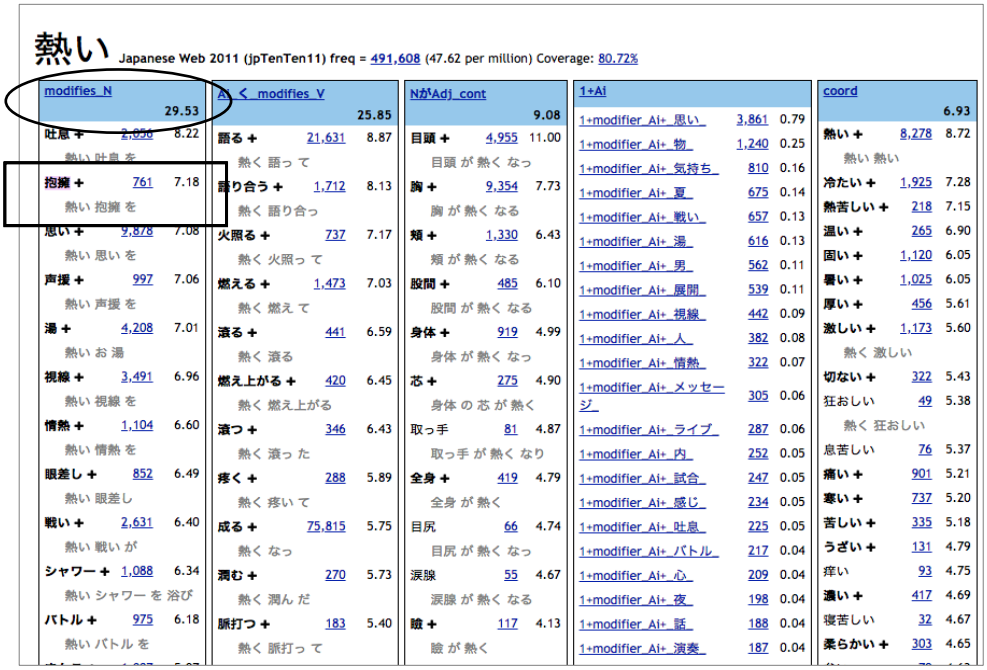


Figure 4.9 A word sketch for the Japanese adjective 熱い *atsui* 'hot/warm'.

The quality of the word sketches, which can be judged by how much of all the possible data they return (recall) and how accurate the data is (precision), is a further critical issue in the automatic summary of large amounts of corpus data (Kilgarriff, 2005). For the more frequent words of a language, a word sketch summarises the grammatical relationships of tens of thousands, sometimes hundreds of thousands, of instances of the word in use. Such extensive data summarisation is all but impossible to do manually, which means the user needs to be able to trust the automatic system to return high quality data. To test the quality of the word sketches, the developers of the system conducted an analysis based on the question of whether the high level of 'recall' of data in a word sketch brings with it a low level of 'precision' because of the increased chance of errors appearing in the data (Kilgarriff et al., 2010, p. 374).

Kilgarriff et al. (2010) investigated this question in the word sketches for the Dutch, English, Slovene, and Japanese TenTen corpora, examining if the word sketches produced in the Sketch Engine software could be judged as suitable for inclusion in a published collocation dictionary. They initiated the Sketch-Eval project to explore the issue and prepared a custom version of the Sketch Engine for each corpus which included the top-20 highest-scoring collocations of a random sample of 42 headwords (a mix of nouns, verbs, and adjectives). For each collocation they created a menu whereby an evaluator could rate the quality of the collocation. For the Japanese word sketch evaluation, the headwords were randomly extracted from three frequency levels (high, mid, and low) of the most common 30,000 nouns, verbs, and adjectives in the jpWaC corpus.[11] Three evaluators were used to evaluate a total of 747 collocations. Two of the evaluators were native speakers of Japanese (both language teachers and linguists) and one a non-native speaker (also a language teacher and linguist). There was a 3-way agreement between the evaluators for under half of the collocations (294), with all three agreeing that 278 (94.6%) were of good quality (i.e. of publishable quality) and 16 (5.5%) were of bad quality (i.e. not publishable). There was 2-way agreement for 690 of the collocations. Of these, 600 (86.95%) were rated as good and 90 (13.05%) as bad.

The main reason for disagreement between the evaluators was seen to be that one evaluator consistently judged semantically incomplete collocations as bad, whilst the others rated this type of collocation as good. An example given by Srdanovic et al. (2011) is *zaikai no yuuryoku* ('influential in financial circles') where the suffix *sha* ('person') has been cut off from *yuuryoku* ('influential') by the morphological analyser.

[11] At the time the Sketch-Eval project was conducted, the jpTenTen11 corpus had not been compiled. The jpWaC corpus is the precursor to the jpTenTen11 corpus.

The full collocation should therefore be *zaikai no yuuryokusha* 'an influential person in financial circles'. As such, examining the concordance lines for this collocation would reveal that it is technically a good collocation, but because it appears as incomplete in the word sketch, one evaluator judged it as bad. The same issue was seen in the other languages in the Sketch-Eval project, for example in the evaluation of the English language word sketches 'put' was judged as a bad collocate of 'cat' when the word 'out' was not included in the evaluation (i.e. 'put the cat out') (Kilgarriff et al., 2010, p. 378). Other problems reported in the Sketch-Eval of the Japanese word sketches were related to part-of-speech tagging errors (e.g. splitting the derivational suffix *-na* from its free morpheme which changes the part of speech of the word from an adjective to a noun), issues with the various scripts which can be used to write Japanese words (e.g. a single word can be written in various scripts), polysemic words (e.g. the word *matsu* can be both a noun with several different meanings as well as a verb) (Srdanovic et al., 2011).

With the fast, automatic processing of large amounts of linguistic data, inaccuracies in the results such as those discussed above are to be expected. However, the Sketch-Eval project not only showed that the vast majority of the collocations where there was 3-way and 2-way agreement were judged as being of publishable quality, but the percentage of publishable-quality word sketches from among the four languages in the project was highest for Japanese (Kilgarriff et al., 2010; Srdanovic et al., 2011). As such, the Japanese word sketches can be judged as high-quality automatic summaries of the grammatical and collocational behaviour of a word, and are used as the main data source for the grammatical behaviour of English loanwords in Japanese for the present study.

4.3.4 *Producing and analysing word sketches for Japanese vocabulary*

In order to address the first research question of what patterns of distribution can be observed in the grammatical relationships of English loanwords in Japanese, a word sketch was produced in the Sketch Engine software using the jpTenTen11 corpus for a sample of English Loanwords in Japanese. This sample contains 587 loanwords and the creation of this list is discussed below in Section 4.4. The sketch grammar used to produce word sketches in the jpTenTen11 corpus contains 157 gramrels (see the previous section), and if a loanword were to occur in all of them, then they would all be listed automatically in the word sketch. Recording every gramrel in which each of the

587 loanwords occurred was considered unnecessary because many of the gramrels were infrequent. For example, the full word sketch for the loanword *bijinesu* has a total of 54 gramrels, many of which occur infrequently. The combined frequency of the collocates that join with *bijinesu* in a [へverb] gramrel, where the loanword is followed by the particle *e* (indicating a direction of motion) and then a verb, is 145; whereas for the [noun/noun] gramrel, where the loanword is the modifying part of a noun compound, the frequency is 473,338.

For this reason, the decision was taken to limit the number of gramrels which would be recorded for each loanword. The word sketch is presented on screen to the user by default in rows of 10 gramrel columns, and for the present study the decision was taken to record the top row of 10 gramrel columns. From a manual check of a sample of 25 word sketches, it was observed that even in just these top-10 gramrels, there was a rapid drop-off in the total frequencies of the collocates in each gramrel. For example, the *bijinesu* word sketch mentioned above is found in a [noun/noun] gramrel 62.1% of the time, decreasing to 2.4% with the tenth gramrel of [*de* verb]. It was therefore decided that recording the top-10 gramrels in the word sketches was not only convenient due to the way in which the word sketch was presented on screen, but that this approach would also cover the most salient aspects of the grammatical distribution of the loanwords (this point will be returned to and further explained in Section 5.2 in the next chapter). In this way, however, not recording every gramrel of each loanword was a limitation of the research, but one which was considered necessary because of the way in which the gramrels were recorded manually.

The top-10 listed gramrels shown in the word sketch for each of the 587 loanwords (i.e. the first ten columns in each of the 587 loanword word sketches) were recorded in a spreadsheet, resulting in a database of 5870 gramrels. As is shown in Figure 4.10, the loanwords are listed vertically in the spreadsheet and then the top-10 grammatical relationships of each loanword are listed horizontally. This means that the spreadsheet can be analysed in several ways to answer the first research question. Firstly, an analysis can be made at the level of each individual loanword. The row of an individual loanword from the 587-word sample can be selected and its top-10 gramrels can be checked. Secondly, an analysis can be made at the level of the whole sample of loanwords. One of the 10 columns can be selected to check all of the different gramrels covering the whole loanword sample. Doing this for each of the columns and removing duplicate gramrels results in a final list of which of the 157 possible gramrels the loanwords were actually found to occur in. Using as an example the sample of 100

gramrels shown in Figure 4.10, such an analysis of the whole loanword sample would result in two different gramrels in the first gramrel column (i.e. G01 in the spreadsheet), three in the next column, and so on. Taking all the columns together and removing the duplicates results in 17 unique gramrels that cover the 10 gramrels of these 10 loanwords. The results of the full analysis are given in the next chapter.

	Loanword	English	G01	G02	G03	G04	G05	G06	G07	G08	G09	G10
1	ページ	Page	particle	noun/noun	o verb	pronom no	no pronom	ni verb	Adn	de verb	o verbsuru	suffix
2	コメント	Comment	particle	distant_Adv	o verb	noun/noun	pronom no	ga verb	ni verb	no pronom	de verb	modifier_Adv
3	システム	System	particle	noun/noun	o verb	no pronom	o verbsuru	pronom no	ni verb	to verb	suffix	ga verb
4	サービス	Service	particle	noun/noun	o verb	distant_Adv	no pronom	o verbsuru	pronom no	ni verb	suffix	to verb
5	ブログ	Blog	particle	noun/noun	o verb	pronom no	de verb	ni verb	Adn	no pronom	o verbsuru	de verbsuru
6	テレビ	TV	particle	noun/noun	de verb	no pronom	o verb	pronom no	ni verb	suffix	coord	de verbsuru
7	センター	Center	particle	noun/noun	no pronom	ni verb	de verb	suffix	o verb	to verb	pronom no	coord
8	ゲーム	Game	particle	noun/noun	o verb	no pronom	pronom no	suffix	Adn	ni verb	de verb	to verb
9	データ	Data	particle	noun/noun	o verb	pronom no	o verbsuru	no pronom	wa verb	ga verb	ni verb	to verb
10	バック	Back	noun/noun	particle	distant_Adv	ni verb	no pronom	o verb	pronom no	de verb	to verb	ga verb

Figure 4.10 The ten most-frequent gramrels of ten English loanwords in Japanese.

The spreadsheet database of 5870 gramrels can then be examined for the frequency of each gramrel. As with the analysis of the types of gramrels just described above, the database can again be examined at the level of an individual loanword and the level of the whole sample of loanwords. Using Figure 4.10 above again as the example, the row of a loanword can be selected and the top-10 gramrels can be checked. For example, selecting and examining the loanword ページ *peeji* 'page' would show that [particle] is the gramrel in gramrel column 1, [noun/noun] is in gramrel column 2 etc. This kind of analysis does not give much information at the level of a single loanword, but when a count is made of the different gramrels in each column from G01 to G10, in other words, the number of gramrel tokens, then a detailed picture of the frequency distribution of the loanwords amongst the different grammatical relationships can be constructed. For example, such an analysis of column 'G01' in Figure 4.10 would show that there are nine tokens of the [particle] gramrel and one of [noun/noun]. Having counted the total number of tokens of each gramrel in each column, an overall count for each gramrel across all of the 587 loanwords can be calculated by combining the number of tokens in all 10 columns. This would create a ranked list of the most common gramrels in which the loanwords occur. The results of this analysis are shown in the next chapter.

To address the second research question of how the patterns of distribution in the grammatical relationships of English loanwords are similar to and different from

patterns of distribution in the grammatical relationships of non-loanwords in Japanese, a word sketch was produced for a sample of non-loanwords in Japanese (i.e. native and Sino-Japanese words). This sample contains 130 non-loanwords and the creation of this list is discussed below. This comparative analysis is necessary because as Hoey (2005) states in his corpus-based analysis of the grammatical relationships of the word 'consequence', "the raw figures or percentages of occurrence in each grammatical position will by themselves tell us little about the colligational [grammatical] preferences of consequence" (2005, p. 45). He explains that the grammatical distribution of a particular search word needs to be compared with that of other similar words in order for meaningful information to be provided. In this way, word sketches produced for the loanwords were to be compared to those created for the sample of the non-loanwords.

It was explained in Section 2.2.4 that the Japanese lexicon is stratified into three main types of vocabulary: native words, Sino-Japanese words, and loanwords. A sample of the most frequent native and Sino-Japanese words was collected, this time using the loanword list published in the Frequency Dictionary of Japanese (Tono, Yamazaki, & Maekawa, 2013). This list was chosen not only because is it derived from the Balanced Corpus of Contemporary Written Japanese, one of the corpora used in the present study to derive the list of loanwords (see Section 4.4 below), but also for the practical reasons that it is available with indexes of words organised by lexical stratum and part-of-speech. For the sample, it was decided to include only native and Sino-Japanese nouns. This is because almost all English loanwords in Japanese are treated as nouns, with other parts-of-speech being realised with grammatical markers appended to the base noun-form, such as *-suru* for verbs and *-na* for adjectives (Irwin, 2011).

Initially, a 100-word sample was considered sufficient for the comparative analysis, as this would result in a database of 1000 grammatical relationships. Therefore, the most frequent 100 non-loanword nouns listed in the dictionary were initially selected. It was then decided to add a stratified sample of the five most frequent nouns at each 100-word frequency band up to the limit of the most frequent 1000 words. This decision was taken so as to include a wider range of frequent native and Sino-Japanese words, and not just the very-most frequent (i.e. only the top-100). Because the most frequent 100 nouns ranged from rank 17 to 350 of all the 5,000 words in the dictionary, the stratified sample of the remaining 100-word frequency bands (up to rank 1000) included an extra 30 words (five words from six frequency bands). In exactly the same way as for the list of English loanwords, a word sketch was produced

for each of the 130 native and Sino-Japanese words, resulting in a database of 1300 non-loanword grammatical relationships. This database could then be examined in the same ways as for the loanword sample described above. The full list of 130 non-loanwords is given in Appendix 3.

To address the third research question of factors which could account for any differences in patterns of distribution in the grammatical relationships of loanwords and non-loanwords in Japanese, the grammatical relationships were examined more closely for more intricate patterns of behaviour. In conducting the analysis to answer the first two research questions, several important features of the loanwords' grammatical behaviour were noticed which affected how the research was to be conducted to answer research question three. Firstly, it was noticed that for the large majority of the loanwords, only two gramrels were observed as the first most preferred and second most preferred gramrels of the loanwords (i.e. the gramrels in the first and second columns in the word sketches). It was also noticed that in the majority of cases, when one of these two gramrels was listed as the first most preferred gramrel, the other was listed as the second most preferred gramrel. This behaviour was different from that of the native and Sino-Japanese words. Furthermore, it was noticed that for some loanwords the first most preferred gramrel was many times more preferred than the second, but for other loanwords the difference was much smaller. The decision was taken to explore this issue in more detail.

To give an example of this issue of degree of difference in frequencies, Figure 4.11 shows the first five gramrel columns of the word sketch for the loanword レンタル *rentaru* 'rental'. The figures in the heading at the top of each column (which is the percentage of all occurrences of this loanword in each gramrel) show that this loanword is found in a [noun/noun] gramrel over 4-times more often than in a [particle] gramrel.

Figure 4.11 The word sketch of *rentaru* 'rental'.

The word sketch shown in Figure 4.12 for the loanword ステーション *suteeshon* 'station', on the other hand, shows a very similar frequency between the first two gramrel columns.

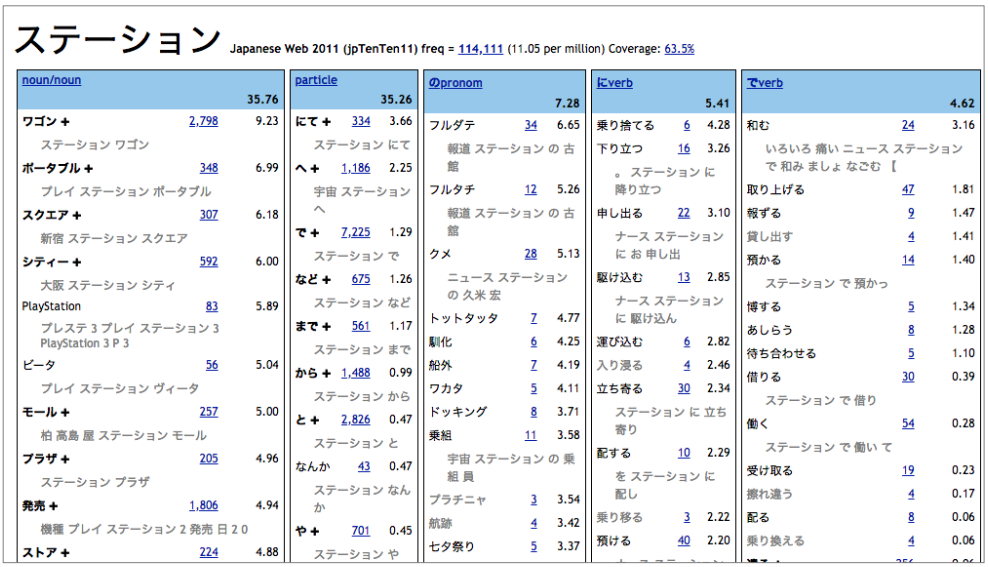


Figure 4.12 The word sketch of *suteeshon* 'station'.

To analyse the extent to which some grammatical relationships are more preferred than others, two analyses can again be conducted using the spreadsheet of gramrels. Figure 4.13 shows the same data as Figure 4.10 with the addition of the total number of occurrences of all the collocates in the first and second most preferred gramrels, and

then a percentage of how much bigger the value of 'FREQ G1' is than 'FREQ G2', calculated by dividing the frequency of G1 by G2, subtracting 1, and turning into a percentage.[12] Using this data, information can again be looked up for an individual loanword by selecting a row and examining the frequencies and percentage difference. For all of the loanwords in the sample, the last column of 'G1:G2', can be used to sort the loanword by the degree of difference between the first and second most preferred gramrels. The results of this analysis are given in Chapter Six.

	Loanword	English	G 01	G 02	G 03	G 04	G 05	G 06	G 07	G 08	G 09	G 10	FREQ G1	FREQ G2	G1:G2
1	ページ	Page	particle	noun/noun	o verb	pronom no	no pronom	ni verb	Adn	de verb	o verbsuru	suffix	900,164	489,960	0.84
2	コメント	Comment	particle	distant_Adv	o verb	noun/noun	pronom no	ga verb	ni verb	no pronom	de verb	modifier_Adv	384,645	190,839	1.02
3	システム	System	particle	noun/noun	o verb	no pronom	o verbsuru	pronom no	ni verb	to verb	suffix	ga verb	801,341	458,975	0.75
4	サービス	Service	particle	noun/noun	o verb	distant_Adv	no pronom	o verbsuru	pronom no	ni verb	suffix	to verb	796,901	468,852	0.70
5	ブログ	Blog	particle	noun/noun	o verb	pronom no	de verb	ni verb	Adn	no pronom	o verbsuru	de verbsuru	877,163	552,800	0.59
6	テレビ	TV	particle	noun/noun	de verb	no pronom	o verb	pronom no	ni verb	suffix	coord	de verbsuru	760,266	719,993	0.06
7	センター	Center	particle	noun/noun	no pronom	ni verb	de verb	suffix	o verb	to verb	pronom no	coord	473,641	440,989	0.07
8	ゲーム	Game	particle	noun/noun	o verb	no pronom	pronom no	suffix	Adn	ni verb	de verb	to verb	916,516	707,851	0.29
9	データ	Data	particle	noun/noun	o verb	pronom no	o verbsuru	no pronom	wa verb	ga verb	ni verb	to verb	712,535	447,529	0.59
10	バック	Back	noun/noun	particle	distant_Adv	ni verb	no pronom	o verb	pronom no	de verb	to verb	ga verb	412,410	280,759	0.47

Figure 4.13 The 10 most preferred gramrels of the 10 most frequent English loanwords, with frequency figures for the first and second most preferred gramrels.

Creating the samples of English loanwords and of native and Sino-Japanese words, producing word sketches for each word in the samples, and recording their most frequent grammatical relationships in a spreadsheet allows for the subsequent searching of patterns of distribution. Details of the patterns of distribution which were identified in these databases are presented in Chapter Five and explored in more detail in Chapter Six. The following section gives details of the creation of the list of frequently-used English loanwords in Japanese for which the word sketches were generated.

4.4 Creating a List of Frequently Used English Loanwords in Japanese

To investigate and account for patterns of distribution in the grammatical relationships of frequent English loanwords in Japanese, three principal language resources were needed: (1) a list of frequently used English loanwords, (2) a large collection of the loanwords' naturally-occurring linguistic contexts, and (3) a comparative sample of frequently used words from other Japanese lexical strata along with their contexts.

[12] In the database, the numbers in the 'G1:G2' column were kept as raw figures (i.e. not multiplied by 100) for ease of reading.

Details of the latter two of these resources were given above, and the following sections focus on the creation of a list of frequently-used English loanwords in Japanese. Two criteria were used for a loanword to be included in the list: (1) a loanword had to appear frequently in natural Japanese language, and (2) it had to appear in three separate Japanese language resources. These criteria were to help ensure that the loanwords were established words in the Japanese language, rather than one-offs, also termed nonce borrowings or occasionalisms (Haspelmath, 2009). Whilst several lists of loanwords in previous research studies and in published lexicographic resources were available for use, the following section discusses why they were considered unsuitable.

4.4.1 *Previous lists of English loanwords in Japanese*

The most comprehensive list of frequent English loanwords available for use at the time of the research was the List of Common Loanwords Corresponding to the British National Corpus 3000, produced in a study into the potential of an in-built English lexicon in the Japanese language (Daulton, 2008). This list contains 1808 English loanwords, grouped into 1356 word-families, which had been extracted from a newspaper corpus and could be matched with one of the 3000 most-frequent word families in the British National Corpus, known as the BNC 3000 (Nation, 2004). The list was made available as an appendix at the back of the study (Daulton, 2008) and represents a substantial sample of English loanwords in Japanese.

The methodology behind the list's construction, however, included several decisions that made it unsuitable as a resource for the present study. Firstly, the loanwords were drawn from a specialised corpus of *Mainichi Shinbun* Japanese-newspaper articles from 2001, meaning they were loanwords from a specialised written genre. Secondly, the list was filtered specifically to retain only those loanwords which corresponded to the most frequent 3000 words from the British National Corpus. Also, 300 loanwords were added to the list based on the intuitive judgements of an individual native-speaker of Japanese, and then filtered based on self-appraisal assessments of whether the words were well-understood by university students. Particularly the latter methodological process and the fact that no frequency data was given for the loanwords meant it was not considered a suitable resource of frequent English loanwords in Japanese.

Other published lists of English loanwords in Japanese were available in works by Miura (1979) and Webb (1990). Miura's collection of over 400 English loanwords is a self-selected list of loanwords which he felt needed a more detailed description than was offered in the *Kadokawa Gairaigo Jiten* and *Konsaisu Gairaigo Jiten* loanword dictionaries, and several other glossaries of loanwords. He sums up his strategy by saying "in the last analysis, however, it was my own subjective judgement that I relied on for the selection of words, for the words chosen were the ones about which I felt there was something worthwhile to say" (1979, p. 15). As such, no statistical information is available in this list about the frequency of the loanwords, or about in what type of Japanese language usage they appear, and the collection was not considered appropriate as a resource of modern contemporary frequent English loanwords. Similarly, Webb's (1990) list includes around 950 words mainly from English, but also from some other European languages, which the author self-selected as being "difficult for a native speaker of English to understand when he/she encounters them for the first time" (p. 7). Although this list is fairly extensive at nearly 1000 loanwords, its lack of statistical information about the words made it inappropriate for use in the present study.

A recently-published general list of Japanese vocabulary was available in A Frequency Dictionary of Japanese (2013), part of the Routledge series of frequency dictionaries of core vocabulary for language learners. This resource gives a list of 5000 of the most commonly-used words in the contemporary Japanese language, based on the 100-million-word Balanced Corpus of Contemporary Written Japanese. In its attempt to offer learners and teachers a resource of the most frequently-used words in modern Japanese, it potentially offered a comprehensive, empirically-derived resource of loanwords. However, it was discovered that the list contained an abnormally low number of loanwords, considering its coverage of 5000 highly-frequent Japanese words. For example, in one of the indexes in the back that lists words by their lexical strata, only eleven words are included under the *gairaigo* category (i.e. non-Chinese loanwords). This represents only a 1.1% sample of the list, whereas many previous studies put the number of loanwords in general Japanese at between 8-12% (Daulton, 2008; Hoffer, 2002; Loveday, 1996; Stanlaw, 2004).

When investigating the reason for the low number of loanwords in the list, a statement was found in the introduction explaining that "words that appeared in the original frequency list but were considered inappropriate for the wordlist were deleted, e.g. archaic words, single letters of the English alphabet, specific company names (e.g.

Sony), personal names (e.g. 信長 [nobunaga], English words (e.g. アンド [ando/and]), too domain-specific terms, etc" (p. 5). The use of the general term 'English words' as a category for words which were deleted is not further qualified in the book, so this issue was further investigated in personal communication with one of the authors. The author explained that English words which were considered to not be fully integrated into the Japanese lexicon, such as *ando* 'and', were excluded from the list. This answer, however, did not include an explanation of the criteria upon which the inclusion or exclusion of a loanword as fully integrated into the Japanese lexicon was made. Because of this uncertainty in the way in which loanwords were handled in the creation of the list, but more importantly because of the very low number of English loanwords appearing on the list, it was considered unsatisfactory for use in the present study.

One other resource of loanwords was potentially available for use. Loanwords in Japan constitute a separate stratum of the Japanese lexicon, known as 外来語 *gairaigo* 'words coming in from outside', meaning that many specialised loanword dictionaries have been published to record the borrowings, such as the Concise Dictionary of Katakana Words (2010). However, as with the other lists described above, there are several issues with loanword dictionaries which made them an unsuitable resource for the current research. Firstly, the majority of the loanword dictionaries group all loanwords together, so the words borrowed from Dutch, Russian, French, and Italian, for example, are grouped together with those from English and ordered in the dictionary alphabetically. Further, the selection of loanwords to include in the dictionaries is often intuition-based, as has been traditionally the case with loanword dictionary construction in general (McEnery et al., 2006, p. 81). Added to this is the fact that frequency data of each loanword in the language is not typically given in the dictionaries, and also that the dictionaries often record tens of thousands of loanwords. For example, the fourth edition of Sanseido's Concise Dictionary of Katakana Words (2010) contains 48,100 loanwords, derived from various different languages, and with no frequency data. Because of this, dictionaries of loanwords in Japan are better positioned as a reference resource for looking up the meaning(s) of the words, rather than for information about their usage in the language. Because of these issues, it was considered necessary to gather together a new collection of frequently-used English loanwords in Japanese. The main advantage of creating a new list was that all stages of the list's construction could be documented, making the reasons for a loanword's inclusion or exclusion explicitly clear. The sections below describe the compilation of a new, corpus-derived list of frequent English loanwords in Japanese.

4.4.2 *Using corpora as a resource of loanwords*

The most common modern method of constructing new word lists is to derive them from corpora (Hunston, 2002; McEnery & Hardie, 2012; Meyer, 2004; Nation & Newton, 1997). Extracting word lists from corpora is considered one of the principal activities in corpus linguistics, with the ability to gain detailed frequency information of lexical items considered to be a major affordance of using computers for the analysis of language (Cheng, 2012, p. 40). Deriving a word list from a corpus (or multiple corpora) allows not only the specified type of words to be extracted (for example, all nouns; transitive verbs; adjectives in an attributive position; core words of the language etc.), but also for frequency information to be attached to each word (Hunston, 2002, p. 67).

The development of corpora and corpus-based lexicographic resources for the Japanese language has been significantly slower than that of other Asian languages such as Chinese (Kilgarriff et al., 2014; Srdanovic et al., 2011). As an example, whilst corpus-based dictionaries of Chinese began to appear in 2007, Kilgarriff notes that even in 2014 "a truly corpus-based monolingual dictionary of Japanese is yet to appear" (Kilgarriff et al., 2014, p. 8). The first large-scale, comprehensive Japanese corpus, the Balanced Corpus of Contemporary Japanese (BCCWJ), was made publicly available in 2011, sixteen years after the general release of the British National Corpus on which the BCCWJ is modelled. Up to this point, the vast majority of Japanese corpora were newspaper archives, as with the 40,000-sentence Kyoto University Text Corpus of Mainichi newspaper articles; specialised text databases, for example the Aozora Bunko corpus which is similar in form to the Project Gutenberg literary text archive; or specialised spoken corpora, as with the Corpus of Spontaneous Japanese (CSJ) which is a 7.5 million-word corpus of 658 hours of transcribed speech from 1,400 Japanese speakers.

Of the limited Japanese corpora available, three were identified as suitable for use in the present study because they were large-scale, easily-accessible (i.e. low cost and available online) and had been used in previous corpus analyses of the Japanese language. The first corpus to be chosen was the 100-million-word Balanced Corpus of Contemporary Written Japanese (BCCWJ). This corpus was constructed along principled methods (see Section 4.4.2.1 below) making it a well-balanced and representative corpus of written Japanese. Word-frequency lists from the corpora have been published online as part of a range of language resources packaged with the release of the corpus. The jpWaC and jpTenTen11 corpora of web-based Japanese were

also chosen because they are very large resources of Japanese, at 400-million words and 8-billion words each and are both included in the Sketch Engine software, making it possible to use the same software and techniques to extract the word-frequency lists from each corpus. Furthermore, the main research of this study uses functions within the Sketch Engine to analyse the English loanwords in their contexts, making it appropriate to use corpora which are already housed in the software. The BCCWJ and jpWaC corpora are described in detail below, and Section 4.3.1 above gives details of the jpTenTen11 corpus.

The Balanced Corpus of Contemporary Written Japanese (BCCWJ) is a widely-used corpus in Japanese lexicographic research (Maekawa et al., 2014). It is a 104.3 million-word corpus of contemporary written Japanese, constructed between 2006 and 2011 by the National Institute for Japanese Language and Linguistics (NINJAL). It covers the years 1976 to 2005 and is comprised of three sub corpora: (a) *publication* (books, magazines, newspapers), (b) *library* (books), and (c) *special purpose* (white papers, textbooks, newsletters, laws etc.). In its design, it was constructed around the three guiding principles of (a) being a corpus of written Japanese only, (b) using random sampling for the inclusion of texts, and (c) being made available to the public (Maekawa et al., 2014). Spoken data was excluded from the corpus for reasons of cost, time, and the fact that it was easier to randomly sample written texts for which the sample population could be better-defined (Maekawa, 2007; Maekawa et al., 2010).

The corpus has been used in a wide range of investigations of the Japanese language. It has been used as a database in the investigation of specific grammatical constructions in Japanese, such as the sentential ending *n deshita* (Nishi, 2017); as a resource of annotated texts in reading-speed eye-tracking research (Asahara, Ono, & Miyamoto, 2016); and as a source of collocational data (Srdanovic, 2014). Whilst the corpus is still one of the only balanced corpora of written Japanese available, one of its drawbacks is its limited size. At around 100-million words, the corpus is only of medium size in comparison with large-scale web corpora such as the jpTenTen11, which contains around 8-billion words. Whilst this was not a problem for the extraction of a list of English loanwords, which because of their frequency in general Japanese discourse are still very likely to appear in the word-frequency list of any sized corpus of Japanese, it was an issue in being able to provide enough instances of the loanwords in context in order to investigate patterns of their grammatical behaviour. As was explained in Section 4.3.1, this was one of several reasons why the jpTenTen11 corpus was used for the research into the loanwords' grammatical distribution.

A more recent trend in Japanese corpora, mirroring that of many other languages, is the development of web corpora. Within the last ten years, this area of corpus linguistics has developed into its own field, known as Web as Corpus (or, WaC), where the focus is on "methods that look at the web as their main resource to implement the corpus linguistic approach" (Gatto, 2014, p. 7). The first large-scale, widely-used web corpus of Japanese was the jpWaC, a 400-million-word corpus derived from 50,000 web pages. The corpus was compiled with the same methods used to create an earlier family of Web as Corpus (WaC) corpora. This family includes web corpora of English (ukWaC), German (deWaC) and French (frWaC), and was developed using open-source tools for web corpora creation including BootCaT to crawl the web for text. The construction process of the jpWaC began with translating the top 500 non-function words in the British National Corpus into Japanese, and turning them into random 4-tuples used to gather URLs (Erjavec et al., 2008, p. 4). This was followed by downloading the web pages, normalising character encodings, extracting the metadata, cleaning up the text, and annotating the corpus with linguistic information (Erjavec et al., 2008, p. 4). The jpWaC corpus was then made publicly available through the web-based Sketch Engine corpus-query software.

Because of the vast scale of web corpora, and the nature of their composition involving a large amount of automated processing and a lack of genre labels for the different texts, there have been concerns over their general validity for linguistic research (Gatto, 2014; McEnery & Hardie, 2012). As such, web corpora are normally subjected to a variety of tests in order to show that the language inside a web corpus is what was intended to be there (i.e. a broad collection of online texts), and not just "a partial and distorted view of a language" (Erjavec et al., 2008, p. 7). For the jpWaC, at the time of construction there was no large-scale balanced corpus of the Japanese language, so a comparison was made by the compilers with one of the more specialised Japanese corpora: The Mainichi Newspaper 2002 corpus. The two corpora were subjected to a frequency-profiling analysis, involving the extraction of a frequency list of lexical items from each corpus and applying a log-likelihood statistic in order to see which words were more or less salient in each corpus.

The analysis found that the jpWaC contained much more informal and interactional material, covering a wider range of content; with the newspaper corpus being much more formal in content as well as style (Erjavec et al., 2008, p. 10). Similar findings were discovered by Sharoff (2006) when investigating the validity of WaC corpora of other languages. For example, in the case of the ukWaC corpus, it was compared with

both a newspaper corpus and the BNC, finding that ukWaC was more similar to the BNC than was the newspaper corpus (Erjavec et al., 2008; Sharoff, 2006). The jpWaC has further been subjected to validity tests in terms of the quality of the collocational data extracted in the word sketches produced from this corpus in the Sketch Engine software. Erjavec et al. (2008) discuss several limitations of the word sketches, such as issues with the flexibility of word order in Japanese (addressed by allowing gaps in the patterns between particles and the search word) and issues with the way in which a word is sometimes tokenised into smaller units (e.g. 女の人 onna no hito 'woman', literally 'girl person' is broken into three separate tokens). Furthermore, they found that part-of-speech tagging errors could sometimes lead to errors in the word sketch output, but in their evaluation of the quality of six word sketches made from the jpWaC corpus, including over 1000 collocations, they found only seven errors of this kind (Erjavec et al., 2008, p. 15).

4.4.3 *Extracting wordlists from Japanese corpora*

The first step in creating a list of frequently used English loanwords in Japanese was to extract the most frequent lexical items from each of the three corpora described above. Three decisions were taken at this stage: (1) deciding whether to extract words or lemmas, (2) deciding whether to extract Short Unit Words (SUW) or Long Unit Words (LUW), and (3) deciding on how many words of the corpus to use for the lists. Theoretically, a list of word forms was considered more appropriate than a list of lemmas because a lemmatised list may lose some differences of the individual loanword forms. As the process of lemmatisation "is a type of annotation that reduces the inflectional variants of words to their respective lexemes (or lemmas) as they appear in dictionary entries" (McEnery et al., 2006, p. 35), some distinctions between loanword forms may be lost if the lemma was selected instead of the word form. For example, lemmatisation would group *menzu* (メンズ 'men's') and *man* (マン 'man') under the lemma of マン 'MAN', obscuring differences in frequency and usage of these individual forms in the corpora.

Indeed, whilst lemmatisation is a convenient method of grouping words (Hunston, 2002, p. 18), when the purpose is to analyse the number and behaviour of individual words it is more appropriate to examine each word form separately. However, for practical reasons it was necessary to extract lemma lists rather than word-form lists. The

wordlist for the BCCWJ corpus was already publicly available and was a list of lemmas rather than word-forms. For the jpWaC and jpTenTen11 corpora, wordlists were not publicly available and so needed to be extracted from the corpora for the present study. In order to maintain consistency between the lists, it was decided to make these lists of lemmas, the same as the BCCWJ list. In practice, a manual analysis of the loanwords on the list found that the lemmatisation of the lists did not actually result in any instances of separate loanword word forms being grouped together under a lemma, except for the example of マン 'man' given above. Furthermore, this loanword ended up being excluded from the final list for unrelated reasons that are explained in Section 4.2.7.

The distinction between short unit words (SUW) and long unit words (LUW) is a very important one in Japanese corpus linguistics. This is due to the fact that words in written Japanese are not separated by white space, meaning that how a sentence is broken down into its constituent parts is open to two main interpretations: a minimal one and a maximal one. In the minimal interpretation, SUWs refer to the smallest individual units of meaning. For example, in a SUW analysis the compound 自動車 *jidousha* 'automobile' would be broken into the two parts of (1) 自動 *jidou* 'automatic', and (2) 車 *sha* 'vehicle'. In the maximal interpretation, a LUW analysis focuses on phrases, and therefore the two parts would be kept together. Generally, the SUW analysis is the one most preferred in Japanese linguistics and this is the principal method of analysis by which the three corpora used for the word lists were constructed (although each one does have a small-sample LUW version available).

The decision to select the most frequent 10,000 word-forms from each corpus was largely a heuristic to try to ensure a large enough sample of the Japanese language from which to extract a significant number of loanwords. As the aim of this research is to investigate the grammatical distribution of the loanwords, a large sample of these words was needed in order to allow the best chance for well-defined patterns to emerge from the data. Several pilot studies conducted for this research, as well as other studies in the academic literature (Daulton, 2008; Irwin, 2011; Kay, 1995; Loveday, 1996; Stanlaw, 2004), have shown that around 7-10% of the Japanese language is made up of loanwords, with over 90% of these being from English. As such, a sample of 10,000 words of Japanese could be expected to include around 700 or more loanwords, which was considered a sufficiently large enough sample for the later investigations. How the loanwords were distributed throughout each of these samples will be considered below.

For the BCCWJ, wordlists were already available as free, downloadable text files from the Center of Corpus Development on the National Institute for Japanese

Language and Linguistics (NINJAL) homepage. The wordlists were available, in SUW and LUW forms, for the entire vocabulary within the corpus, for specialised sub corpora, or by parts of speech. The SUW list of the entire corpus vocabulary was downloaded and the first 10,000 words from the text file were copied into a spreadsheet. NINJAL had taken the decision to rank words with the same frequency at the same shared rank on the list. For example, five words were grouped together at rank 4477 because they each had a frequency of 1,699 (16.3 per million). Rather than using this ranking to select the top 10,000 words from the BCCWJ, the individual word forms were separated out and 10,000 individual instances were given their own individual rank. This decision was taken in order to align the list with two others to be created from the jpWaC and jpTenTen11 corpora.

The BCCWJ wordlist had also been tagged by NINJAL for lexical information with the use of the freeware UniDic-mecab electronic dictionary and morphological analyser software (version 2.1.1). One of the features of the words tagged with information was the lexical stratum of each word. This tagging is an automated process and inevitably introduces some errors into the results. However, the wordlist had been cleaned by NINJAL before publication and in my analysis of the list only a few minor tagging errors still remained, such as the tagging of the three black squares of '■■■' as part of the *gairaigo* stratum (rank 9848 in the list of 10,000 words).[13]

For the jpWaC and jpTenTen11 corpora, there were no existing wordlists available. Therefore, a methodology was adopted in order to construct wordlists that would be equivalent to the one for the BCCWJ created by NINJAL. Both corpora are accessible through the online Sketch Engine corpus analysis software and can be analysed using tools such as the wordlist, word sketch, thesaurus, and sketch difference functions. Using the wordlist function, the most frequent 10,000 lemmas were extracted from each corpus as a text file and copied into a spreadsheet. Figure 4.14 shows a screenshot of the wordlist function of the Sketch Engine, with the search attribute set to 'lemma' (to match the BCCWJ list of lemmas). All other options were kept at the default settings. The lists were then analysed with the UniDic-mecab software (version 2.1.1) to tag words with their lexical stratum (native, Sino-Japanese, or loanword). This was the same software tool used by NINJAL to tag words in the BCCWJ wordlist so ensures consistency between how the words were tagged in each wordlist.

[13] This kind of error is commonly related to the font used to show Japanese texts in digital documents.

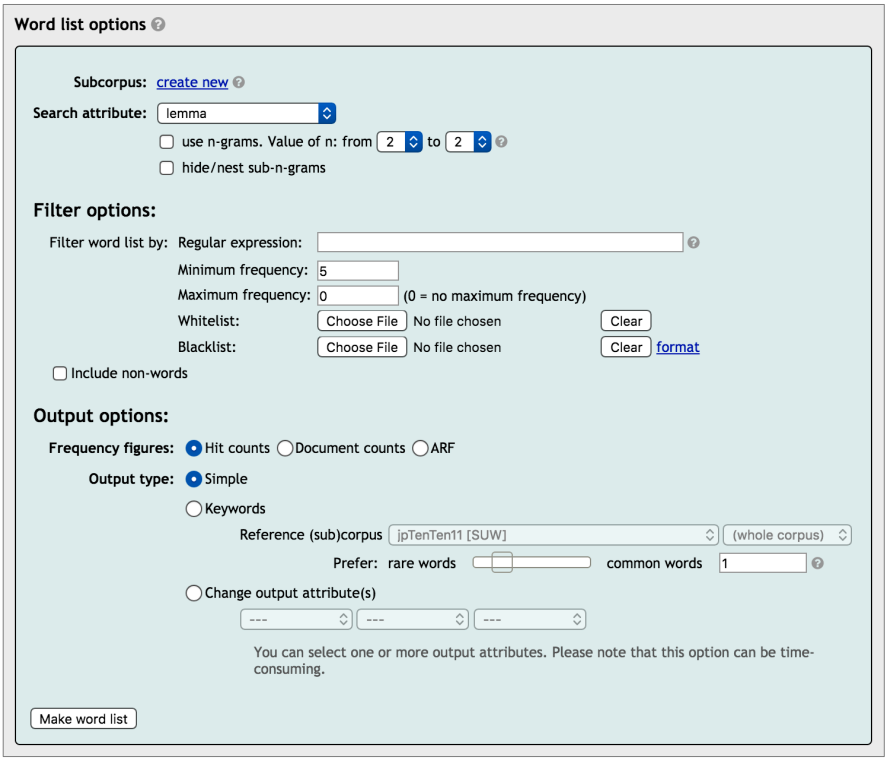


Figure 4.14 The wordlist function of the Sketch Engine with search attribute set to lemma.

4.4.4 *Combining the wordlists*

Tagging the lemmas in the jpWaC and jpTenTen11 corpora for their lexical strata with UniDic resulted in some errors, caused by the assignment of incorrect tags. As a result, each wordlist needed manual checking and cleaning in order to bring the number of items in each list to 10,000. For jpWaC there were 883 instances of where the UniDic software had problems with confidently tagging the words, and these are conveniently given a different code from the words which are assumed by the software to be analysed correctly. This means they can relatively-easily be filtered, manually analysed, and corrected. An initial check of the 883 potential errors revealed that the most common cause of tagging error was the software splitting words into further morphological parts. For example, the word として (*toshite* 'as, for, by way of') had been split into 3 separate morphemes と (*to*), し (*shi*), and て (*te*). Each of these had been incorrectly assigned a different rank in the list because they were analysed by the software as individual morphemes. Correcting this error involved changing the 3 separate words created by the UniDic analysis back into the original one word in the list extracted from

the corpus. Further correction was then needed to reanalyse the individual word and check that it had been assigned the correct lexical stratum tag. Of the 883 errors, 22 involved errors relating to *gairaigo* stratum words, and in only 1 instance did the tag need to be manually changed. This was for T シャツ (*Tshatsu* 't-shirt') which had been split into T and シャツ and then the 'T' had been incorrectly assigned a 'symbol' tag rather than *gairaigo*. The manual check of the UniDic analysis errors brought the list back to 10,000 individual items. The same process was followed for the jpTenTen11 wordlist where the UniDic analysis resulted in 213 potential errors. These were filtered, reanalysed, and corrected, leaving a final list of 10,000 lemmas.

The next stage in the process involved sorting and filtering the lists for the words tagged as being part of the *gairaigo* lexical stratum. First, the lists were sorted in a spreadsheet by the various lexical strata tags which had been automatically assigned by the UniDic-mecab software. Then, the lists were filtered for only the words which had been assigned a *gairaigo* tag. Filtering by *gairaigo* tag produced a final list of 1035 *gairaigo* lemmas for the BCCWJ wordlist, 1630 for jpTenTen11, and 1092 for jpWaC. Table 4.3 summarises these results.

Table 4.3 The number of loanwords extracted from the most frequent 10,000 lemmas of each corpus.

Corpus	**Total lemmas**	***Gairaigo***	
		no. lemmas	*%*
BCCWJ	10,000	1035	10.35%
jpWaC	10,000	1630	16.30%
jpTenTen11	10,000	1092	10.92%

The final goal was to produce a single list of English loanwords, and this was achieved by combining the three lists and retaining only those loanwords which appeared in all three corpora. This increased the likelihood that each loanword was indeed frequently used in the contemporary Japanese language, because each loanword needed to appear in three different corpora of the present-day Japanese language. The process of creating this single list is described below.

The first step in combining the three loanword lists was to compare the lemmas in each one. Firstly, the BCCWJ and jpTenTen11 lists were compared. This was done with a MATCH formula in a spreadsheet that searched for members of one list within the members of the other. The lists were compared in both directions to ensure that all loanwords on both lists could be included in the comparison. This produced 960 lemma

matches when the jpTenTen11 loanwords were compared to the BCCWJ, but produced 953 matches when the BCCWJ loanwords were compared to the jpTenTen11. It therefore revealed that there was an issue with seven of the loanwords on the BCCWJ list. On further analysis, it was found that the BCCWJ list had separated out seven instances of loanword homographs.

Because of the restricted phonology of the Japanese language compared with English, some words which are phonologically and graphologically distinguished in English (such as 'link' and 'rink') are combined into a single *gairaigo* form (リンク *rinku* 'link' or 'rink'). These homograph instances had been separated out in the BCCWJ wordlist, but this process was not part of the creation of the jpTenTen11 and jpWaC lists. Therefore, for the sake of convenience, the decision was taken to remerge the separated-out loanword homographs in the BCCWJ list and assign them a new rank based on their combined frequencies. These homographs were eventually excluded from the final list because of issues with how they could be analysed in the software. A further decision was taken at this point to remove three loanwords which had been tagged as *gairaigo*, but which are more typically written with *kanji* and/or *hiragana* characters than *katakana* (天ぷら *tenpura* 'tempura'; 旦那, *danna* 'husband/master'; 達磨 *daruma* 'dharma').

Remerging the loanword homographs and removing the three words written in *kanji* produced a final list of 950 loanwords appearing in both the BCCWJ and jpTenTen11 corpora. This list was then taken forward to compare with the jpWaC loanword list, following the same process described above. This produced a final list of 782 loanwords which appear in the top 10,000 lemmas of all three corpora. As these 782 loanwords appear at different ranks within each of the wordlists, their frequencies were combined and averaged in order to create a new, combined ranking of their frequency across all three corpora. No weighting was considered necessary in the combined ranking.

At this stage of the analysis a check was done on the distribution in each word list of the words which had been tagged as loanwords by the UniDic-mecab software. This was in order to build up a general picture of how loanwords are distributed in the Japanese lexicon in terms of whether they are mainly concentrated at a certain word-frequency level, such as being very frequent or very infrequent, or instead more evenly distributed throughout different frequency levels. Table 4.4 shows the latter to be the case, with the loanwords being distributed throughout the 10,000 words in all three lists. The combining of the lists into a single one, described above, led to the distribution data

not being further investigated in the analysis of the grammatical behaviour of the loanwords. Section 7.5, however, discusses how further research into the grammatical behaviour of loanwords could investigate a possible relationship between patterns of behaviour and frequency distribution in the lexicon.

Table 4.4 The frequency distribution of the 782 loanwords in each corpus.

Frequency Rank	jpTenTen11	*BCCWJ*	*jpWaC*
1-999	67	17	31
1000-1999	126	62	70
2000-2999	166	92	92
3000-3999	126	112	104
4000-4999	132	100	108
5000-5999	75	105	96
6000-6999	43	92	99
7000-7999	28	87	88
8000-8999	12	60	46
9000-10000	7	55	48
Total	**782**	**782**	**782**

As discussed in Chapter Two, not all loanwords in Japanese are borrowed from the English language. The UniDic software only tags words as part of the general *gairaigo* (loanword) lexical stratum and does not include specific information of which language the loanword has been borrowed from. No software could be found which would automatically tag each loanword with etymological information, and therefore this process was conducted manually by looking up each of the 782 loanwords in the Daijirin dictionary (3rd ed.) (2006), and the Concise Dictionary of Katakana words (4th ed.) (2010). These dictionaries were accessed through the online service Web Dictionary from the dictionaries' publisher Sanseido, as it allows the looking-up of words in multiple dictionaries at one time. The etymological information for each loanword was checked and the loanwords were manually tagged for the language from which they had been borrowed. This process resulted in a list of 748 loanwords, meaning that 95.7% of the initial 782 loanwords could be classified as English loanwords.

4.4.5 *Further filtering the list of English loanwords*

The final stage in the creation of the list was to remove English loanwords which would be potentially problematic for analysis within the Sketch Engine. The most likely potential cause of problems was considered to be the numerous homophonic and homographic English loanwords. This issue of homophonic and homographic English loanwords in Japanese is a complex one, because of the multiple variations involved. The issue is discussed in detail in Barrs (2015) and the discussion below is a brief overview of the kinds of issues which homophonic and homographic loanwords raise in regards to their corpus analysis.

Homographs are problematic in corpus analyses of language because of the difficulty in automatically disambiguating the different senses of words with different meanings but the same spellings (Kilgarriff, 2005). Explaining the problem with a basic example from English: how is the computer to know that two or more words with the same spelling (e.g. a money 'bank' and a river 'bank') are actually different words? This is a particular problem with a corpus-derived list of English loanwords in Japanese because homographs are numerous. First of all, some words which are homophonic homographs in English (i.e. words with the same sound and spelling but different meaning, such as 'match' related to fire and 'match' related to a game) are also homophonic homographs when borrowed into Japanese (i.e. マッチ *macchi* is used for both of the English words).

Furthermore, because of the limited phonemic inventory of Japanese compared to English, as was discussed in Section 2.4.1, homophonic hetereographs (same sound, different spelling) can become homophonic homographs (same sound and spelling) in their loanword form. For example, 'route' and 'root' become ルート *ruuto*. Complicating the situation even further, even hetereophonic hetereographs (i.e. different words in sound and spelling) can become homophonic homographs in Japanese. Using the previous example, not only 'route' and 'root' but also 'loot' takes the loanword form of ルート *ruuto*. In extreme cases, three or more phonologically and orthographically distinct words in English get represented as a single loanword in Japanese, as is the case with the English words 'lunch', 'launch', and 'ranch' all being represented as the English loanword ランチ *ranchi*. Because of these issues, the list of

748 English loanwords was manually analysed and loanwords with a single form for several different meanings were removed.[14]

Another type of loanword excluded from the final list were those which consisted of only one or two morae, which is a light form of a syllable and the more regular terminology used in Japanese phonology (Irwin, 2011). The reason for taking this decision was that a manual analysis of some of the word sketches of the short loanwords showed that the software sometimes was unsure which word was represented by the loanword. This is a particular problem in the computer-based analysis of languages which do not have white space between the words. For these languages, words can often be tokenized incorrectly where a loanword is split into smaller parts. For example, プロ is the loanword representation of the English word 'pro', as in 'professional', but is also used in Japanese as a shortening of other loanwords such as プログラム *puroguramu* 'program' and プロスティテュート *purosutityuuto* 'prostitute'. Another example is マン *man* 'man' which was analysed as both an independent loanword meaning 'man', and also a dependent part of other loanwords such as ウーマン *uuman* 'woman'. For this reason, this kind of short one or two morae loanwords were excluded from the list (other examples being ペン *pen*, ロー *roo*, and わん *wan*). Taking this decision to only include medium to long-length loanwords on the final list was not thought, however, to completely solve the problem of how accurately the software tokenises the loanwords. Indeed, the tokenisation of Japanese in general is a major issue in Japanese corpus linguistics (Tono et al., 2013). However, it was felt that these short loanwords had a particularly high probability of their word sketches containing a large number of errors, so removing them was considered appropriate.

A manual analysis of all 748 English loanwords identified 161 of them as potentially problematic for analysis in the word sketch function of the Sketch Engine. Consequently, they were removed to leave a final list of 587 frequent loanwords in the written Japanese language. This final list of 587 English loanwords is given in Appendix 2. To summarise the contents of this list, it was created with two guiding principles in mind: (1) to gather a sample of loanwords with a high frequency in the language so that a large amount of data of their linguistic contexts could be processed in their word sketches within the Sketch Engine, and (2) to try as much as possible to limit

[14] A precise number of how many homographs were removed from the list is not given because the exclusion of a loanword was often made for multiple reasons.

errors and ambiguities in the word sketches by removing loanwords which the Sketch Engine software would likely have difficulty in accurately analysing.

4.5 Reading the Tables of Gramrels

Table 4.5 is an excerpt of the data in Table 5.2 discussed in the next chapter. It is used here to summarise the preceding discussion and explain how the data in the tables shown in the next chapter should be read. The far-left column gives the rank of the gramrel types which are listed in the second column. As was described in Section 4.3.4, the top-10 most preferred gramrels shown in the word sketch of each of the 587 loanwords and the 130 non-loanwords were recorded and counted. These top-10 gramrels are indicated by the columns labelled '1st' to '10th', read horizontally across each table. The figures that are plotted in the main data area of the table indicate the number of times that a word sketch showed a gramrel appearing as one of the top-10 most preferred gramrels (i.e. gramrel tokens), and the numbers in parentheses show this data as a percentage of the total number of words in each sample (587 loanwords and 130 non-loanwords). In the next chapter, the term 'gramrel tokens' is used for the number of times that a word occurred in a specific preferred gramrel. As an example of how to read the data for each gramrel, the [particle] gramrel which is ranked first in Table 4.5 was the first most preferred gramrel for 358 (61%) of the 587 loanwords, the second most preferred gramrel for 189 (32.2%) of the loanwords, and so on across the table. The column on the far-right labelled 'Total' gives the number of tokens for that gramrel across all of the top-10 most preferred gramrels in the word sketches.

Table 4.5 An excerpt of the data from Table 5.2 showing the top-10 gramrel types found in the loanword database.

	Gramrel	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	TOTAL
1	particle	358 (61)	189 (32.2)	17 (2.9)	3 (0.6)	2 (0.4)						569 (97)
2	noun/noun	212 (36.2)	274 (46.7)	51 (8.7)	17 (2.9)	10 (1.8)	4 (0.7)	3 (0.6)	4 (0.7)	1 (0.2)	1 (0.2)	577 (98.3)
3	N_Adj	7 (1.2)	15 (2.6)	15 (2.6)	7 (1.2)	4 (0.7)	8 (1.4)	10 (1.8)	8 (1.4)	15 (2.6)	17 (2.9)	106 (18.1)
4	na_modifies_N	4 (0.7)	9 (1.6)	4 (0.7)	4 (0.7)	2 (0.4)	1 (0.2)		1 (0.2)		1 (0.2)	26 (4.5)
5	distant_Adv	2 (0.4)	11 (1.9)	59 (10.1)	25 (4.3)	8 (1.4)		1 (0.2)				106 (18.1)
6	noun o	2 (0.4)	2 (0.4)	5 (0.9)	6 (1.1)	3 (0.6)	7 (1.2)	5 (0.9)	3 (0.6)	5 (0.9)	2 (0.4)	40 (6.9)
7	suffix	1 (0.2)	1 (0.2)	14 (2.4)	29 (5)	35 (6)	52 (8.9)	53 (9.1)	80 (13.7)	72 (12.3)	47 (8.1)	384 (65.5)
8	Ana_ni modifies_V	1 (0.2)	1 (0.2)	1 (0.2)	4 (0.7)	3 (0.6)	2 (0.4)	2 (0.4)	3 (0.6)		3 (0.6)	20 (3.5)
9	o verb		58 (9.9)	173 (29.5)	109 (18.6)	63 (10.8)	49 (8.4)	46 (7.9)	18 (3.1)	10 (1.8)	4 (0.7)	530 (90.3)
10	pronom no		13 (2.3)	85 (14.5)	153 (26.1)	158 (27)	79 (13.5)	47 (8.1)	9 (1.6)	7 (1.2)	3 (0.6)	554 (94.4)

4.6 Chapter Summary

This chapter has given details of the corpora, corpus tools, and word lists used in the investigation of patterns of distribution in the grammatical relationships of frequent English loanwords in Japanese. It was explained how three main language resources are needed for the research: (1) a list of frequently used English loanwords, (2) a large collection of the loanwords' naturally-occurring linguistic contexts, and (3) a comparative sample of frequently used words from other Japanese lexical strata along with their contexts. These resources were discussed with reference to critical issues and limitations in the application of a corpus linguistics methodology, including the representativeness of corpora and the precision and recall of automated summaries of corpus-based data. It was then explained how the examination of the grammatical distribution of the loanwords requires the production of word sketches for each of the 587 English loanwords and the 130 native and Sino-Japanese words, and details were given of how the grammatical relationships in each word sketch were recorded, analysed, and combined into a database of over 5870 grammatical relationships for the English loanwords, and over 1300 for the non-loanwords. Details were also given of the creation of the list of 587 frequently-used English loanwords. The next chapter presents the results of the analysis of the word sketches of the 587 English loanwords and 130 non-loanwords.

5 Findings and Discussion 1: Patterns of Distribution

5.1 Chapter Overview

This chapter compares the grammatical relationships of the 587 English loanwords with those of the 130 Sino-Japanese and native words. Its purpose is for the grammatical behaviour of the Sino-Japanese and native words (hereafter, non-loanwords) to act as baseline comparative data for the analysis of the grammatical behaviour of the English loanwords. It was explained in Chapter Four that a set of word sketches was produced in the Sketch Engine software for the 587 loanwords and the 130 non-loanwords and the top-10 most preferred grammatical relationships (hereafter, gramrels) in each word sketch were recorded and turned into a database of 5870 gramrels for the loanwords and 1300 gramrels for the non-loanwords (see Section 4.3.4 for details of the databases). The following sections of this chapter analyse and compare the two databases in order to investigate patterns of distribution in the grammatical relationships of the English loanwords which are similar to and different from the patterns of distribution in the non-loanwords. The first analysis compares the total set of gramrels in each database. The second analysis focuses specifically on the gramrel types which appear as the first, second, and third most preferred gramrels in the loanword and non-loanword databases (i.e. the first three gramrel columns shown in a word sketch).[15] The last analysis focuses on a marked pattern of grammatical behaviour seen in the loanword database.

5.2 General Summary of the Loanword and Non-loanword Gramrel Databases

The first analysis to be conducted on the two databases is to compare the entire set of loanword gramrels with the entire set of non-loanword gramrels. In this section, details of the 1300 non-loanword gramrels are presented first, as this is baseline comparative data to see where the behaviour of the set of grammatical relationships of the 587 English loanwords is similar to and where it is different from that of the non-loanwords. Also, in this section the precise grammatical relationship which is represented by each gramrel (e.g. subject of the sentence, suffix, etc.) is not explained. The focus is on

[15] See Section 4.3.4 as well as the discussion below for an explanation of this terminology.

giving an overview of the main similarities and differences in the two data sets. Section 5.4 is where the explanations of the gramrels will be given.

Table 5.1 presents the whole set of 1300 gramrels of the 130 frequently-used non-loanwords. The table was created by analysing the database of 1300 gramrels, described in Section 4.3.4 above, to determine how many distinct gramrels were found in the 130 word sketches (i.e. gramrel types) and how many times in total each of these gramrels occurred in the set of 130 word sketches (i.e. gramrel tokens). The theoretical maximum number of times any one gramrel can appear in the database is 130, because a gramrel can only occur once in a single word sketch. It can be seen in the far-left column of the table that 61 out of a possible 157 gramrel types (38.9%) were observed (see Section 4.3.3 for details of all 157 gramrels). These 61 types are ordered in the table by the number of tokens of each type in the database, in high to low order. The number of tokens of each gramrel is shown in the far-right column labelled 'Total' and is given in parentheses as a percentage of the whole set of 130 non-loanwords. As an example, the [particle] gramrel is ranked first because there are 122 tokens of a possible 130 (93.9%) of this gramrel, meaning that only eight of the 130 non-loanwords (6.1%) did not have [particle] listed as one of its top-10 most preferred gramrels' The term 'most preferred gramrel' was introduced in Section 4.3.4 and because of the importance of this term in helping to explain the results, it will be explained in more detail here.

The term 'most preferred gramrels' is used in this study to refer to how the gramrel columns are automatically rank-ordered in the Sketch Engine word sketches. A word sketch is a table of a word's collocates distributed into columns based on a pre-defined set of grammatical relationships that link a word with its collocates. The algorithms used in the Sketch Engine to create Japanese word sketches present the columns on screen in frequency order of how often a word links with its collocates in one of 157 pre-defined grammatical relationships (which the Sketch Engine calls gramrels). Therefore, the first column presented in a word sketch (viewing the word sketch from left to right) shows the most frequent way in which the word is grammatically linked with its collocates from amongst all of its occurrences in the corpus. This can be thought of as the grammatical relationship which is 'most preferred' by the word, and as such, in the present study this first column in a word sketch is described as the 'first most preferred gramrel' of the word. Accordingly, the phrase 'top-10 most preferred gramrels' refers to the first 10 columns in a word sketch.

Table 5.1 The 61 gramrel types observed in the 130 non-loanword word sketches ordered by number of tokens (percentage of all 130 non-loanwords in parentheses).

	Gramrel	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	TOTAL
1	particle	106 (81.6)	12 (9.3)	1 (0.8)	1 (0.8)			1 (0.8)		1 (0.8)		122 (93.9)
2	pronom no		11 (8.5)	18 (13.9)	21 (16.2)	22 (17)	17 (13.1)	10 (7.7)	4 (3.1)	3 (2.4)	6 (4.7)	112 (86.2)
3	ni verb		10 (7.7)	14 (10.8)	13 (10)	17 (13.1)	14 (10.8)	20 (15.4)	14 (10.8)	2 (1.6)	2 (1.6)	106 (81.6)
4	noun/noun	8 (6.2)	19 (14.7)	11 (8.5)	18 (13.9)	9 (7)	15 (11.6)	5 (3.9)	10 (7.7)	5 (3.9)	5 (3.9)	105 (80.8)
5	no pronom		14 (10.8)	25 (19.3)	17 (13.1)	11 (8.5)	9 (7)	9 (7)	8 (6.2)	6 (4.7)	4 (3.1)	103 (79.3)
6	o verb		21 (16.2)	14 (10.8)	6 (4.7)	9 (7)	7 (5.4)	7 (5.4)	7 (5.4)	15 (11.6)	6 (4.7)	92 (70.8)
7	ga verb		1 (0.8)	1 (0.8)	10 (7.7)	7 (5.4)	14 (10.8)	12 (9.3)	9 (7)	13 (10)	8 (6.2)	75 (57.7)
8	de verb		3 (2.4)	5 (3.9)	4 (3.1)	8 (6.2)	5 (3.9)	9 (7)	13 (10)	10 (7.7)	6 (4.7)	63 (48.5)
9	wa verb	1 (0.8)	2 (1.6)	2 (1.6)	7 (5.4)	6 (4.7)	7 (5.4)	7 (5.4)	4 (3.1)	9 (7)	16 (12.4)	61 (47)
10	Adn			5 (3.9)	5 (3.9)	7 (5.4)	9 (7)	11 (8.5)	6 (4.7)	6 (4.7)	9 (7)	58 (44.7)
11	suffix	1 (0.8)	3 (2.4)	6 (4.7)	6 (4.7)	6 (4.7)	4 (3.1)	2 (1.6)	5 (3.9)	5 (3.9)	4 (3.1)	42 (32.4)
12	X					6 (4.7)	6 (4.7)	6 (4.7)	6 (4.7)	6 (4.7)	6 (4.7)	36 (27.7)
13	to verb			1 (0.8)	1 (0.8)		1 (0.8)	4 (3.1)	7 (5.4)	7 (5.4)	14 (10.8)	35 (27)
14	suffix_base	7 (5.4)	11 (8.5)	4 (3.1)	1 (0.8)	1 (0.8)	1 (0.8)	1 (0.8)		2 (1.6)		28 (21.6)
15	prefix	1 (0.8)	7 (5.4)	4 (3.1)		1 (0.8)	3 (2.4)	3 (2.4)	3 (2.4)	3 (2.4)	2 (1.6)	27 (20.8)
16	modifier_Ana		1 (0.8)		2 (1.6)	1 (0.8)	2 (1.6)	3 (2.4)	1 (0.8)	6 (4.7)	7 (5.4)	23 (17.7)
17	N_Adj		2 (1.6)			2 (1.6)	4 (3.1)	3 (2.4)	5 (3.9)	2 (1.6)	2 (1.6)	20 (15.4)
18	kara verb			1 (0.8)	2 (1.6)	1 (0.8)	2 (1.6)	1 (0.8)	4 (3.1)	4 (3.1)	2 (1.6)	17 (13.1)
19	ni verbsuru				2 (1.6)	1 (0.8)		5 (3.9)	3 (2.4)	3 (2.4)	3 (2.4)	17 (13.1)
20	modifier_Ai					3 (2.4)	1 (0.8)		3 (2.4)	3 (2.4)	6 (4.7)	16 (12.4)
21	distant_Adv	1 (0.8)	7 (5.4)	5 (3.9)	2 (1.6)		1 (0.8)					15 (11.6)
22	coord		1 (0.8)	6 (4.7)	1 (0.8)	2 (1.6)			1 (0.8)	1 (0.8)	2 (1.6)	14 (10.8)
23	wa verbsuru					2 (1.6)	1 (0.8)		3 (2.4)		2 (1.6)	8 (6.2)
24	made verb			1 (0.8)	2 (1.6)				1 (0.8)	3 (2.4)		7 (5.4)
25	noun o			1 (0.8)		2 (1.6)		1 (0.8)	1 (0.8)	1 (0.8)	1 (0.8)	7 (5.4)
26	prefix_base		1 (0.8)			1 (0.8)		2 (1.6)			2 (1.6)	6 (4.7)
27	N-Ana_modifies_N						1 (0.8)	1 (0.8)		3 (2.4)	1 (0.8)	6 (4.7)
28	N ga no Ai				5 (3.9)							5 (3.9)
29	N-Ai_modifies_N							2 (1.6)	2 (1.6)	1 (0.8)		5 (3.9)
30	o verbsuru								1 (0.8)	2 (1.6)	2 (1.6)	5 (3.9)
31	distant_V	4 (3.1)										4 (3.1)
32	distant_N+suru		3 (2.4)	1 (0.8)								4 (3.1)
33	noun ga			1 (0.8)						3 (2.4)		4 (3.1)
34	V_te Aux				1 (0.8)	1 (0.8)	2 (1.6)					4 (3.1)
35	noun ni	1 (0.8)						1 (0.8)		1 (0.8)		3 (2.4)
36	modifies_V		1 (0.8)	2 (1.6)								3 (2.4)
37	modifies_N+suru				1 (0.8)	1 (0.8)			1 (0.8)			3 (2.4)
38	ga Adj_concl					2 (1.6)			1 (0.8)			3 (2.4)
39	modifier_Adv					1 (0.8)	1 (0.8)				1 (0.8)	3 (2.4)
40	Ana						2 (1.6)	1 (0.8)				3 (2.4)
41	modifier_N-Ai								2 (1.6)		1 (0.8)	3 (2.4)
42	na_modifies_N			1 (0.8)	1 (0.8)							2 (1.6)
43	noun de							2 (1.6)				2 (1.6)
44	1+Ai							1 (0.8)	1 (0.8)			2 (1.6)
45	noun wa								1 (0.8)		1 (0.8)	2 (1.6)
46	e verb										2 (1.6)	2 (1.6)
47	made verbsuru										2 (1.6)	2 (1.6)
48	noun e				1 (0.8)							1 (0.8)
49	Ana_ni modifies_V						1 (0.8)					1 (0.8)
50	2+Ai								1 (0.8)			1 (0.8)
51	de verbsuru								1 (0.8)			1 (0.8)
52	noun to								1 (0.8)			1 (0.8)
53	Ai+1									1 (0.8)		1 (0.8)
54	ga Ai+N									1 (0.8)		1 (0.8)
55	kara verbsuru									1 (0.8)		1 (0.8)
56	made no pronom									1 (0.8)		1 (0.8)
57	modifier_N-Ana										1 (0.8)	1 (0.8)
58	N ga no Ai+1										1 (0.8)	1 (0.8)
59	noun kara										1 (0.8)	1 (0.8)
60	pronom to no										1 (0.8)	1 (0.8)
61	wa Adj_concl										1 (0.8)	1 (0.8)

Continuing with the example of how the data is read in Table 5.1, at the bottom end of the table the gramrel [*wa* Adj_concl] is ranked 61st because it has the (shared) lowest number of tokens of just one (0.8%), meaning that 129 of the 130 loanwords (99.2%) did not have this gramrel appearing anywhere in the top-10 most preferred gramrels in their word sketches. This gramrel is ranked at the very bottom of the table because of the fact that it occurs as the tenth most preferred gramrel and also comes last in the a-z ordering which has been used to rank the gramrels that have a shared number of tokens at the same time as appearing in the same position within the top-10 most preferred gramrels. The columns of data in Table 5.1 labelled 1st to 10th show how each gramrel was distributed across the top-10 most preferred gramrels of the entire set of 130 non-loanwords. Taking the [particle] gramrel again, this was the first most preferred gramrel of 106 (81.6%) non-loanwords, the second most preferred gramrel of 12 non-loanwords (9.3%), and so on.

Moving to the description of the database of loanword grammatical relationships, Table 5.2 was created in the same way as Table 5.1, by analysing the database of 5870 English loanword gramrels and determining the number of gramrel types and tokens. In this table, the theoretical maximum number of times any one individual gramrel can appear is 587, again because a gramrel can only occur once in a single word sketch. The far-left column of Table 5.2 shows that 57 of a possible 157 gramrel types (36.3%) were found in the database. These 57 gramrel types are ordered in the same way as in Table 5.1, in a high-low order of the total number of gramrel tokens across all top-10 most preferred gramrels, shown in the far-right column and given in parentheses as a percentage of the whole set of 587 loanwords. As an example of reading the data in this table, [noun/noun] is ranked first, with 577 gramrel tokens (98.3%). This means that only 10 of the 587 loanwords (1.7%) did not have the [noun/noun] gramrel appearing anywhere in the first 10 gramrel columns of their word sketches. Ranked at the bottom of the list of gramrel types is [*toshite* verb*suru*], with a single gramrel token in the tenth most preferred gramrel position. Like in Table 5.1, an a-z ordering was used for the gramrels which share the same number of tokens at the same time as appearing in the same position within the top-10 most preferred gramrels. Using the same terminology as for the description of the non-loanwords above, it can be seen in Table 5.2 that [noun/noun] was the first most preferred gramrel of 212 (36.2%) loanwords, the second most preferred gramrel of 274 (46.7) loanwords, etc.

Table 5.2 The 57 gramrel types observed in the 587 English loanword word sketches ordered by number of tokens (percentage of all 587 English loanwords in parentheses).

	Gramrel	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	TOTAL
1	noun/noun	212 (36.2)	274 (46.7)	51 (8.7)	17 (2.9)	10 (1.8)	4 (0.7)	3 (0.6)	4 (0.7)	1 (0.2)	1 (0.2)	577 (98.3)
2	particle	358 (61)	189 (32.2)	17 (2.9)	3 (0.6)	2 (0.4)						569 (97)
3	pronom no		13 (2.3)	85 (14.5)	153 (26.1)	158 (27)	79 (13.5)	47 (8.1)	9 (1.6)	7 (1.2)	3 (0.6)	554 (94.4)
4	o verb		58 (9.9)	173 (29.5)	109 (18.6)	63 (10.8)	49 (8.4)	46 (7.9)	18 (3.1)	10 (1.8)	4 (0.7)	530 (90.3)
5	no pronom		5 (0.9)	109 (18.6)	123 (21)	114 (19.5)	80 (13.7)	40 (6.9)	29 (5)	10 (1.8)	14 (2.4)	524 (89.3)
6	ni verb		3 (0.6)	16 (2.8)	44 (7.5)	65 (11.1)	100 (17.1)	104 (17.8)	65 (11.1)	36 (6.2)	31 (5.3)	464 (79.1)
7	suffix	1 (0.2)	1 (0.2)	14 (2.4)	29 (5)	35 (6)	52 (8.9)	53 (9.1)	80 (13.7)	72 (12.3)	47 (8.1)	384 (65.5)
8	ga verb		1 (0.2)	3 (0.6)	10 (1.8)	24 (4.1)	45 (7.7)	45 (7.7)	80 (13.7)	73 (12.5)	71 (12.1)	352 (60)
9	de verb		3 (0.6)	19 (3.3)	21 (3.6)	33 (5.7)	48 (8.2)	58 (9.9)	47 (8.1)	49 (8.4)	29 (5)	307 (52.3)
10	to verb			3 (0.6)	6 (1.1)	8 (1.4)	23 (4)	33 (5.7)	47 (8.1)	34 (5.8)	55 (9.4)	209 (35.7)
11	o verbsuru				5 (0.9)	8 (1.4)	20 (3.5)	37 (6.4)	37 (6.4)	41 (7)	42 (7.2)	190 (32.4)
12	coord			1 (0.2)	2 (0.4)	6 (1.1)	22 (3.8)	28 (4.8)	33 (5.7)	38 (6.5)	54 (9.2)	184 (31.4)
13	N_Adj	7 (1.2)	15 (2.6)	15 (2.6)	7 (1.2)	4 (0.7)	8 (1.4)	10 (1.8)	8 (1.4)	15 (2.6)	17 (2.9)	106 (18.1)
14	distant_Adv	2 (0.4)	11 (1.9)	59 (10.1)	25 (4.3)	8 (1.4)		1 (0.2)				106 (18.1)
15	Adn				3 (0.6)	5 (0.9)		13 (2.3)	13 (2.3)	36 (6.2)	34 (5.8)	104 (17.8)
16	X			1 (0.2)	3 (0.6)	3 (0.6)	4 (0.7)	12 (2.1)	16 (2.8)	21 (3.6)	32 (5.5)	92 (15.7)
17	wa verb					1 (0.2)	2 (0.4)	7 (1.2)	14 (2.4)	39 (6.7)	29 (5)	92 (15.7)
18	modifier_Ana					7 (1.2)	4 (0.7)	8 (1.4)	23 (4)	22 (3.8)	20 (3.5)	84 (14.4)
19	prefix			3 (0.6)		4 (0.7)	4 (0.7)	4 (0.7)	13 (2.3)	9 (1.6)	10 (1.8)	47 (8.1)
20	noun o	2 (0.4)	2 (0.4)	5 (0.9)	6 (1.1)	3 (0.6)	7 (1.2)	5 (0.9)	3 (0.6)	5 (0.9)	2 (0.4)	40 (6.9)
21	de verbsuru						1 (0.2)	4 (0.7)		11 (1.9)	13 (2.3)	29 (5)
22	Adv		2 (0.4)	4 (0.7)	2 (0.4)	9 (1.6)	2 (0.4)	1 (0.2)		4 (0.7)	4 (0.7)	28 (4.8)
23	ni verbsuru					1 (0.2)	2 (0.4)	3 (0.6)	5 (0.9)	6 (1.1)	11 (1.9)	28 (4.8)
24	na_modifies_N	4 (0.7)	9 (1.6)	4 (0.7)	4 (0.7)	2 (0.4)	1 (0.2)		1 (0.2)		1 (0.2)	26 (4.5)
25	V_te Aux				4 (0.7)		5 (0.9)	5 (0.9)	2 (0.4)	5 (0.9)	3 (0.6)	24 (4.1)
26	noun ni			1 (0.2)	2 (0.4)	1 (0.2)	2 (0.4)	2 (0.4)	3 (0.6)	3 (0.6)	8 (1.4)	22 (3.8)
27	Ana_ni modifies_V	1 (0.2)	1 (0.2)	1 (0.2)	4 (0.7)	3 (0.6)	2 (0.4)	2 (0.4)	3 (0.6)		3 (0.6)	20 (3.5)
28	modifier_Ai					2 (0.4)	2 (0.4)		1 (0.2)	7 (1.2)	8 (1.4)	20 (3.5)
29	modifier_Adv				1 (0.2)		2 (0.4)	4 (0.7)	2 (0.4)	4 (0.7)	6 (1.1)	19 (3.3)
30	no_modifies_N			1 (0.2)	3 (0.6)	4 (0.7)	5 (0.9)	2 (0.4)	1 (0.2)	1 (0.2)	1 (0.2)	18 (3.1)
31	kara verb						1 (0.2)	1 (0.2)	6 (1.1)	5 (0.9)	4 (0.7)	17 (2.9)
32	noun de					2 (0.4)		2 (0.4)	3 (0.6)	5 (0.9)	2 (0.4)	14 (2.4)
33	ga verbsuru								1 (0.2)	3 (0.6)	8 (1.4)	12 (2.1)
34	Adj+V						6 (1.1)		2 (0.4)	2 (0.4)	1 (0.2)	11 (1.9)
35	toshite verb						1 (0.2)	1 (0.2)	5 (0.9)	1 (0.2)	2 (0.4)	10 (1.8)
36	noun ga						2 (0.4)	3 (0.6)	3 (0.6)		1 (0.2)	9 (1.6)
37	e verb									4 (0.7)	4 (0.7)	8 (1.4)
38	N-Ai_modifies_N								2 (0.4)	1 (0.2)	2 (0.4)	5 (0.9)
39	Ana_modifies_N+suru				1 (0.2)	1 (0.2)		1 (0.2)	1 (0.2)			4 (0.7)
40	Ana_ni modifies N+suru			2 (0.4)					1 (0.2)			3 (0.6)
41	to verbsuru									1 (0.2)	2 (0.4)	3 (0.6)
42	noun kara					1 (0.2)			1 (0.2)			2 (0.4)
43	pronom to no						1 (0.2)		1 (0.2)			2 (0.4)
44	promom e no							2 (0.4)				2 (0.4)
45	noun wa								1 (0.2)	1 (0.2)		2 (0.4)
46	modifier_Ano								1 (0.2)		1 (0.2)	2 (0.4)
47	noun to								1 (0.2)		1 (0.2)	2 (0.4)
48	de no pronom									1 (0.2)	1 (0.2)	2 (0.4)
49	gaAdj_cont									1 (0.2)	1 (0.2)	2 (0.4)
50	ga Adj_concl									2 (0.4)		2 (0.4)
51	modifier_N-Ai						1 (0.2)					1 (0.2)
52	pronom kara no								1 (0.2)			1 (0.2)
53	made verb									1 (0.2)		1 (0.2)
54	Ai+1										1 (0.2)	1 (0.2)
55	modifier_Ai_ku										1 (0.2)	1 (0.2)
56	modifier_Ana_ni										1 (0.2)	1 (0.2)
57	toshite verbsuru										1 (0.2)	1 (0.2)

5.3 Pattern 1: Distribution of Gramrel Tokens

To help in uncovering patterns in the data, Tables 5.1 and 5.2 have been divided into three groups: a group of gramrels with a high number of tokens, a group with a medium number of tokens, and a group with a low number of tokens. The groupings are somewhat arbitrary, in that their boundaries are not statistically defined but rather drawn at lines that represent important general divisions in the data. In the qualitative approach adopted for this analysis of the gramrel databases (see Section 4.2 in the previous chapter), this drawing of lines in the tables of data is an example of what Richards (2003) refers to as "breaking down and recombining the data in an effort to build a picture that will respond to the aims of the research" (p. 270).

The first line is drawn at the point where the number of gramrel tokens (shown in the 'Total' column) drops below 50%. This is the point at which the gramrels start only being in the top-10 most preferred gramrels of half of the words in each database. In Table 5.1, this first line is drawn between [*ga* verb] and [*de* verb] and in Table 5.2 it is between [*de* verb] and [*to* verb]. The second line has been drawn where the number of gramrel tokens drops down to a single token. In Table 5.1 this is between [*made* verb*suru*] and [noun *e*] and in Table 5.2 it is between [*ga*Adj_Concl] and [modifier_N-Ai]. In Table 5.1 this creates a top group of 7 gramrels (11.5%) with a high number of tokens (i.e. the gramrels in this group were very common in the database), a middle group of 40 gramrels (65.6%) with a medium number of tokens, and a bottom group of 14 gramrels (22.9%) with a low number of tokens. In Table 5.2 the top group has 9 gramrels (15.78%), the middle group has 41 (71.92%), and the bottom group has 7 (12.28%).

Although the groupings are for the most part arbitrary, in both tables they match well with the broad changes in the colour-gradient conditional formatting which was applied to each of the database spreadsheets. This conditional formatting is based on a value range from the maximum number of words in the database (i.e. the gramrel appears in all 130 non-loanword or 587 loanword word sketches) down to a minimum of one (the gramrel appears in only one of the word sketches). The conditional formatting automatically colours the data cells with a dark red shading if the gramrels have a high number of tokens, a lighter red shading for the gramrels with a medium number of tokens, and white for the gramrels with a low number of tokens. In other words, the darker the red shading in a table cell, the higher the number of times the

gramrel was found appearing as the first, second, third (etc.) most preferred gramrel of the words.

The boundary where the line has been drawn in both tables making the group of gramrels with a high number of tokens matches well with the darkest area of red shading in the tables. This is also true for where the boundaries of the middle group are defined, with the shading going from light red down to white, and the bottom group where there is only white shading. Whilst these groupings and shadings are not directly comparable because of the different number of words in each database (130 non-loanwords and 587 loanwords), it is interesting to observe a similar general pattern in the sets of gramrels: that a small number of gramrel types appear to account for the majority of the gramrel tokens. This indicates that the most meaningful grammatical behaviour in both sets of data is concentrated in a small number of gramrel types.

This pattern can be confirmed in each table with a calculation of the total number of gramrel tokens in each of the three groups. For the non-loanwords in Table 5.1, the top group of seven gramrels account for 715 of the 1300 gramrel tokens (55%), the middle group of 40 gramrels account for 571 gramrel tokens (43.92%), and the bottom group of 14 gramrels account for 14 gramrel tokens (1.07%). For the loanword data in Table 5.2, the top group of nine gramrels account for 4261 of the 5870 gramrel tokens (72.58%), the middle group of 41 gramrels account for 1602 tokens (27.29%), and the bottom group of seven gramrels account for (0.11%). This data is shown in Table 5.3 with the top group in both data sets highlighted showing that this group accounts for the majority of gramrel tokens in both databases.

Table 5.3 The number of loanword and non-loanword gramrel tokens accounted for by the top, middle, and bottom groups of gramrel types.

Group	**Loanwords**			**Non-Loanwords**		
	Gramrel Types	Gramrel Tokens	% of 5870 Tokens	Gramrel Types	Gramrel Tokens	% of 1300 tokens
Top	**9**	**4261**	**72.58**	**7**	**715**	**55**
Middle	41	1602	27.29	40	571	43.92
Bottom	7	7	0.11	14	14	1.07
Total	*57*	*5870*	*100*	*61*	*1300*	*100*

To summarise this initial pattern, even though the 587 loanwords participate in 57 gramrel types and the 130 non-loanwords participate in 61 gramrel types, only a very small number of the gramrel types in each database have a high number of gramrel tokens. This section has given a general overview of the two databases, showing the

number of gramrel types, the total number of gramrel tokens of each type, and their distribution across the first 10 columns of the word sketches. It is the case, however, that the gramrels appearing in the first few columns of a word sketch are more meaningful in the present study than those appearing later. The next section discusses this issue in more detail and focuses the analysis on the top-3 most preferred gramrels.

5.4 Pattern 2: Dominance of Several Gramrel Types

It was explained in Chapter Four that the columns of gramrels appearing first in a word sketch will often be of more interest to the researcher than the columns of gramrels appearing later.[16] This is because these columns in the Japanese word sketches are organised in a high-low frequency order, meaning that the column which appears first is the grammatical relationship in which the search word most frequently appears in the corpus. As the present study is focused on understanding the grammatical behaviour of English loanwords in Japanese, these initial columns of the word sketches are of most interest because they show the most typical grammatical behaviour of the loanwords. The discussion above of the similar pattern of behaviour in the sample of non-loanwords and loanwords was focused on the overall number of tokens of each gramrel type. However, it is theoretically possible for any specific gramrel type to have a very high number of tokens, but for these the majority of these tokens to be occurring as one of the lesser preferred gramrels in the word sketches, for example the eighth, ninth, or tenth most preferred gramrel. These would not be as meaningful to an overall understanding of the behaviour of the sample of words as would the tokens occurring as the first, second, and third most preferred gramrel. Moreover, it can be seen from the data and colour-shading in Tables 5.1 and 5.2, that the frequency with which a search word occurs in a grammatical relationship drops rapidly even among the top-10 most preferred gramrels. For this reason, the discussion below focuses attention on the gramrels appearing in the first three gramrel columns in the word sketches.

The subsections below give a side-by-side comparison of the gramrel types which appear as the first, second, and third most preferred gramrels of the loanwords and non-loanwords. It could be argued that focusing solely on only the first gramrel column in the word sketches would be sufficient, because this represents the most frequent

[16] However, this depends on the research question being asked when examining the word sketches. Attention might be focused, for example, on exploring an uncommon grammatical relation.

grammatical behaviour of the loanwords. However, it was discussed in Section 4.3.4 that an observation of the frequencies of the first and second gramrel columns showed that in many cases there was very little difference in the numbers. This means that for these words, whilst a particular gramrel would be ranked in first place, in practice the overall frequency of the gramrel may differ little from the gramrel ranked in second place, making it more preferable to examine both gramrels as if they were sharing a frequency rank. From the same observation of the gramrel column frequencies, it was observed that there was often a considerable drop in frequency between the second and third gramrel columns. Taking these issues into consideration, the following discussion focuses specifically on the gramrels which appear in the first three gramrel columns in the word sketches, with the rest of the data on the gramrels in the other seven gramrel columns available in Tables 5.1 and 5.2 above and in a short summary in Section 5.4.2.

5.4.1 *The first most preferred gramrels*

Table 5.4 shows a side-by-side comparison of the first most preferred gramrel types and their number of tokens in the database of 130 non-loanwords and the database of 587 loanwords. Colours have been used to highlight a matching gramrel between the non-loanword and loanword data sets.[17] It can be noticed straight away that despite the differences in the number of words in each database, there is a very similar number of gramrel types appearing as the first most preferred gramrel, with nine gramrel types in the non-loanword database and eight in the loanword database. Furthermore, around half of these gramrels are the same in each data set. Out of the total of 61 gramrel types seen in the non-loanword database and the 57 gramrel types in the loanword database, this shows that in both data sets the words occur most frequently in a small number of these gramrel types, and furthermore that many of these types are the same between the two sets of words. For the non-loanwords, only 14.75% of all the 61 gramrels observed in the database appear as first most preferred, and the figure is 15.68% of the 57 gramrels for the loanwords. Overall, this is a similar finding to that of the first pattern discussed above in Section 5.3 which showed a small number of gramrel types accounting for a majority of gramrel tokens, with the difference being that the focus here is just on the first most preferred gramrel rather than all 10 gramrels combined.

[17] In Table 5.4 it just so happens that the gramrel matches occur in the same rows, but this is not a condition of the colour coding. This is shown in Table 5.5 where matches occur in different rows.

Table 5.4 The first most preferred gramrels of the non-loanwords and loanwords.

	Non-Loanwords		Loanwords	
	Gramrel Type	***Gramrel Token (%)***	***Gramrel Type***	***Gramrel Token (%)***
1	particle	106 (81.6)	particle	358 (61)
2	noun/noun	8 (6.2)	noun/noun	212 (36.1)
3	suffix_base	7 (5.4)	N_Adj	7 (1.2)
4	distant_V	4 (3.1)	na_modifies_N	4 (0.7)
5	distant_Adv	1 (0.8)	distant_Adv	2 (0.3)
6	noun ni	1 (0.8)	noun o	2 (0.3)
7	prefix	1 (0.8)	Ana_ni modifies_V	1 (0.2)
8	suffix	1 (0.8)	suffix	1 (0.2)
9	wa verb	1 (0.8)	-	-
	Total	*130 (100)*		*587 (100)*

Another interesting pattern in Table 5.4 is that the first two gramrel types with the highest number of tokens in both databases are the same: [particle] and [noun/noun]. This suggests that the grammatical behaviour of the two sets of words are similar. However, a point of particular interest is the high number of gramrels of [noun/noun] in the loanword database compared to the much lower number in the non-loanword database. At 212 tokens of the [noun/noun] gramrel for the loanwords, this is 36.1% of the maximum of 587 tokens, whilst for the non-loanwords, the number of tokens is only eight, representing only 6.2% of the maximum 130. A chi squared test was run on this difference to check for statistical significance, and the difference was found to be significant at $p < 0.05$. Furthermore, the percentage of gramrel tokens of [noun/noun] is not markedly different from the other first most preferred gramrels for the non-loanwords but is substantially different in the case of the loanwords. This much higher percentage for [noun/noun] in the loanword data compared with the non-loanword data warrants further exploration and will be taken up below in Section 5.5.

5.4.2 *The second most preferred gramrels*

Looking at data of the second most preferred gramrels in Table 5.5, it can be seen from the same colouring used in Table 5.4 to highlight the matching gramrels that many of the gramrel types are the same between the two sets of words. Whilst for the first most

preferred gramrels around half of the gramrel types matched, in this case the majority of the gramrels in both data sets match. For this second most preferred gramrel, however, there is a different ordering of the matching gramrels between each dataset. For example [*o* verb] is the gramrel with the highest number of tokens for the non-loanwords but has the third highest number for the loanwords. This is an indication that with this second most preferred gramrel, apart from the matching gramrel types between the non-loanwords and loanwords, similarities between the non-loanword and loanword gramrels begin to fade.

Table 5.5 The second most preferred gramrels of the non-loanwords and loanwords.

	Non-Loanwords		**Loanwords**	
	Gramrel Type	***Gramrel Token (%)***	***Gramrel Type***	***Gramrel Token (%)***
1	o verb	21 (16.2)	noun/noun	274 (46.7)
2	noun/noun	19 (14.7)	particle	189 (32.2)
3	no pronom	14 (10.8)	o verb	58 (9.9)
4	particle	12 (9.3)	N_Adj	15 (2.6)
5	pronom no	11 (8.5)	pronom no	13 (2.2)
6	suffix_base	11 (8.5)	distant_Adv	11 (1.9)
7	ni verb	10 (7.7)	na_modifies_N	9 (1.5)
8	distant_Adv	7 (5.4)	no pronom	5 (0.9)
9	prefix	7 (5.4)	de verb	3 (0.5)
10	de verb	3 (2.4)	ni verb	3 (0.5)
11	distant_N+suru	3 (2.4)	Adv	2 (0.3)
12	suffix	3 (2.4)	noun o	2 (0.3)
13	N_Adj	2 (1.6)	Ana_ni modifies_V	1 (0.2)
14	wa verb	2 (1.6)	ga verb	1 (0.2)
15	coord	1 (0.8)	suffix	1 (0.2)
16	ga verb	1 (0.8)	-	-
17	modifier_Ana	1 (0.8)	-	-
18	modifies_V	1 (0.8)	-	-
19	prefix_base	1 (0.8)	-	-
	Total	*130 (100)*		*587 (100)*

For the non-loanwords, no particular gramrel stands out as having a markedly higher number of tokens than another, with there instead being a small and steady decline in the number of tokens of each gramrel. There is only about a 15% overall

decrease in the number of tokens between [*o* verb] with the highest number of 21 (16.2%), and the five gramrels at the bottom of the table with only one token (0.8%). The data for the loanwords, however, is very different, with the behaviour seen above in Table 5.4 of a clear tendency for just two gramrels to have a particularly high rate of occurrence continuing here in the behaviour of the second most preferred gramrels. Furthermore, these two gramrels in Table 5.5 with the highest number of tokens are the same as the two gramrels in Table 5.4 with the highest number of tokens, just with their order switched. So, whilst [particle] is the first most preferred gramrel for the largest number of loanwords (Table 5.4) and [noun/noun] is in second place, [noun/noun] is the second most preferred gramrel for the largest number of loanwords (Table 5.5) and [particle] is in second place. This is further evidence that [noun/noun] seems to be a particularly dominant grammatical relationship in the behaviour of the loanwords, but not so dominant in the behaviour of the non-loanwords.

Looking at other aspects of the data in Table 5.5, whilst the non-loanword data displays a slow and steady decline in the number of tokens of each gramrel, there is instead a sharp drop in the number of tokens in the loanword data between the first two gramrels and the remaining 13. This is shown by the fact that in the non-loanword data, the largest difference between the number of tokens of each gramrel is around 4%, between [noun/noun] and [*no* pronom]. However, in the loanword data there is a 22% drop between [particle] and [*o* verb]. Moreover, whilst as just mentioned there is a 15% overall decrease in the number of tokens in the non-loanword data, it is around 46% for the loanwords. This could be one reason why the number of second most preferred gramrel types is lower for the loanwords than the non-loanwords, despite the fact that the sample of loanwords is considerably bigger than the non-loanwords (which would potentially allow more grammatical relationships the chance to appear). This provides even more evidence that the [particle] and [noun/noun] gramrels are dominant grammatical relationships of the loanwords.

5.4.3 *The third most preferred gramrels*

Table 5.6 shows comparative data of the third most preferred gramrel of the non-loanwords and loanwords, using the same colour-coding as the previous two tables to show matching gramrels. Interestingly, this data reveals more consistency between the behaviour of this third most preferred gramrel in the two datasets than with the previous two most preferred gramrels. Firstly, there are a similar number of gramrel types, which

is again despite the large difference in the number of words in each word-sample. Secondly, the three gramrels with the highest number of tokens in each database are the same, just in a different order. Most importantly, looking at the percentages of the gramrel tokens, there does not appear to be a gramrel in either database with a considerably higher number of tokens than any other. Looking back at Tables 5.1 and 5.2 of the data of all 10 most preferred gramrels, it can be seen that the pattern of one or two gramrels having a distinctly higher number of tokens than the others is replaced with a pattern of a large number of gramrel types having a medium or low number of tokens. Expressed in another way, the dominance of [particle] and [noun/noun] in the database is giving way to other gramrels rising to the top of the list, such as [*o* verb]. This is because, as Tables 5.4 and 5.5 have shown, the large majority of instances of the [particle] and [noun/noun] gramrels in the loanword database are when they occur as the first and second most preferred gramrels of the loanwords. Once again this shows the dominance of these two gramrels in the grammatical behaviour of the loanwords.

The data in Table 5.6 reveals a further pattern of the same set of five gramrels in each database consistently having a high number of tokens as the top-three most preferred gramrels. Two of these are [particle] and [noun/noun], and the other three are [*o* verb] [*no* pronom] and [pronom *no*]. In all three tables of data so far described here, these five gramrels have continually appeared at the top of the tables for both datasets, meaning that they are very frequent grammatical relationships of both the non-loanwords and the loanwords. For this reason, the type of grammatical relationship which each of these represents will be explained in more detail below. Most attention, however, will be given to [particle] and [noun/noun] because these are the two gramrels with the highest combined number of tokens as the first and second most preferred gramrels for both the non-loanwords and loanwords. Before focusing on these gramrels, the next subsection gives a short summary of the data of the fourth to tenth most preferred gramrels.

Table 5.6 The third most preferred gramrels of the non-loanwords and loanwords.

	Non-Loanwords		Loanwords	
	Gramrel Type	***Gramrel Token (%)***	***Gramrel Type***	***Gramrel Token (%)***
1	no pronom	25 (19.3)	o verb	173 (29.5)
2	pronom no	18 (13.9)	no pronom	109 (18.6)
3	o verb	14 (10.8)	pronom no	85 (14.5)
4	ni verb	14 (10.8)	distant_Adv	59 (10.1)
5	noun/noun	11 (8.5)	noun/noun	51 (8.7)
6	suffix	6 (4.7)	de verb	19 (3.2)
7	coord	6 (4.7)	particle	17 (2.9)
8	distant_Adv	5 (3.9)	ni verb	16 (2.7)
9	de verb	5 (3.9)	N_Adj	15 (2.6)
10	Adn	5 (3.9)	suffix	14 (2.4)
11	suffix_base	4 (3.1)	noun o	5 (0.9)
12	prefix	4 (3.1)	na_modifies_N	4 (0.7)
13	wa verb	2 (1.6)	Adv	4 (0.7)
14	modifies_V	2 (1.6)	ga verb	3 (0.5)
15	particle	1 (0.8)	prefix	3 (0.5)
16	distant_N+suru	1 (0.8)	to verb	3 (0.5)
17	ga verb	1 (0.8)	Ana_ni modifies N+suru	2 (0.3)
18	na_modifies_N	1 (0.8)	X	1 (0.2)
19	kara verb	1 (0.8)	Ana_ni modifies V	1 (0.2)
20	made verb	1 (0.8)	coord	1 (0.2)
21	noun ga	1 (0.8)	no_modifies_N	1 (0.2)
22	noun o	1 (0.8)	noun ni	1 (0.2)
23	to verb	1 (0.8)	-	-
	Total	*130 (100)*		*587 (100)*

5.4.4 *The fourth to tenth most preferred gramrels*

Looking back again at Tables 5.1 and 5.2 of the data of all the gramrels in the two databases, the stepped pattern in the data caused by where the data points appear in the table shows that there are a much larger number of tenth most preferred gramrels than first most preferred gramrels. For the non-loanword data in Table 5.1, there are nine first most preferred gramrel types, rising by 12 to make 21 second most preferred gramrels, then a gradual increase across the other positions to end up with 34 tenth most preferred gramrel types. Amongst these 34 gramrel types, the number of tokens of each one is relatively low, ranging from 16 with [*wa* verb] (12.4%) to just one of [*wa* Adj_concl] (0.8%). For the nine gramrel types in the first column, however, the range is from 106 tokens of [particle] (81.6%) down to one of [*wa* verb] (0.8%), which makes this difference 7-times larger. For the loanword data in Table 5.2, there is a similar stepped pattern. There are eight first most preferred gramrel types, increasing steadily to 45 tenth most preferred gramrel types, with each of these 45 gramrel types having a fairly low number of tokens. The number of tokens ranges from the one of [*toshite* verb*suru*] (0.2%) to 71 of [*ga* verb] (12.1%). Looking back at the first column, here the number of tokens of the eight gramrel types range from the 358 tokens of [particle] (61%) to the one token of [Ana_*ni* modifies_V] (0.2%). This is a 5-times bigger difference than with the tenth most preferred gramrels. This gives further empirical evidence of the non-loanwords and loanwords occurring most-frequently in just a few gramrel types. The next section will discuss and compare these gramrel types in detail.

5.5 Pattern 3: A Marked Grammatical Behaviour of the Loanwords

Summarising the above findings, it has been shown that both the non-loanwords and loanwords are most frequently found in a [particle] grammatical relationship in the jpTenTen11 corpus. Furthermore, it has been found that the [noun/noun] gramrel is also frequently observed in both databases, but is particularly prevalent in the loanword gramrel database where it accounts for very close to 10% of all the 5870 gramrel tokens in the database. In fact, the [noun/noun] gramrel appeared within the first 10 gramrel columns of 577 of the 587 English loanword word sketches, and 486 of these tokens (84%) were within the first two gramrel columns (i.e. first and second most preferred gramrels). So far in this chapter, the actual grammatical behaviour expressed by the

gramrels has not been explained, because of the focus on broadly comparing the overall data of gramrel types and gramrel tokens in the two databases and searching for patterns. In this Section, these gramrels are now explained to better understand the patterns of grammatical behaviour of the English loanwords.

In the Japanese sketch grammar used in the Sketch Engine to build the word sketches, the [particle] gramrel represents a behaviour significantly different from the other gramrels. It is defined as a relationship whereby the search word (which in the present study means any of the words in the non-loanword and loanword samples) is followed by one or more of over 50 different post-positional particles. Particles in Japanese are the most frequent words in the language and "mark the functions of elements immediately before them" (Cipris & Hamano, 2002, p. 23). They are commonly grouped into eight types: case markers, parallel markers, sentence-ending particles, interjectory particles, adverbial particles, binding particles, conjunctive particles, and phrasal particles (Hasegawa, 2015). They express a wide range of grammatical functions holding between the words/phrases which come before and those which come after the particle, such as: location + particle *de* + action; direction + particle *ni* + action; a rhetorical question + particle *ne*; an invitation + particle *ka*, a reason + particle *kara* or *node*, and assertiveness + particle *yo* (examples of some of these particles in the jpTenTen11 corpus will be given below).

The [particle] gramrel in the Japanese sketch grammar is unique within the set of 157 Japanese gramrels in that it is the only one which groups together into one large set other individual gramrels which all display a related type of grammatical behaviour. The same type of gramrel can also be found in the English sketch grammar used for word sketches made from the enTenTen13 corpus. In this case the gramrel is called [prepositional phrases] and groups together individual preposition gramrels such as 'in', 'of', 'with', 'for', and 'from' (Thomas, 2017, p. 182). To more clearly explain the behaviour of the [particle] gramrel in the Japanese word sketches, Figure 5.1 shows part of the word sketch for the English loanword *kamera* 'camera', with the names of the gramrels in the first four columns highlighted in a coloured box. It can be seen that this loanword occurs in a [particle] grammatical relationship more frequently than any of the other gramrels in its word sketch, shown by the fact that it appears in the first column and has the figure 56.82 at the top of the column meaning that almost 57% of its 833,163 occurrences in the corpus are found in this grammatical relationship (compared with the 22.37% of the [noun/noun] gramrel).

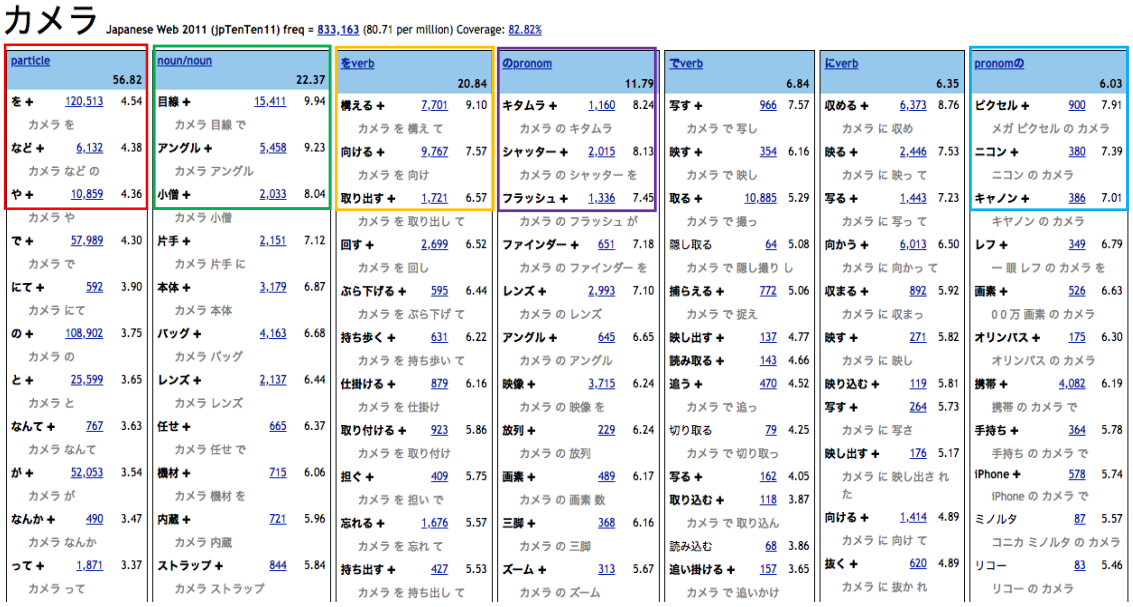


Figure 5.1 An excerpt of the word sketch for the loanword カメラ *kamera* 'camera' with the first four gramrel columns highlighted.

The top-three listed collocates of this loanword, highlighted within the red box, are カメラを *kamera o*; カメラなど *kamera nado*; and カメラや *kamera ya*. The *o* particle indicates that a transitive verb is 'being done' to the noun *kamera*. The first concordance line in the corpus of this collocational pair is:

カメラを用意 すれ ば 、 顔 を 見 ながら の 会話 も可能 です

which transliterates as *kamera o yoi sureba kao o minagara no kaiwa mo kanou desu* and translates as 'if a camera is prepared/set up then it is possible to have a conversation whilst also seeing the speaker's face'. The second collocational pair of *kamera* with *nado* indicates that *kamera* is marked as one example from a category, for example コンピューター、カメラなど *konpyuutaa, kamera nado* 'things like computers and cameras', and the third indicates a more basic listing of items, for example *dejitaru kamera ya geemu ki* 'digital cameras and game machines'.

Overall, rather than expressing one specific grammatical relationship, as is done by the other gramrels in the Japanese sketch grammar, the [particle] gramrel expresses that the search word is used in a variety of different grammatical patterns in the corpus. Taking the *kamera* example again, the fact that [particle] gramrel is by far its first most preferred gramrel reflects the wide variety of ways it is being used in the corpus, including most frequently as the object of a verb (when followed by the *o* particle), but also within lists of items (with the use of the *ya* particle), and as an example of an item

(with the *nado* particle). Other particles in the first column of the word sketch show it being used as the subject of a sentence (with the *ga* particle, for example *kamera ga hatsubai sareta* 'the camera was put on sale'), quoted within a phrase (with the *-tte* particle, for example *konpakuto kamera tte takusan aru to omou* 'speaking of digital cameras, I think there are a lot around), and as a method/means of doing something (with the *de* particle, for example *futsuu no kamera de* 'with a regular camera').

Because so many individual grammatical relationships are grouped within this single [particle] gramrel, it is highly likely that in the analysis of the gramrels of Japanese vocabulary [particle] will come out as the most-frequent gramrel of the majority of the words. Indeed, this is exactly what is shown in both the non-loanword and loanword gramrel databases. Tables 5.7 and 5.8 are extracts of the data in Tables 5.1 and 5.2 respectively, this time just showing the top groups of gramrels with the highest number of tokens and used here to draw attention to the [particle] gramrel in both databases. Looking at Table 5.7, the darkest-shaded red cell of data is for [particle] as the first most preferred gramrel. It is therefore very clearly the grammatical relationship in which the loanwords participate most often. However, as was just explained, it is not actually an individual grammatical relationship. Therefore, to get a fairer, more informative view on the data, it would be useful to remove this [gramrel] from the analysis. In doing so, and then examining the remaining data in the table, particularly in the top-two most preferred gramrel columns, the shading in the table does not immediately bring out any other data point as more noticeable than another.

Table 5.7 Extract of data from Table 5.1 showing the group of non-loanword gramrels with the highest number of tokens.

	Gramrel	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	TOTAL
1	particle	106 (81.6)	12 (9.3)	1 (0.8)	1 (0.8)			1 (0.8)		1 (0.8)		122 (93.9)
2	pronom no		11 (8.5)	18 (13.9)	21 (16.2)	22 (17)	17 (13.1)	10 (7.7)	4 (3.1)	3 (2.4)	6 (4.7)	112 (86.2)
3	ni verb		10 (7.7)	14 (10.8)	13 (10)	17 (13.1)	14 (10.8)	20 (15.4)	14 (10.8)	2 (1.6)	2 (1.6)	106 (81.6)
4	noun/noun	8 (6.2)	19 (14.7)	11 (8.5)	18 (13.9)	9 (7)	15 (11.6)	5 (3.9)	10 (7.7)	5 (3.9)	5 (3.9)	105 (80.8)
5	no pronom		14 (10.8)	25 (19.3)	17 (13.1)	11 (8.5)	9 (7)	9 (7)	8 (6.2)	6 (4.7)	4 (3.1)	103 (79.3)
6	o verb		21 (16.2)	14 (10.8)	6 (4.7)	9 (7)	7 (5.4)	7 (5.4)	7 (5.4)	15 (11.6)	6 (4.7)	92 (70.8)
7	ga verb		1 (0.8)	1 (0.8)	10 (7.7)	7 (5.4)	14 (10.8)	12 (9.3)	9 (7)	13 (10)	8 (6.2)	75 (57.7)

The high gramrel count of [particle] as actually a parent gramrel grouping together many individual particle-based gramrels would suggest that some of these individual particle-based gramrels should be appearing high up in the table. This is indeed the case, with Table 5.7 showing [*ni* verb] at rank 3, [*o* verb] at rank 6, and [*ga* verb] at rank 7. These are just 3 of many different particle-based gramrels that are subsumed

into the [particle] gramrel: [*ni* verb] represents the grammatical relationship of 'direction', for example *Tokyo ni ikimasu* 'go to Tokyo'; [*o* verb] expresses the word acting as the object of a verb phrase, such as *gohan o taberu* 'eat dinner'; and [*ga* verb] indicates the word is the subject of a sentence, such as *kamera ga takai* 'the camera is expensive'. With both the non-loanwords and loanwords, [*o* verb] is the most frequent particle-based gramrel and will be discussed in more detail below.

Table 5.8 shows that for the gramrels of the loanwords, if the [particle] gramrel is temporarily discounted from the analysis, it results in a situation which is markedly different from that just described for the non-loanwords. Again, in the loanword database, just as for the non-loanwords, [particle] has an extremely high number of tokens, being a gramrel which appears as the first or second most preferred gramrel of 547 of the 587 loanwords (93%). However, if this gramrel is excluded, two data points are still left behind in the top-two most preferred gramrel columns with a much darker shade of red than the others, signifying a much higher number of tokens than the other gramrels. These two data points are for the [noun/noun] gramrel, and account for 486 of its 587 gramrel tokens (83%). It can be said therefore, that the very high number of occurrences of the [noun/noun] particle in the loanword database expresses a dominant grammatical behaviour of the group of English loanwords in Japanese which is not observed as a dominant behaviour in the group of native and Sino-Japanese words.

Table 5.8 Extract of data from Table 5.2 showing the group of loanword gramrels with the highest number of tokens.

	Gramrel	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	TOTAL
1	noun/noun	212 (36.2)	274 (46.7)	51 (8.7)	17 (2.9)	10 (1.8)	4 (0.7)	3 (0.6)	4 (0.7)	1 (0.2)	1 (0.2)	577 (98.3)
2	particle	358 (61)	189 (32.2)	17 (2.9)	3 (0.6)	2 (0.4)						569 (97)
3	pronom no		13 (2.3)	85 (14.5)	153 (26.1)	158 (27)	79 (13.5)	47 (8.1)	9 (1.6)	7 (1.2)	3 (0.6)	554 (94.4)
4	o verb		58 (9.9)	173 (29.5)	109 (18.6)	63 (10.8)	49 (8.4)	46 (7.9)	18 (3.1)	10 (1.8)	4 (0.7)	530 (90.3)
5	no pronom		5 (0.9)	109 (18.6)	123 (21)	114 (19.5)	80 (13.7)	40 (6.9)	29 (5)	10 (1.8)	14 (2.4)	524 (89.3)
6	ni verb		3 (0.6)	16 (2.8)	44 (7.5)	65 (11.1)	100 (17.1)	104 (17.8)	65 (11.1)	36 (6.2)	31 (5.3)	464 (79.1)
7	suffix	1 (0.2)	1 (0.2)	14 (2.4)	29 (5)	35 (6)	52 (8.9)	53 (9.1)	80 (13.7)	72 (12.3)	47 (8.1)	384 (65.5)
8	ga verb		1 (0.2)	3 (0.6)	10 (1.8)	24 (4.1)	45 (7.7)	45 (7.7)	80 (13.7)	73 (12.5)	71 (12.1)	352 (60)
9	de verb		3 (0.6)	19 (3.3)	21 (3.6)	33 (5.7)	48 (8.2)	58 (9.9)	47 (8.1)	49 (8.4)	29 (5)	307 (52.3)

If for the moment both [particle] and [noun/noun] are removed from the analysis of both databases, it leaves three other gramrels which have a particularly high number of tokens and which are two of the more preferred gramrels for both the non-loanwords and loanwords. One of these is [*o* verb], which is related to the [particle] gramrel, and the other two are [*no* pronom] and [pronom *no*], which are connected to the [noun/noun] gramrel. As was introduced above, the [*o* verb] gramrel is one of the

members of the [particle] gramrel, as well as being a gramrel within its own right. Loanwords occurring in this relationship are nouns acted upon by the subsequent transitive verb. In Figure 5.1, the top-three collocates in this third most preferred gramrel (highlighted within the orange box) are *kamera o kamaeru* 'setting up/holding a camera', *kamera o mukeru* 'to face/turn towards a camera', and *kamera o toridasu* 'to take out a camera'. The list of collocates in this particular gramrel are especially informative for giving an overall sense of the meaning of the word under investigation. This is because in expressing the different actions which can be done to the word, the [*o* verb] gramrel helps disambiguate words which have various meanings. For example, whilst all instances of the loanword *kamera* in its word sketch are not necessarily referring to the actual technological object of a 'camera', because the word can also be used in contexts such as brand names, like the company BicCamera, the most salient verb collocates listed in the [*o* verb] gramrel (as ranked by the logDice statistic) help show that *kamera* typically does refer to the technological device, such as 'take out', 'hang', 'turn', 'forget', 'walk with', and 'hold'.

The [*no* pronom] gramrel indicates that the word being analysed modifies a following noun in the formation of a noun phrase. The difference between this gramrel and [noun/noun] is that the [*no* pronom] gramrel forms a more open-ended noun phrase whilst [noun/noun] forms a compound noun. The two nouns in the phrase can relate to each other in a variety of ways, such as noun A being the possessor of B (e.g. 'the cat's tail'), noun A being the location of noun B (e.g. 'the top of the table'), and noun B being about noun A (e.g. 'a book on/about sports'). This means that the form 'noun A *no* noun B' by itself is ambiguous and needs further context in order for the meaning to be correctly discerned. For example, a noun phrase such as パイロットの本 *pairotto no hon* could mean 'the pilot's book (which he/she is reading now)', 'the books the pilot owns', and 'a book about pilots'. In the *kamera* word sketch in Figure 5.1, [*no* pronom] is the fourth most preferred gramrel and the top-3 collocates (highlighted within the purple box) are *kamera no Kitamura* 'literally, Kitamura who is involved with cameras (the name of a nationwide camera shop using the family name 'Kitamura'), *kamera no shattaa* 'a camera's shutter', and *kamera no furasshu* 'a camera's flash'. The [pronom *no*] gramrel is the reverse of [*no* pronom], with the search word being specialised by a noun or noun phrase occurring before it. The top-three collocates of this gramrel in the *kamera* word sketch above (highlighted within the blue box) are *pikuseru no kamera* 'pixel camera' (e.g. 'a 10 megapixel camera'), *Nikon no camera* 'a Nikon camera' (i.e. a 'Nikon' brand camera'), and *kyanon no kamera* (a 'Canon brand camera').

Returning to the fact that [noun/noun] is shown in Table 5.8 (and in more detail in Table 5.2) to be a very highly frequent grammatical relationship in which the English loanwords participate in the corpus, it is important to examine the specific grammatical relation expressed by this gramrel in the Japanese sketch grammar. [noun/noun] represents a grammatical relationship whereby a search term occurs as a modifying noun in a compound noun. The basic word structure of the Japanese language places the categorical 'head' (the part of a compound indicating the lexical category of the whole compound) on the right-hand side and the modifier on the left. This makes Japanese a 'head-final' language in terms of the head directionality parameter (Namiki & Kageyama, 2016). Because of this, the modifying element of a compound noun in Japanese consistently precedes the head noun (Namiki & Kageyama, 2016, p. 203). Whilst left-headed compound structures do exist, these are restricted to compounds consisting of a Sino-Japanese verbal element and nominal element, such as 帰国 *kikoku* 'return to one's own country' where the verbal element 帰 on the left is the head of the compound. Double-headed compound nouns also exist, such as 日英 *nichiei* 'Japan and the UK' but these are considered to be rare (Namiki & Kageyama, 2016, p. 214). The algorithm which defines the [noun/noun] gramrel in the Japanese sketch grammar is structured in such a way that it picks up the search word as this left-hand modifying part of a noun compound.

The principal feature of this gramrel is that it indicates that the search word is being used as a noun and is followed by another noun in the formation of a noun + noun compound. For example, in Figure 5.1 above of the loanword *kamera* 'camera', it can be seen within the green-coloured box that [noun/noun] is the second most preferred gramrel. The top-three collocates in this relationship (as ranked by the logDice measure) are カメラ目線 *kamera mesen* 'a line of sight into a camera'; カメラアングル *kamera anguru* ' a camera angle'; and カメラ小僧 *kamera kozou* 'an amateur photographer' (literally, a boy/man who likes using a camera to take pictures of female models and celebrities). In these three instances, *kamera* is being used as an attributive noun. In the first example, the head of the compound noun is 'line of sight', in the second example, the head is 'angle', and in the third the head is 'amateur or boy/man'. All three head nouns are then being attributively modified by the loanword *kamera*. This compound structure of a loanword occurring on the left and functioning to modify the noun on the right can be seen for all the other collocates in this gramrel and is the regular pattern across all instances of this gramrel in the word sketches. The [noun/noun] gramrel therefore signifies a specific, restricted type of grammatical behaviour in comparison to

the broad range of grammatical relationships expressed by the [particle] gramrel, which is better thought of as a large collection of individual grammatical relationships.

Looking again at Tables 5.7 and 5.8, it can be seen that the [noun/noun] gramrel appears in both of them, meaning there is an overall high number of tokens of this gramrel in both the non-loanword and loanword gramrel databases. The very important difference, however, is in how the tokens of this gramrel are distributed across the gramrel columns in the word sketches. Table 5.7 shows that for the non-loanwords, the [noun/noun] gramrel appears with a somewhat haphazard distribution across the top-10 most preferred gramrels. There is no sharp decline, for example, in the number of tokens of this gramrel as it becomes the non-loanwords' 'lesser-preferred' gramrel. Indeed, the number of tokens of [noun/noun] as the eighth most preferred gramrel (10 tokens) is higher than as the first (eight tokens). In Table 5.8 the picture is a different one. Here there is an obvious sharp decline in the number of tokens, particularly when comparing the number of tokens of [noun/noun] as the second most preferred gramrel (274 tokens) and as the third (51 tokens). This decline continues with there being only one token as the tenth most preferred gramrel.

This finding that the [noun/noun] gramrel is a dominant grammatical relationship among the English loanwords, but not a dominant one among the non-loanwords, is a very important one because it suggests a major difference in the grammatical behaviour of the English loanwords in Japanese compared to the native and Sino-Japanese words. The exploration of this finding will address the third research question of this study: what factors appear to account for this difference in the patterns of distribution in the grammatical relationships of the loanwords and non-loanwords. This is the subject of the next chapter.

5.6 Chapter Summary

In exploring answers to the questions put forward at the end of Chapter Three, of what patterns of distribution can be observed in the grammatical relationships of English loanwords in Japanese and in what ways these patterns of distribution are similar to and different from patterns of distribution in the grammatical relationships of non-loanwords in Japanese, this chapter has compared data on the gramrels of 130 frequently-used native and Sino-Japanese words and 587 English loanwords in the Japanese language. Three significant patterns of behaviour were found in this

comparison. The first pattern expresses a similarity in the distribution of the grammatical behaviour of the loanwords and non-loanwords, with a small number of gramrel types accounting for a large number of the 1300 non-loanword gramrel tokens and the 5870 loanword gramrel tokens. Viewed from a different perspective, the majority of the gramrels in both datasets were seen to have a low number of tokens, meaning that the loanwords and non-loanwords participated in a large number of gramrel types infrequently, and a small number of gramrel types frequently.

The second pattern concerns the gramrels occurring as the first, second, and third most preferred gramrels of the two sets of words, and again shows similarities of grammatical behaviour in that a very similar small number of gramrel types with a high number of tokens were found in the two gramrel databases. In particular, the [particle] gramrel was found to be an extremely frequent grammatical relationship of both sets of words. It was discussed, however, that the [particle] gramrel is special in that it combines together a wide range of individual particle-based gramrels. When this gramrel was discounted from the analysis of both datasets, a very important behavioural difference in the loanwords compared to the non-loanwords was revealed. This difference was discussed as the third major pattern of behaviour found in the comparison of the gramrel databases, in that the [noun/noun] gramrel was seen to be a much more dominant gramrel of the loanwords than of the non-loanwords. The next chapter discusses what factors appear to account for this difference.

6 Findings and Discussion 2: Accounting for Differences

6.1 Chapter Overview

This chapter analyses the collocational behaviour of a sample of the 587 English loanwords in Japanese discussed in the previous chapter. Its purpose is for the collocational behaviour of the English loanwords to be used in helping to account for the differences found in the previous chapter of the patterns of distribution in the grammatical relationships of loanwords and non-loanwords in Japanese. In particular, patterns in the collocational behaviour are used to try to account for the high incidence of loanwords having the [noun/noun] grammatical relationship as their first or second-preferred grammatical relationship in the word sketches. This behaviour was not found with the non-loanwords and will be explored in this chapter with a collocational analysis of a sub-sample of the loanwords. Following this analysis, an attempt will be made to account for the behaviour of the loanwords by referring to a theory of the functions of loanwords in relation to the categories ‘catachrestic’ and ‘non-catachrestic’.

6.2 Non-Loanwords with [noun/noun] as the Most Preferred Gramrel

It was explained in the previous chapter that because the [particle] gramrel in the Japanese word sketches displays a special behaviour in how it subsumes a large range of individual particle-based gramrels, it can be expected to be the first gramrel to occur in a very large number of the word sketches of Japanese vocabulary (i.e. the grammatical relationship in which words most frequently participate). This was shown to indeed be the case in the non-loanword gramrel database, with 81.6% of the non-loanword word sketches (106 of 130) showing [particle] to be the first listed gramrel column in the word sketches. For the remaining 24 word sketches, the first listed gramrel was [noun/noun] for 6.2% (eight of 130), [suffix_base] for 5.4% (seven), [distant_V] for 3.1% (four), and [wa verb], [suffix], [prefix], [distant_Adv], and [noun ni] for 0.8% (one). By contrast, in the loanword database, 36.2% of the loanwords (212 of 587) had [noun/noun] as the first most preferred gramrel. This means that [noun/noun] is around six times more often seen as the first most preferred gramrel of the loanwords than it is of the non-loanwords. Furthermore, when the data of the second most preferred gramrel is added in, the difference becomes much more striking. For the

non-loanwords, [noun/noun] appears in either the first or second gramrel column in 20.76% of the word sketches (27 of 130). For the loanwords it is 82.79% (486 out of the 587). In order to account for this difference, it is first of all necessary to examine the non-loanword gramrels in more detail.

Table 6.1 shows the eight non-loanwords with the [noun/noun] gramrel as the first listed gramrel in their word sketches.

Table 6.1 Eight non-loanwords with [noun/noun] as their first most preferred gramrel.

	Kanji	Romanisation	English
1	日本	*nihon*	Japan
2	一	*ichi*	one
3	二	*ni*	two
4	三	*san*	three
5	町	*chou/machi*	town/city
6	四	*shi/yon*	four
7	一度	*ichido*	once
8	六	*roku*	six

Five of these words are of the same type, being the Sino-Japanese numbers one, two, three, four, and six. An examination of the word sketches of these five words showed that they are very similar in terms of gramrel types and the ordering of the gramrels. For example, each of these words is found in the [noun/noun] gramrel about 20 times as often as the their second listed gramrel, which for all of them is [prefix], and they all have just four gramrel columns shown in their word sketches (whereas word sketches typically show more than 30 gramrel columns). Figure 6.1 shows an example of this with the word sketch of 一 *ichi* ‘one’ only containing four gramrel columns.

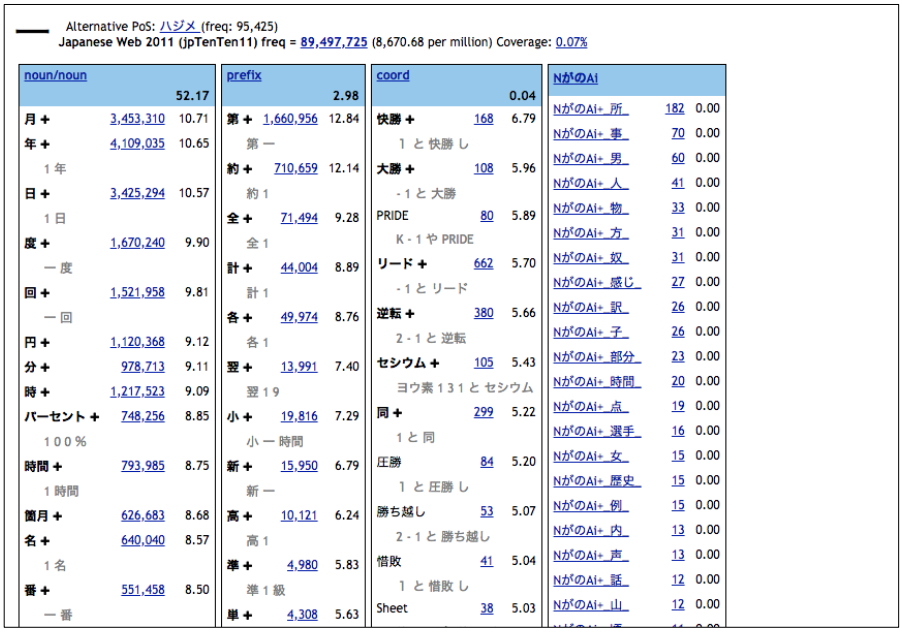


Figure 6.1 The word sketch of *ichi* 'one' showing that it contains only 4 gramrel columns.

The word sketches for these five number-words reflect their specific function in the Japanese language, which is that they are used in conjunction with counter words (called *josuushi* in Japanese) when counting things such as units of time (e.g. *ichi nen* 'one year'), units of money (*ichi en* 'one yen'), and units of measurement (e.g. *ichi guramu* 'one gram'). It is this function of the number typically combining with something being counted that explains why the [noun/noun] gramrel, which expresses a compound noun structure, and [prefix] which expresses the number being bound to a prefix (such as *yaku* 'about') are the two most preferred gramrels of these numbers, rather than the [particle] gramrel.

Two of the other non-loanwords in Table 6.1 display a behaviour similar to one another, and this behaviour explains why their first most preferred gramrel is [noun/noun]. The word 町 *chou/machi* means 'town/city/street' and is very commonly used in addresses and on forms asking people to write their addresses. In the [noun/noun] gramrel column of this word in its word sketch, the most salient collocation is 村 *mura* 'village'. The 'longest-commonest match' for this collocational pair (i.e. the most common realisation of this collocational pair in the corpus, see Section 4.3.4 for the explanation of this terminology) is 市町村 *shichouson* 'city, town, village' which is used on application forms to indicate where to write out one's full address. As such, it very commonly is used in a compound noun structure. Similarly,

日本 *nihon* is the word used to mean ‘Japan’ and therefore appears very frequently in a compound noun structure. In its word sketch, the top-three most salient [noun/noun] collocations are *nihongo* ‘Japanese language’, *nihon daihyou* ‘Japanese representative’, and *nihon shakai*, ‘Japanese society’. The last remaining word in Table 6.1 is 一度. This word is somewhat of an anomaly in this set of words because it is actually a very frequent collocational realisation of the number ‘one’ with the counter of ‘time’. This word is an example of the difficulty in tokenising words in a corpus, in terms of whether certain words are analyzable into smaller words, which is especially problematic for languages without white spaces between words (Tono et al., 2013). The word sketch for 一度, shown in Figure 6.2, is not very informative, with only very low frequencies for several collocates in three gramrels, suggesting that it is most commonly treated by the software as a collocational pair, or multiword expression (Kilgarriff et al., 2014), rather than a single word.

Figure 6.2 The word sketch for *ichido* ‘one time’.

The above discussion has shown that the eight non-loanwords in Table 6.1 with [noun/noun] as their first most-preferred gramrel, which is behaviour that is different from the other 122 words in the 130-word sample, can be accounted for by their specific function in the language. However, another analysis needs to be conducted on this sample of words because, as discussed in Section 4.3.4, it could be the case that there is a very small difference in frequencies between the gramrels listed first and second in a word sketch. As a highly-exaggerated theoretical example, whilst only eight non-

loanwords have [noun/noun] appearing as the first gramrel column in their word sketches, there could be the very small chance that all the other 122 word sketches with [particle] as the first gramrel column have [noun/noun] appearing in the second gramrel column with only a minimal difference in frequencies. This, however, was shown not to be the case in Table 5.1 in the previous chapter, which showed that [noun/noun] only occurred as the second most preferred gramrel of 19 of the 130 words (14.7%). Furthermore, Table 6.2 below shows that when [particle] is the first most preferred gramrel, there is no discernible large difference in the number of times that the gramrels which are second most preferred occur. To explain this more clearly, Table 6.2 shows that 21 of the 130 words (16.2%) had the one-two gramrel pairing of [particle] + [*o* verb], 19 (14.7%) had the pairing of [particle] + [noun/noun], 14 had [particle] + [*no* pronom], etc. None of these pairings appear particularly remarkable.

Table 6.2 Rates of occurrence of first and second most preferred non-loanword gramrels.

Pairing of Gramrel 1 + Gramrel 2	Number of Tokens of this Pairing
particle	**106**
o verb	21
noun/noun	19
no pronom	14
pronom no	11
suffix_base	11
ni verb	10
distant_Adv	7
de verb	3
suffix	2
prefix	2
wa verb	2
prefix_base	1
N_Adj	1
ga verb	1
modifier_Ana	1
noun/noun	**8**
prefix	5
particle	2
N_Adj	1
suffix_base	**7**
particle	7
distant_V	**4**

distant_N+suru	3
modifies_V	1
noun ni	**1**
coord	1
wa verb	**1**
particle	1
suffix	**1**
particle	1
distant_Adv	**1**
particle	1
prefix	**1**
suffix	1
Total	**130**

This exploration of the [noun/noun] gramrel in the non-loanword database has revealed overall a mostly predictable grammatical behaviour of the native and Sino-Japanese words. The fact that 122 of the 130 words (93.9%) participate most frequently in a [particle] gramrel is a mostly predictable outcome of [particle] exhibiting the special characteristic of subsuming a large number of particle-based gramrels. Moreover, the eight words occurring most frequently in a [noun/noun] grammatical relationship can be attributed to their specific function in the language. Lastly, apart from this [particle] gramrel, there is no other indication from the tables of data discussed above and in the last chapter (especially in the conditional colour-formatting of Table 5.1) of a particular aspect of the grammatical behaviour of the non-loanwords being anything different from what would be expected of frequently-used nouns in a language. To explain this last point in more detail, because nouns play a whole range of syntactic and morphological roles in a language, such as subject, direct object, adjectival noun, and prefix, it is expected that an examination of the grammatical behaviour of a large sample of the most-frequently used nouns in a language would reveal the kind of behaviour that has been found in the native and Sino-Japanese words in this study. The question then arises as to what can account for the extraordinarily high occurrence of the [noun/noun] gramrel in the word sketches of the English loanwords compared to the non-loanwords.

6.3 Loanwords with [noun/noun] as the Most Preferred Gramrel

Chapter Five showed that the most significant difference in the grammatical behaviour of the English loanwords compared to the non-loanwords is the very high frequency of the [noun/noun] gramrel in the loanword database. Summarising this behaviour again, in the database of 5870 loanword gramrels, [noun/noun] is observed 577 times out of a maximum possible 587 (98.3%) (i.e. once per loanword word sketch). In comparison, it was observed 105 times out of a maximum possible 130 (80.8%) in the non-loanword database. These numbers alone are not vastly different, but for the non-loanwords the 105 appearances of [noun/noun] are spread out among the top-10 gramrel columns in the word sketches. This means that whilst non-loanwords do participate quite frequently in this grammatical relationship in the jpTenTen11 corpus, this participation is not, for the majority of non-loanwords, their most preferred grammatical relationship. This is different for the loanwords, however, which display a very strong preference for participating in the [noun/noun] grammatical relation. Of the 577 times that this gramrel is observed in the loanword database, 91.48% (537) occur in the first three gramrel columns of the word sketches, compared to 29.23% for the non-loanwords.

Using the same analysis as in Table 6.2 above, Table 6.3 analyses the gramrel pairings of the first + second most preferred gramrels in the total set of 587 loanword word sketches. There are two data points in the figures of the second most preferred gramrel which are considerably more prominent than the others, shown in the table with boxes around the number. Looking at the [particle] gramrel, for the 358 loanwords with this as the first most preferred gramrel, 273 (76.3%) have [noun/noun] as the second-preferred gramrel. Then looking at when [noun/noun] is the first most preferred gramrel, 187 of the 212 loanwords (88.2%) have [particle] as the second-preferred gramrel. The only other pairing of gramrels with a rate of occurrence approaching, but still a long way off, these two patterns is that of [*o* verb] occurring as the second most preferred gramrel for 58 of the 358 loanwords (16.2%) when [particle] is first most preferred. Therefore, the two patterns highlighted in Table 6.3 of (1) [noun/noun] + [particle], and (2) [particle] + [noun/noun], are clearly dominant pairings. This is in contrast to the non-loanword database where it was discussed above that no gramrel pairing was noticeably much more dominant than another. The question then arises of what factors can be seen to account for this marked grammatical behaviour of the English loanwords.

Table 6.3 Rates of occurrence of first and second most preferred loanword gramrels.

Pairing of Gramrel 1 + Gramrel 2	Number of Tokens of this Pairing
particle	**358**
noun/noun	273
o verb	58
pronom no	8
distant_Adv	6
no pronom	5
de verb	3
ni verb	3
N_Adj	1
ga verb	1
noun/noun	**212**
particle	187
N_Adj	12
pronom no	5
distant_Adv	4
na_modifies_N	2
suffix	1
noun o	1
N_Adj	**7**
na_modifies_N	5
Ana_ni modifies_V	1
particle	1
na_modifies_N	**4**
Adv	2
N_Adj	2
distant_Adv	**2**
noun o	1
noun/noun	1
noun o	**2**
distant_Adv	1
particle	1
Ana_ni modifies_V	**1**
na_modifies_N	1
suffix	**1**
na_modifies_N	1

The discussion above of the non-loanwords showed that five of the eight words with [noun/noun] appearing in the first gramrel column of their word sketches (or five out of seven if *ichido* 'once' is reanalyzed as a frequent collocation using the numeral *ichi* 'one') have a special function in Japanese: as Sino-Japanese numerals used in combination with counters. The very large number of loanwords with [noun/noun] in the first gramrel column, however, suggests that this behaviour cannot be accounted for in the same way. For example, it is unlikely that the majority of the loanwords are numerals. Furthermore, many of them may display behaviour similar to the non-loanwords *nihon* 'Japan' and *chou/machi* 'town/city', in that they are frequent nouns which because of their meaning often get compounded with other nouns, but again the low incidence of this behaviour with the non-loanwords would suggest there is a different factor involved in the high incidence of this behaviour with the loanwords.

To investigate this further, Table 6.4 shows the first 20 loanwords from the list of 212 English loanwords which have the [noun/noun] gramrel as the first gramrel column in their word sketches. It can be seen that unlike the words in Table 6.1, there is no obvious way in which the words are related to one another, or in which they can be categorized into groups with a special function such as with the non-loanword numerals. The loanwords instead refer to a disparate range of things, such as technological concepts/objects (*intaanetto* 'internet', *nettowaaku* 'network', *bideo* 'video'), spatial positions (*toppu* 'top', *bakku* 'back') and activities (*supootsu* 'sports', *bijinesu* 'business'). For this reason, other factors need to be found to account for the special grammatical behaviour of the loanwords. For this, three theories will be returned to which were discussed in the review of previous research on loanwords in Chapter Three.

Table 6.4 The first 20 loanwords in the list of 212 English loanwords which have [noun/noun] as their first most preferred gramrel.

	Top-20 English Loanwords with [noun/noun] as the first most frequent gramrel		
	Katakana character	Romaji	English
1	バック	*bakku*	back
2	ソフト	*sofuto*	soft[18]
3	セット	*setto*	set
4	トップ	*toppu*	top
5	スポーツ	*supootsu*	sports
6	インターネット	*intaaneto*	internet
7	ビジネス	*bijinesu*	business
8	ネットワーク	*nettowaaku*	network
9	アクセス	*akusesu*	access
10	スタート	*sutaato*	start
11	タイム	*taimu*	time
12	ブランド	*burando*	brand
13	ビデオ	*bideo*	video
14	スーパー	*suupaa*	super(market)
15	オープン	*oopun*	open
16	デジタル	*dejitaru*	digital
17	フリー	*furii*	free
18	レビュー	*rebyuu*	review
19	サポート	*sapooto*	support
20	パワー	*pawaa*	power

[18] *sofuto* (‘soft’) is an example of English word which is an adjective in its original English form but treated as a noun in its Japanese loanword form. To be used as an adjective in Japanese, it needs to be followed by the grammatical marker *-na*, as in *sofuto na kanji* ‘a soft feeling’. *furii* (‘free’) further down the table is another example of this behaviour of an adjective in English being treated as a noun loanword in Japanese, as in *furii na hito* ‘a free person’.

6.4 Accounting for the Grammatical Behaviour of the Loanwords

From the review in Chapter Three of research carried out in the field of lexical borrowing, there were several studies putting forward theories concerning the behaviour of loanwords which may help in accounting for the behaviour of the English loanwords uncovered in Chapter Five. Loveday (1996), for example, put forward the theory that there are Japanizing and Westernizing patterns of loanword usage in Japan, whereby some loanwords refer primarily to the Western conceptualisation of an object, such as a wooden door (*doa*) rather than the traditional Japanese *shoji* ('paper screen'). As such, it may be the case that the loanwords participating in the [noun/noun] gramrel are acting as such Western-object referents and this may explain their constrained usage in a compound noun. Secondly, Doi (2014) posited a theory of the process of loanword naturalisation, explaining that some loanwords are restricted to attributive (i.e. modifier) usage in compounds. This seems to relate well to the behaviour of the English loanwords found in Chapter Five. There was also a discussion of the theory of catachrestic and non-catachrestic innovations, put forward by Onysko and Winter-Froemel (2011) and ultimately based on a century-old traditional distinction between necessary and luxury loanwords. It could be that the English loanwords occurring most frequently in the [noun/noun] gramrel are non-catachrestic/luxury loanwords, which could account for their constrained grammatical behaviour. These three theories will be discussed in more detail to see how well they are able to account for the behaviour of the English loanwords in the present study.

6.4.1 *The theory of Japanizing and Westernizing patterns*

Loveday (1996) theorises how some loanwords in Japanese are used to refer to Western concepts and objects in comparison to native equivalent expressions for the loanwords which refer to the basic concept. He argues that these loanwords are thereby signifiers of material transformation, where Japanese cultural concepts have been materially transformed by processes of modernisation and have resulted in the proliferation of loanwords used to describe the transformed concepts. He gives the examples of *sutoroberii* and *biifu* being loanwords referring to westernised versions of the basic concepts of 'a strawberry' and 'beef'. It could be the case then that the extraordinary grammatical behaviour of the English loanwords found in the present study can be

accounted for by this theory: that the [noun/noun] behaviour of the loanwords represents their use as referring to Western concepts, or what Loveday (1996) calls Westernizing patterns of behaviour.

However, in a discussion of Loveday's theory by Bordilovskaya (2016), which is also commented upon in Chapter Three, she points out two major weaknesses. She discusses how his discussion is only of concrete nouns, such as the aforementioned 'strawberry' and 'beef' and others such as 'rice' and 'door', with no indication of whether the theory applies to other more abstract types of loanwords such as *risuku* 'risk' and other parts of speech such as loanword adjectives. Bordilovskaya's further criticism of the theory is that it is not based on systematic evidence, but rather on anecdotal observations of how loanwords are used. Furthermore, the theory is premised on the assumption of loanwords having a complementary native equivalent expression with which to distinguish between foreign references (e.g. *sutoroberii sheeku* 'strawberry shake', where strawberry is a flavour) and native references (e.g. *ichigo* 'strawberry', where strawberry is the fruit).

These three issues make Loveday's theory unable to account adequately for the [noun/noun] gramrel behaviour of the loanwords in the present study. Not only does the theory lack a solid empirical basis, but as Table 6.4 above shows, not all of the 587 loanwords in the present study are concrete nouns. As a result, whilst it may be true that some of the loanwords are behaving in such a way of referring to Western concepts, many of the 587 loanwords will fall outside of the scope of the theory.

6.4.2 *The theory of the naturalisation of loanwords*

Chapter Three also reviewed Doi's (2014) investigation of the processes of naturalisation of Japanese loanwords into the English language, where he comments that at one stage during the naturalisation process, Japanese loanwords are often restricted to an attributive usage whereby the loanword functions as a quasi-adjective, acting to modify a following noun (p. 682). He gives the example of 'tatami room', where the Japanese loanword 'tatami' is used to describe the type of room. This idea of restricted attributive usage seems to relate to the findings of the present study, in that the [noun/noun] gramrel represents such a restricted behaviour of loanwords being used attributively in compound nouns (see section 6.5 for examples of this compound noun behaviour of English loanwords).

The problem, however, with using Doi's naturalisation theory to account for the marked behaviour of the loanwords is that it is premised upon a scale of gradual incorporation of the Japanese loanwords into the linguistic structure of English. At one end of the scale the loanwords are considered "totally foreign", and at the other end "fully incorporated/native" (p. 677). This scale is related to the frequency of the loanword in English, with the totally foreign loanwords being infrequent and therefore needing paraphrases, glossing, and italicising to explain their meanings. In this way, they have not undergone extensive phonological, orthographical, morphological, and semantic integration into the language. The fully incorporated loanwords, on the other hand, by way of their higher frequency in the language, do not need these markings because they have been extensively adapted to fit the English linguistic system. Because of this, they can be used productively, for example in derivative structures such as 'kimonoed beauties of Japan' (2014, p. 684). It was explained in Chapter Four, however, that all of the 587 loanwords in the present study are frequently-used English loanwords in Japanese which have been extensively adapted for their incorporation into the language (e.g. they are written in *katakana* rather than the Latin alphabet). This means that the set of loanwords in the present study cannot be analysed along this continuum of totally foreign to fully incorporated.

Furthermore Doi states in his paper that "when discussing loanwords from a particular language, generalisations should primarily be based on that single language; loanwords from other languages might behave differently" (2014, 676). He makes this comment in relation to an earlier four-stage process of naturalisation he found in a paper focusing on Malayan loanwords. The author of this framework believed that it could also be applied to Japanese loanwords, a point which Doi disagrees with when he states that "although there should be general tendencies among languages, it is only after the mechanisms for both source languages are determined that an overview of the similarities and dissimilarities between the naturalisation process can be made" (2014, p. 676). As such, it is not fitting to lift the naturalisation process which Doi discusses in his paper on Japanese loanwords in English and apply it to the situation of English loanwords in Japanese.

6.4.3 *The theory of catachrestic and non-catachrestic innovations in a language*

There was one other theory introduced in Chapter Three that may be able to account for the behaviour of the English loanwords seen in Chapter Five. This is the theory of catachrestic and non-catachrestic innovations in a language, explored most extensively in the work of Onysko and Winter-Froemel (2011). This theory was briefly discussed in Section 3.7 and will be explored more extensively here. The origins of the theory have a long tradition in lexical borrowing research, and discuss how the functions of loanwords in a language can be conceptualised from an either/or perspective: either they fill a lexical-gap as words for newly introduced cultural concepts, or they give a special-effect as words used for their pragmatic contrast with native equivalents (Durkin, 2014; Haspelmath, 2008; Inagawa, 2010, 2012; Rebuck, 2002; Takashi, 1990). This is the classical model of categorisation, seen since the influential works of Haugen (1950) and Weinreich (1953) in the mid-twentieth century, and has become a scholarly tradition across the whole field of lexical borrowing research (Diez-Arroyo, 2016, p. 611).

These works of Haugen (1950) and Weinreich (1953) renewed interest in the topic of lexical borrowing (see Section 3.2), and helped the labels necessary and luxury used as far back as 1886 become the standard expressions to describe the functions of loanwords (Onysko & Winter-Froemel, 2011). Deroy (1956), cited in Onysko and Winter-Froemel (2011), for example, followed on from the luxury/necessary distinction and divided loanwords into those borrowed for affective reasons and those for practical necessity (Onysko & Winter-Froemel, 2011, p. 1551). The necessary loans are the ones seen to fill gaps in the recipient language opened up by the introduction of new cultural items and concepts for which there are no viable existing expressions. Luxury loans, on the other hand, are those which already have similar expressions in the recipient language, used mainly for adding pragmatic effects (i.e. as euphemistic terms and words which can be used in place of taboo words in the recipient language).

Dissatisfaction with the terms of necessary and luxury loans, in that they imply that luxury loans are not necessary in a language, has encouraged some researchers to develop alternative names for the categories. Myers-Scotton (2006) re-conceptualised the categories as cultural borrowings and core borrowings in her work on bilingualism and code-switching. Cultural borrowings, in her terminology, are the traditional category of necessary loans, in that they "fill gaps in the recipient language's store of words because they stand for objects or concepts new to the language's culture" (Myers-Scotton, 2006, p. 212). Core borrowings are the luxury loans, in that they

"duplicate elements that the recipient language already has in its word store. They are gratuitous – by definition, another layer on the cake, because the recipient language always has viable equivalents" (2006, p. 215). These new terms were then further updated in the work of Onysko and Winter-Froemel (2011), who analysed the functions of Anglicisms (i.e. English loanwords) in German. Using the literary term of catachresis which has traditionally been applied to the necessity of a metaphor in the rhetorical tradition, they put forward the categories of catachrestic innovations for Myers-Scotton's cultural borrowings or the earlier-named necessary loans; and non-catachrestic innovations for core borrowings or luxury loans (Onysko & Winter-Froemel, 2011). Table 6.5 summarises the historical development of the various names for these two categories of loanword function.

Table 6.5 Names given to the two main functional categories of loanwords (partly adapted from Onysko & Winter-Froemel, 2011).

Research Study	**No equivalent expression(s) in the borrowing language**	**Equivalent expression(s) in the borrowing language**
Paul (1886)	Necessary loans	Luxury loans
Deroy (1956)	Borrowings for practical necessity	Borrowings for affective reasons
Takashi (1990)	Lexical-gap fillers	Special-effects-givers
Myers-Scotton (2006)	Cultural borrowings	Core borrowings
Onysko & Winter-Froemel (2011)	Catachrestic innovations	Non-catachrestic innovations

In Onskyo and Winter-Froemel's (2011) study, the authors aimed to find a theoretical framework within which to explain "how far pragmatic and stylistic characteristics of loans can be classified on a unified theoretical basis" (2011, p. 1551). The theoretical framework they chose was to use the distinction between inferences of informativeness and manner from Levinson's theory of presumptive meanings (Levinson, 2000). For their terminology, as was discussed above, Onysko and Winter-Froemel proposed the terms catachrestic and non-catachrestic innovations to refer to loanwords displaying *i*-implicatures or *m*-implicatures respectively (Onysko & Winter-Froemel, 2011). The catachrestic innovations, in being words that introduce a new concept into the language, become the normal way of speaking about things, and

thereby convey implicatures of informativeness (*i*-implicatures). The non-catachrestic innovations, on the other hand, in being words for which there is already an equivalent expression, are the marked way of speaking about things, and thereby convey implicatures of manner (*m*-implicatures). Whilst Onysko and Winter-Froemel updated the terminology and the theoretical framework within which to discuss the functions of loanwords in a language, they did not, however, change the methodological approach of establishing the categorisation based on a dictionary check of whether the loanword has an equivalent expression in the language into which the word is being borrowed.

Looking at the grammatical behaviour of the loanwords analysed in the present study, it appears that this theoretical categorisation into loanwords conveying informativeness implicatures and those conveying manner implicatures could account for the two types of grammatical behaviour discussed in Chapter Five. It could be, for example, that the loanwords most frequently found in the present study occurring in a [particle] grammatical relationship are what Onysko and Winter-Froemel call catachrestic innovations, being used as the normal way of communicating about things because of their position of introducing a new concept into the language, just like the majority of the non-loanwords. Then, the large number of loanwords most frequently found occurring in a [noun/noun] grammatical relationship, which was a behaviour not seen with the non-loanwords, could be described as non-catachrestic innovations, being used as the marked way of communicating about things because of their position as having equivalent expressions already available in Japanese.

The following sections explore whether the behaviour of the English loanwords in the present study can be accounted for by this theory of the distinction between catachrestic and non-catachrestic innovations. This exploration begins with creating two lists of loanwords out of the findings from Chapter Five: one consisting of the 358 loanwords with [particle] as their first most preferred gramrel, and one consisting of the 212 loanwords with [noun/noun] as their first most preferred gramrel. Next, the loanwords in each list are ranked by the degree of difference between their first and second most preferred gramrels (see Chapter Four for an explanation) (the full [particle] list is given in Appendix 4 and the [noun/noun] list is given in Appendix 5). This was considered important because a loanword which has a first most preferred gramrel with a frequency many times more than that of the second most preferred gramrel could be said to have less variation in its grammatical behaviour than a loanword with a more equal frequency between the first and second most preferred gramrels. The issue of variation in grammatical behaviour is then to be examined with a detailed analysis

conducted on the word sketches of a sample of the loanwords taken from three different points on the ranked lists (see below for details of this sample). This analysis involves examining the collocational behaviour of the sampled loanwords to explore patterns between the lists in what kind of collocates the loanwords join with in their different grammatical relationships.

The reason for conducting an analysis of the loanwords' collocational behaviour to better understand their grammatical behaviour is because of similar work previously carried out by Bordilovskaya (2012, 2016), whose work was discussed at the end of Chapter Three. Focusing on English loanword adjectives, such as *nagai* 'long' and *atsui* 'hot', Bordilovskaya found a tendency for what she terms homogenous collocations, whereby some loanwords were seen to collocate more strongly with other loanwords rather than native and Sino Japanese words. She then uses this differentiation in collocational behaviour to account for a difference of the loanwords' functions in Japanese, with the loanwords collocating most strongly with other loanwords being used for a stylistic function, such as conveying positive connotations of modernity and fashion. She found that these loanwords often had an alternative expression in Japanese. The loanwords collocating more freely with native and Sino-Japanese words, on the other hand, she categorises as words that have become the standard means of communication, including the adjectives *infoomaru* 'informal' and *iregyuraa* 'irregular'. These loanwords, she found, often did not have an equivalent expression in Japanese.

This varied collocational behaviour of English loanwords with words from other strata in the Japanese lexicon therefore seems to be able to reveal general behavioural differences of the loanwords in the language. As such, the following section will take a sub-sample of the loanwords from the [noun/noun] list and the [particle] list and explore whether patterns can be identified in whether they collocate more often with other loanwords or with other non-loanwords. The results of the collocational analysis will then be discussed in light of the theory discussed above of the distinction between catachrestic and non-catachrestic innovations, to see if this theory can indeed account for the marked [noun/noun] grammatical behaviour of the loanwords.

6.5 Collocational Behaviour of the English Loanwords

The lists of 212 loanwords with [noun/noun] as their first most preferred gramrel and 358 loanwords with [particle] as their first most preferred gramrel were ranked in high-low order of the degree of difference between the first and second most preferred gramrels (hereafter, [noun/noun] list and [particle] list). A sample of five loanwords was taken from the top, middle, and bottom of each list in order to represent different points on the scale of frequency variation and thereby obtain an unbiased selection of loanwords. More specifically, the first five, middle five, and last five loanwords in each ranked list were selected. After each of the tables of collocates below, there is a summary of the main findings and then an overall summary in Table 6.18 of the general behaviour of the collocates from both lists. To begin the analysis, Table 6.6 shows the five loanwords from the top of the [noun/noun] list. The 'Difference' column shows how many times bigger the frequency of the first gramrel ('Gramrel 1 Freq.') is compared with the second ('Gramrel 2 Freq.'), and is used to rank the two lists in high-low order of degree of difference. The loanword *ando* 'and', for example, occurs in the [noun/noun] gramrel almost 33 times more often than in its second most preferred gramrel.

Table 6.6 The top-5 English loanwords from the [noun/noun] list.

	Katakana	Romaji	English	Gramrel 1 Freq.	Gramrel 2 Freq.	Difference
1	アンド	*ando*	and	52,290	1,552	32.7
2	バイオ	*baio*	bio	105,332	6,103	16.3
3	ニュー	*nyuu*	new	330,414	24,516	12.5
4	オート	*ooto*	auto	171,849	13,038	12.2
5	フォト	*foto*	photo	120,640	9,646	11.5

Table 6.7 below presents the top-10 most salient collocates of each word in Table 6.6 above. As explained in Chapter Four, these collocates are not specific to one grammatical relationship but rather are the most salient collocates from amongst all the collocates across all the grammatical relationships. This analysis can be done in the Sketch Engine by de-selecting the option to structure word sketch by gramrels (see Figure 4.7 in Chapter Four). Doing so presents a single list of collocates, ranked by the logDice score, rather than an output structured into columns of gramrels with the collocates ranked within the gramrels. In this single list of collocates, data is also given

of the gramrel in which the collocate was found to be most salient, as well as the Longest-Commonest Match of the search word + collocate pair, which is the longest, most common realisation in the corpus of the triple of the search word + gramrel + collocate. This data is shown in Table 6.7, which also includes a column of an English translation of the longest-commonest match.

The colour-coding of Table 6.7 is central to an overall understanding of the collocational behaviour of the loanword sample, and is the same colour coding used in the other tables of collocates in this section. The use of red text highlights the collocates which are also loanwords in the Japanese language (typically written in *katakana* or in some cases Romanised), and the use of green text highlights the collocates which are native or Sino-Japanese words (typically written in *kanji* and/or *hiragana* and in a few rare cases Romanised). As was explained in Chapter Four, the fact that written Japanese uses the *katakana* script as the standard form of writing loanwords means that this distinction between loanword and non-loanword collocates could be made easily. However, there are some cases in written Japanese where scripts are used in creative ways which can mean that loanwords are written in scripts other than *katakana*, and also that non-loanwords can be written in *katakana*. Each collocate was therefore double-checked to ensure it was colour-coded correctly.

The light-grey colouring of several collocates (but with no strike-through) means they have been discounted from the analysis because they represent problematic collocations. This could be that the collocate is an international name, such as 'Photoshop' and 'World Trade Center' (and therefore does not represent a Japanese usage of the loanword), or a non-English loanword collocation, such as the loanword *ooto* which can be English 'art' as well as French 'haute', as in *ooto kuchuuru* 'haute couture' (expensive clothing). To explain this method of exclusion in more detail, each collocation was individually investigated on the internet to analyse its meaning, with inclusions or exclusions based on the results of the internet analysis. In Table 6.9 for example, the collocation スターウォーズ 'Star Wars' was excluded but the collocation スターオーシャン 'Star Ocean' was retained. This is because 'Star Wars' is an international brand name whereas 'Star Ocean' is a Japanese usage of English for the naming of a video game. As a further example, Table 6.11 includes the collocations 'Station Wagon', which was removed from the analysis, and 'PlayStation', which was retained. In this case, 'Station Wagon' is an international usage of the word 'station' whilst 'PlayStation' is a Japanese usage because the device was originally developed and branded in Japan, by Sony Corporation.

The collocates with light-grey shading and a strike-through represent what I judged to be an error in the automated word sketch output. Such errors are an inevitable part of the automation of huge amounts of corpus-based data, not only in the accuracy of the initial part-of-speech tagging in the corpus, but especially in the tokenisation of languages such as Japanese without white spaces between words. These issues are discussed in Chapter Four along with the steps taken by the Sketch Engine developers to limit these issues, however only a minimal number of collocates in each table were judged as errors. Furthermore, a second rater of these errors was not used in this study because the vast majority of cases were clear from a check of the translations of the Longest-Commonest Matches.

Table 6.7 The most salient collocates and Longest-Commonest Match for the top-five English loanwords on the [noun/noun] list.

Salient Collocates	Gramrel	logDice	Longest-Commonest Match	English Translation
[noun/noun] List Top Word 1: *ando* 'and'				
テーク	*noun/noun*	8.63	ギブ アンド テイク (14%)	give and take
アウェー	*noun/noun*	8.04	ヒット アンド アウェイ (12%)	hit and away
ドロップ	*noun/noun*	8.02	ドラッグ アンド ドロップ で (6%)	drag and drop
フィール	*noun/noun*	7.57	の ルック アンド フィール (14%)	look and feel
~~ロペ~~	~~*noun/noun*~~	~~7.4~~	~~アンド ロペ ニス (17%)~~	~~andropenis~~
レスポンス	*noun/noun*	7.26	の コール アンド レスポンス (4%)	call and response
ライド	*noun/noun*	7.21	パーク アンド ライド (17%)	park and ride
ゴー	*noun/noun*	7.18	タッチ アンド ゴー (6%)	touch and go
ペースト	*noun/noun*	7.12	コピー アンド ペースト (13%)	copy and paste
~~ラス~~	~~*noun/noun*~~	~~7.11~~	~~アンド ラス (16%)~~	~~andruss~~
[noun/noun] List Top Word 2: *baio* 'bio'				
エタノール	*noun/noun*	9.89	バイオ エタノール (18%)	bioethanol
燃料	*noun/noun*	9.65	バイオ 燃料 (17%)	biofuel
ディーゼル	*noun/noun*	9.25	バイオ ディーゼル 燃料 (8%)	biodiesel fuel
メトリクス	*noun/noun*	8.37	バイオ メトリクス (16%)	biometrics
マーカー	*noun/noun*	8.36	バイオ マーカー (19%)	biomarker
インフォーマティック	*noun/noun*	8.09	バイオ インフォマティクス (19%)	bioinformatics
サイエンス	*noun/noun*	8.06	バイオ サイエンス 研究 (8%)	bioscience research
フィルム	*noun/noun*	7.94	バイオ フィルム (16%)	biofilm
ジェル	*noun/noun*	7.78	バイオ ジェル (8%)	bio gel
ベンチャー	*noun/noun*	7.75	バイオ ベンチャー (20%)	bio venture
[noun/noun] List Top Word 3: *nyuu* 'new'				
ハーフ	*noun/noun*	*10.58*	ニュー ハーフ (11%)	new half
タウン	*noun/noun*	*10.1*	ニュー タウン (16%)	new town
速	*noun/noun*	*9.81*	ニュー 速 (11%)	new speed
オオタニ	*noun/noun*	*9.65*	ホテル ニュー オータニ (7%)	Hotel New Otani
イヤー	*noun/noun*	*9.09*	ニュー イヤー (18%)	new year
ウエーブ	*noun/noun*	*8.9*	ニュー ウェーブ (7%)	new wave
アルバム	*noun/noun*	*8.74*	の ニュー アルバム (6%)	new album
シングル	*noun/noun*	*8.4*	の ニュー シングル (11%)	new single
エージ	*noun/noun*	*8.26*	ニュー エイジ (18%)	new age
フェース	*noun/noun*	*8.18*	ニュー フェイス (22%)	new face

[noun/noun] List Top Word 4: *ooto* 'auto'				
バックス	*noun/noun*	10.22	オート バックス (14%)	AUTOBACS
フォーカス	*noun/noun*	9.85	オート フォーカス (18%)	autofocus
クチュール	*noun/noun*	9.48	オート クチュール (21%)	haute couture
キャンプ	*noun/noun*	9.21	オート キャンプ 場 (5%)	auto camping ground
ロック	*noun/noun*	8.67	オート ロック (15%)	auto lock
ボット	*noun/noun*	8.58	オート ボット (22%)	autobot
ポリス	*noun/noun*	8.1	オート ポリス (16%)	autopolis
チャージ	*noun/noun*	7.96	オート チャージ (13%)	auto charge
メッセ	*noun/noun*	7.95	大阪 オート メッセ (17%)	Osaka Auto Messe
サロン	*noun/noun*	7.76	東京 オート サロン (7%)	Tokyo Auto Salon
[noun/noun] List Top Word 5: *foto* 'photo'				
フレーム	*noun/noun*	9.97	デジタル フォト フレーム (3%)	digital photo frame
ギャラリー	*noun/noun*	9.4	フォト ギャラリー (11%)	photo gallery
ブック	*noun/noun*	9.12	フォト ブック (15%)	photobook
フェイシャル	*noun/noun*	9.08	フォト フェイシャル (16%)	photo facial
コンテスト	*noun/noun*	8.84	フォト コンテスト (10%)	photo contest
ショップ	*noun/noun*	8.69	フォト ショップ (18%)	Photoshop
リーディング	*noun/noun*	7.83	フォト リーディング (17%)	photo reading
ウエディング	*noun/noun*	7.57	フォト ウェディング (9%)	photo wedding
アルバム	*noun/noun*	7.51	フォト アルバム (17%)	photo album
レタッチ	*noun/noun*	7.43	フォト レタッチ (12%)	photo retouch

In Table 6.7 above, 44 of the 50 collocates analysed (88%) are loanword collocates and 2 (4%) are non-loanword collocates, with 4 (8%) being excluded. This leads to an overwhelmingly dominant red-colouring of the text in the table, showing that in the jpTenTen11 corpus, the five loanwords which most prefer to participate in a [noun/noun] gramrel also have a very strong tendency to collocate with other loanwords rather than with native and/or Sino-Japanese words. For each of the five loanwords, the vast majority of their top-10 most salient collocates (salience is shown by the figure in the logDice column) are other loanwords occurring in a compound noun relationship (shown in the Gramrel column), with only two collocates in the entire table being non-loanwords. The Longest Commonest-Matches (hereafter LCMs) show a wide range of behaviours, from brand namings (e.g. 'Hotel New Otani') to concrete objects ('photo album') to concepts (e.g. 'biometrics').

The first loanword in the table, *ando* 'and', shows a fairly unique type of behaviour in that it is a function word in English and function words are rarely borrowed into Japanese (Irwin, 2011). In its Japanese form, however, *ando* is used as a noun and very commonly in fixed three-part English expressions, such as *doraggu ando duroppu* 'drag and drop'. Because of this, it is for the most part unsurprising that it only collocates with other loanwords. *Baio* 'bio' appears to be behaving in Japanese like it does in English: as an attributive adjective prefixed to other words, such as 'biofuel', and there is a hint from this collocation that, unlike *ando*, it can collocate with non-loanwords.

The data in Table 6.7 takes on more meaning when it is compared to words appearing further down in the [noun/noun] list (i.e. loanwords which are less confined to the [noun/noun] gramrel). Tables 6.8 and 6.9 show the data for five loanwords from the middle of the list.

Table 6.8 The middle-five English loanwords from the [noun/noun] list.

	Katakana	Romaji	English	Gramrel 1 Freq.	Gramrel 2 Freq.	Difference
1	テニス	*tenisu*	tennis	116,728	66,019	0.77
2	スター	*sutaa*	star	218,418	124,743	0.75
3	ダイヤモンド	*daiyamondo*	diamond	81,588	46,613	0.75
4	セール	*seeru*	sale	99,935	57,805	0.73
5	コミック	*komikku*	comic	92,509	53,717	0.72

Table 6.9 The most salient collocates and Longest-Commonest Match for the middle-five English loanwords on the [noun/noun] list.

Salient Collocates	Gramrel	logDice	Longest-Commonest Match	English Translation
[noun/noun] List Middle Word 1: *tenisu* 'tennis'				
オウジ	*no pronom*	9.84	テニス の 王子 様 (12%)	The Prince of Tennis
コート	*noun/noun*	9.72	テニス コート (20%)	tennis court
ラケット	*noun/noun*	9.12	テニス ラケット (12%)	tennis racket
スクール	*noun/noun*	8.21	テニス スクール (12%)	tennis school
ウエア	*noun/noun*	7.82	テニス ウェア (7%)	tennis clothing
シューズ	*noun/noun*	7.52	テニス シューズ (9%)	tennis shoes
協会	*noun/noun*	7.52	テニス 協会 (20%)	tennis association
肘	*suffix*	7.42	テニス 肘 (20%)	tennis elbow
プレーヤー	*noun/noun*	7.23	テニス プレーヤー (13%)	tennis player
部	*noun/noun*	7.13	テニス 部 (23%)	tennis club
[noun/noun] List Middle Word 2: *sutaa* 'star'				
ウォーズ	*noun/noun*	9.69	スター ウォーズ (18%)	Star Wars
オーシャン	*noun/noun*	7.7	スター オーシャン (13%)	Star Ocean
アライアンス	*noun/noun*	7.44	スター アライアンス (18%)	Star Alliance
マイン	*noun/noun*	7.34	スター マイン (11%)	Star Mine
ライト	*noun/noun*	7.27	スター ライト (18%)	starlight
トレック	*noun/noun*	7.22	「 スター トレック (22%)	Star Trek
スクリーム	*noun/noun*	7.14	スター スクリーム (17%)	Star Scream
フライアー	*noun/noun*	6.67	スター フライヤー (16%)	Star Flyer
選手	*noun/noun*	6.66	の スター 選手 (5%)	star players
フォックス	*noun/noun*	6.61	スター フォックス (22%)	Star Fox
[noun/noun] List Middle Word 3: *daiyamondo* 'diamond'				
ダスト	*noun/noun*	8.18	ダイヤモンド ダスト (10%)	diamond dust
シティー	*noun/noun*	7.77	ダイヤモンド シティ (20%)	Diamond City
ネックレス	*noun/noun*	7.68	ダイヤモンド ネックレス (9%)	diamond necklace
リング	*noun/noun*	7.52	ダイヤモンド リング (7%)	diamond ring
バックス	*noun/noun*	7.38	ダイヤモンド バックス (11%)	Diamond Backs
パイソン	*noun/noun*	7.13	ダイヤモンド パイソン (10%)	diamond python
ペンダント	*noun/noun*	7.03	ダイヤモンド ペンダント (13%)	diamond pendant
輝き	*no pronom*	6.96	ダイヤモンド の 輝き を (4%)	the radiance of diamonds
ジュエリー	*noun/noun*	6.92	ダイヤモンド ジュエリー (16%)	diamond jewelry
クレバス	*noun/noun*	6.8	ダイアモンド クレバス (12%)	Diamond Crevasse

[noun/noun] List Middle Word 4: *seeru* 'sale'				
SALE	*noun/noun*	8.93	セール SALE (17%)	sale SALE
ストーク	*noun/noun*	8.66	の セール ストーク (5%)	sale stock
特価	*noun/noun*	7.67	セール 特価 (13%)	sale special price
OFF	*noun/noun*	7.35	セール % OFF (23%)	sale % OFF
開催	*noun/noun*	6.92	セール 開催 中 ！ (5%)	sale in progress!
品	*suffix*	6.65	セール 品 (14%)	sale items
激安	*noun/noun*	6.64	SALE セール 激安 (13%)	SALE sale cheap
アウトレット	*noun/noun*	6.6	セール アウトレット (10%)	sale outlet
価格	*noun/noun*	6.56	セール 価格 (23%)	sale price
還元	*noun/noun*	5.92	1トレンド セール 円高 還元 (9%)	trend sale strong yen reduction
[noun/noun] List Middle Word 5: *komikku* 'comic'				
マーケット	*noun/noun*	9.08	コミック マーケット 7 (6%)	Comic Market 7
巻	*noun/noun*	7.34	-	(comic volume)
スピリット	*noun/noun*	7.06	ビッグ コミック スピリッツ (17%)	Big Comic Spirits
アンソロジー	*noun/noun*	6.73	コミック アンソロジー (13%)	comic anthology
シティー	*noun/noun*	6.67	コミック シティ (17%)	Comic City
版	*noun/noun*	6.67	コミック 版 (11%)	comic version
ボンボン	*noun/noun*	6.59	コミック ボンボン (12%)	Comic Bonbon
ブレード	*noun/noun*	6.48	月刊 コミック ブレイド (5%)	Monthly Comic Blade
誌	*suffix*	6.42	コミック 誌 (13%)	comic magazine
エッセー	*noun/noun*	6.06	コミック エッセイ (18%)	comic essay

In this table, 30 of the 50 collocates (60%) are loanword collocates and 15 (30%) are non-loanword collocates, with 5 (10%) being excluded. This leads to the dominant red-colouring of Table 6.7 starting to give way to a larger amount of green colouring, meaning that these loanwords have more of a tendency than the ones at the top of the list to collocate with non-loanwords as well as other loanwords. A few of the collocates are coloured grey because they express international names, such as 'Star Wars'. There is also a tendency for the collocates to be occurring in more diverse grammatical relationships with the loanwords than merely [noun/noun]: In Table 6.7 the Gramrel column consisted 100% of [noun/noun] gramrels, whereas in Table 6.9, five of the collocates were found in other relationships such as [suffix]. Interestingly however, these other relationships are all similar in nature to [noun/noun]. The [*no* pronom] gramrel appearing with the loanword *daiyamondo* 'diamond' for example, represents a possessive noun phrase relationship, as in 'the radiance of'. Functionally, this is very similar to a compound noun relationship. The [suffix] gramrel, similarly, represents a compounding type of behaviour. As such, a comparison of Tables 6.7 and 6.9 shows a greater tendency for the words in the middle of the list to collocate with non-loanwords, but the overall grammatical behaviour across the two sets of loanwords is very similar. Tables 6.10 and 6.11 allow a comparison with the loanwords that have a much weaker overall preference for the [noun/noun] gramrel.

Table 6.10 The bottom-five English loanwords from the [noun/noun] list.

	Katakana	Romaji	English	Gramrel 1 Freq.	Gramrel 2 Freq.	Difference
1	インターネット	*intaanetto*	internet	284,607	278,883	0.021
2	ラジオ	*rajio*	radio	176,407	173,117	0.019
3	トレード	*toreedo*	trade	51,957	51,037	0.018
4	ストップ	*sutoppu*	stop	32,380	31,940	0.014
5	ステーション	*suteeshon*	station	40,693	40,241	0.011

Table 6.11 The most salient collocates and Longest-Commonest Match for the bottom-five English loanwords on the [noun/noun] list.

Salient Collocates	Gramrel	logDice	Longest-Commonest Match	English Translation
[noun/noun] List Bottom Word 1: *intaanetto* 'internet'				
接続	*noun/noun*	9.15	インターネット 接続 (14%)	internet connection
上	*suffix*	7.71	インターネット 上 で (14%)	on the internet
回線	*noun/noun*	7.62	インターネット 回線 (22%)	internet line
経由	*noun/noun*	7.57	インターネット 経由 で (22%)	over the internet
調べる	*de verb*	7.45	インターネット で 調べ て (6%)	examine it on the internet
普及	*no pronom*	7.33	インターネット の 普及 に (15%)	diffusion of the internet
サービス	*noun/noun*	7.31	インターネット サービス (9%)	internet service
カフェ	*noun/noun*	7.3	インターネット カフェ (17%)	internet cafe
バンキング	*noun/noun*	7.29	インターネット バンキング (19%)	internet banking
通ずる	*o verb*	7.28	インターネット を 通じ て (24%)	through the internet
[noun/noun] List Bottom Word 2: *rajio* 'radio'				
体操	*noun/noun*	9.43	ラジオ 体操 (13%)	radio exercises
番組	*noun/noun*	8.8	の ラジオ 番組 (5%)	radio program
局	*noun/noun*	8.4	ラジオ 局 (20%)	radio station
ネーム	*noun/noun*	7.61	ラジオ ネーム (18%)	radio name
放送	*noun/noun*	7.44	ラジオ 放送 (17%)	radio broadcasting
NIKKEI	*noun/noun*	7.36	ラジオ NIKKEI (17%)	Radio NIKKEI
ペンチ	*noun/noun*	7.12	ラジオ ペンチ (23%)	radio pliers
ドラマ	*noun/noun*	7.1	ラジオ ドラマ (15%)	radio drama
ヘッド	*noun/noun*	6.91	レディオ ヘッド の (5%)	radiohead
パーソナリティー	*noun/noun*	6.85	ラジオ パーソナリティ (14%)	radio personality
[noun/noun] List Bottom Word 3: *toreedo* 'trade'				
オフ	*noun/noun*	7.7	の トレード オフ (10%)	trade-off
証券	*noun/noun*	7.33	イー ・ トレード 証券 (4%)	E · Trade Securities
手法	*noun/noun*	6.9	トレード 手法 (22%)	trade method
ショー	*noun/noun*	5.87	トレード ショー (20%)	trade show
要員	*noun/noun*	5.75	トレード 要員 に (5%)	to trade personnel
センター	*noun/noun*	5.5	ワールド トレード センター (9%)	World Trade Center
移籍	*de verb suru*	5.5	トレード で 移籍 し (15%)	transfer with trading
シグナル	*noun/noun*	5.47	トレード シグナル (15%)	trade signal
ツール	*noun/noun*	5.18	トレード ツール (14%)	trade tool
回数	*noun/noun*	5.15	トレード 回数 (22%)	trade frequency
[noun/noun] List Bottom Word 4: *sutoppu* 'stop'				
モーション	*noun/noun*	8.32	ストップ モーション (14%)	stop motion
安	*suffix*	8.12	ストップ 安 (11%)	stop low
ロス	*noun/noun*	7.86	ストップ ロス (24%)	stop loss

ランプ	*noun/noun*	7.36	ハイ マウント ストップ ランプ (5%)	high mount stop lamp
作戦	*noun/noun*	6.62	1 ストップ 作戦 (7%)	1-stop strategy
オーバー	*noun/noun*	6.17	ストップ オーバー (17%)	stopover
高	*suffix*	6.14	ストップ 高 (14%)	stop height
機構	*noun/noun*	5.73	アイドリング ストップ 機構 (11%)	idling stop mechanism
温暖	*N_Adj*	5.54	「 ストップ 温暖 化 (5%)	"stop warming
ボタン	*noun/noun*	5.36	ストップ ボタン を (13%)	stop button
[noun/noun] List Bottom Word 5: *suteeshon* 'station'				
ワゴン	*noun/noun*	8.91	ステーション ワゴン (17%)	Station Wagon
ポータブル	*noun/noun*	6.09	プレイ ステーション ポータブル (8%)	Play Station Portable
シティー	*noun/noun*	6.01	大阪 ステーション シティ (14%)	Osaka Station City
スクエア	*noun/noun*	5.71	新宿 ステーション スクエア (20%)	Shinjuku Station Square
発売	*noun/noun*	5.35	機種 プレイ ステーション 2 発売 日 2 0 (2%)	Model Play Station 2 Release Date 20
ビル	*noun/noun*	5.25	ステーション ビル (11%)	station building
モール	*noun/noun*	5.21	柏 高島 屋 ステーション モール (10%)	Kashiwa Takashimaya store station mall
ストア	*noun/noun*	5.03	プレイ ステーション ストア (22%)	Play Station Store
プラザ	*noun/noun*	5	ステーション プラザ (12%)	Station Plaza
ホテル	*noun/noun*	4.94	ステーション ホテル (10%)	Station Hotel

In Table 6.11, 24 of the 50 collocates analysed (48%) are loanword collocates and 23 (46%) are non-loanword collocates, with 3 (6%) being excluded. The increase of green text in Table 6.11 compared to Tables 6.7 and 6.9 shows that loanwords with a more evenly-balanced frequency between the [noun/noun] gramrel and their second most preferred gramrel have a much stronger tendency than the loanwords higher up on the list to collocate with non-loanwords. There is also the first appearance of gramrels unrelated to the compound noun structure, such as [*de* verb] (which expresses the means of doing something) with the loanword *intaanetto* 'internet' and the LCM of 'examine it on the internet'. Although there are only a few examples of these types of grammatical relationships, they are evidence of these loanwords occurring in a wider, less restricted range of grammatical behaviour patterns. It is interesting to see how this behaviour compares against loanwords on the [particle] list, which is shown in Tables 6.12 to 6.15.

Table 6.12 The top-five English loanwords from the [particle] list.

	Katakana	Romaji	English	Gramrel 1 Freq.	Gramrel 2 Freq.	Difference
1	シーン	*shiin*	scene	712,021	127,837	4.6
2	テンポ	*tenpo*	tempo	84,268	17,113	3.9
3	セクション	*sekushon*	section	33,110	6,764	3.9
4	カップル	*kappuru*	couple	105,742	24,254	3.4
5	コーナー	*koonaa*	corner	712,021	127,837	3.3

Table 6.13 The most salient collocates and Longest-Commonest Match for the top-five English loanwords on the [particle] list.

Salient Collocates	Gramrel	logDice	Longest-Commonest Match	English Translation
[particle] List Top Word 1: *shiin* 'scene'				
様々	*modifier_Ana*	7.1	様々 な シーン で (12%)	in various scenes
活躍	*de verb suru*	6.57	シーン で 活躍 し (18%)	active in the scene
あらゆる	*Adn*	6.47	あらゆる シーン で (18%)	in every scene
彼の	*Adn*	6.18	、 あの シーン (9%)	, that scene
合わせる	*ni verb*	6.08	シーン に 合わせ て (9%)	according to the scene
最後	*pronom no*	5.79	最後 の シーン (24%)	the last scene
別れ	*pronom no*	5.61	の 別れ の シーン (19%)	farewell scene
思い出す	*o verb*	5.59	シーン を 思い出し (22%)	I remember the scene
名	*prefix*	5.53	の 名 シーン (9%)	name scene
どんな	*Adn*	5.51	どんな シーン で も (11%)	in any scene
[particle] List Top Word 2: *tenpo* 'tempo'				
~~ラリー~~	~~*noun/noun*~~	~~6.45~~	~~テンポ ラリ (16%)~~	~~temporary~~
早め	*modifier_Ano*	6.44	速め の テンポ で (7%)	with a fast tempo
遅め	*modifier_Ano*	6.27	遅め の テンポ で (13%)	with a slow tempo[19]
早める	*o verb*	5.74	テンポ を 速め て (4%)	increase the tempo
落とす	*o verb*	5.61	テンポ を 落とし て (5%)	lower the tempo
弾く	*de verb*	5.2	-	(play to the tempo)
スロー	*modifier_Ana*	5.17	スロー な テンポ で (8%)	with a slow tempo
曲	*no pronom*	5.08	アップ テンポ の 曲 (7%)	up tempo song
早い	*modifier_Ai*	5.07	速い テンポ で (6%)	with a fast tempo
会話	*pronom no*	4.99	会話 の テンポ が (6%)	the tempo of a conversation
[particle] List Top Word 3: *sekushon* 'section'				
各	*prefix*	5.51	各 セクション の (7%)	of each section
区切り	*noun/noun*	5.39	セクション 区切り を (6%)	section break
分かれる	*ni verb*	5.32	セクション に 分かれ て (17%)	divided into sections
ヘッダー	*noun/noun*	4.86	セクション ヘッダー (22%)	section header
毎	*suffix*	4.63	セクション ごと に (11%)	by section
参照	*o verb suru*	4.36	の セクション を 参照 し て(9%)	see section
分ける	*ni verb*	4.27	セクション に 分け (19%)	divide into sections
分け	*noun/noun*	4.14	セクション 分け (18%)	section division
従える	*o verb*	4.08	セクション を 従え た (15%)	followed the section
設ける	*o verb*	3.77	セクション を 設け て (8%)	specify a section

[19] Comparing this LCM with one 4 lines below, they both have the same English translation ('with a slow tempo'). However, the green text shows the collocate is the non-loanword 遅め *osome* 'slow' and the red text shows the collocate is the English loanword スロー *suroo* 'slow'. This interesting behaviour will be returned to in Chapter Seven.

[particle] List Top Word 4: *kappuru* 'couple'				
成立	*noun/noun*	6.92	カップル 成立 (21%)	couple formation
似合い	*pronom no*	6.86	お 似合い の カップル (16%)	suitable couple
若い	*modifier_Ai*	6.67	若い カップル が (6%)	young couple
喫茶	*noun/noun*	6.63	カップル 喫茶 (14%)	couples' cafe
組	*pronom no*	6.54	組 の カップル が (8%)	a couple of groups
ゲイ	*pronom no*	5.76	ゲイ の カップル (22%)	gay couples
プラン	*noun/noun*	5.68	カップル プラン (7%)	coupled plan
~~だらけ~~	~~*suffix*~~	~~5.6~~	~~カップル だらけ (21%)~~	~~lots of couples~~
~~だらけ~~	~~*N_Adj*~~	~~5.6~~	~~カップル だらけ (21%)~~	~~lots of couples~~
男女	*pronom no*	5.13	男女 の カップル が (5%)	a couple of men and women
[particle] List Top Word 5: *koonaa* 'corner'				
キック	*noun/noun*	7.99	コーナー キック から (7%)	from the corner kick
設ける	*o verb*	6.98	コーナー を 設け て (5%)	specify a corner
リング	*noun/noun*	6.55	コーナー リング (20%)	corner ring
ソファー	*noun/noun*	6.31	コーナー ソファ (7%)	corner sofa
曲がる	*o verb*	5.85	コーナー を 曲がる (13%)	turn the corner
設ける	*ga verb*	5.67	コーナー が 設け られ て (19%)	a corner was specified
手前	*noun/noun*	5.59	コーナー 手前 で (6%)	before the corner
回る	*o verb*	5.54	コーナー を 回っ (24%)	round the corner
進入	*noun/noun*	5.53	コーナー 進入 時 (11%)	when entering a corner
にて	*particle*	5.41	コーナー にて (23%)	in the corner

In this first table of collocates from the [particle] list, the balance of loanword to non-loanword collocates has changed with there now being a majority of non-loanword collocates (i.e., more green than red text). 40 of the 50 collocates (80%) are non-loanword collocates and 7 (14%) are loanword collocates, with 3 (6%) being excluded. This shows that these loanwords are much less restricted in their collocational behaviour than the loanwords from the top of the [noun/noun] list, in that they have a greater propensity to collocate with non-loanwords. Furthermore, there is a much greater degree of grammatical flexibility in how the loanwords occur with their collocates: there are a wide range of different gramrels and only a few instances of the [noun/noun] gramrel. This is expressed in the length of the LCMs which give a much more insightful view into the realisations of the collocations in the corpus. For example, it can be seen that the loanword *tenpo* 'tempo' occurs in phrases such as 'with a slow tempo' and 'increase the tempo'. Overall, the behaviour of the data in this table is almost the exact opposite of the data in Table 6.7, expressing the very different natures of the [noun/noun] and [particle] gramrels. Tables 6.14 and 6.15 explore how the behaviour develops with the loanwords in the middle of the [particle] list.

Table 6.14 The middle-five English loanwords from the [particle] list.

	Katakana	Romaji	English	Gramrel 1 Freq.	Gramrel 2 Freq.	Difference
1	ヒント	*hinto*	hint	78,750	43,206	0.823
2	マスク	*masuku*	mask	93,604	51,494	0.823
3	カーテン	*kaaten*	curtain	71,134	39,253	0.818
4	プロジェクト	*purojyekuto*	project	198,800	109,938	0.812
5	スケール	*sukeeru*	scale	58,975	32,637	0.808

Table 6.15 The most salient collocates and Longest-Commonest Match for the middle-five English loanwords on the [particle] list.

Salient Collocates	Gramrel	logDice	Longest-Commonest Match	English Translation
[particle] List Middle Word 1: *hinto* 'hint'				
得る	*o verb*	6.69	ヒント を 得 て (17%)	get a hint
与える	*o verb*	6.18	ヒント を 与え て (17%)	give a hint
隠す	*ga verb*	5.71	ヒント が 隠さ れ て いる (21%)	hints are hidden
鏤める	*ga verb*	5.41	ヒント が 散りばめ られ て (13%)	scattered with hints
探る	*o verb*	5.38	の ヒント を 探る (9%)	explore hints
見付かる	*ga verb*	5.19	ヒント が 見つかる (16%)	hints are found
掴む	*o verb*	4.97	-	(hold a mask)
満載	*noun/noun*	4.97	の ヒント 満載 (8%)	full of hints
集	*suffix*	4.95	の ヒント 集 (9%)	hints collection
解決	*pronom no*	4.84	解決 の ヒント を (8%)	a hint of the solution
[particle] List Middle Word 2: *masuku* 'mask'				
着用	*noun/noun*	7.65	マスク 着用 (14%)	mask wearing
被る	*o verb*	7.61	マスク を 被っ (16%)	cover with a mask
着用	*o verb suru*	6.94	マスク を 着用 し て (7%)	wearing a mask
外す	*o verb*	6.23	マスク を 外し (19%)	remove a mask
マン	*noun/noun*	6.14	マスク マン (12%)	Maskman
マスク	*noun/noun*	5.92	-	(mask mask)
着用	*no pronom*	5.87	、 マスク の 着用 (5%)	, wearing a mask
甘い	*modifier_Ai*	5.85	甘い マスク (18%)	sweet mask (beautiful)
越し	*suffix*	5.73	マスク 越し に (10%)	through a mask
装着	*o verb suru*	5.69	マスク を 装着 し て (3%)	wear a mask
[particle] List Middle Word 3: *kaaten* 'curtain'				
コール	*noun/noun*	10.08	カーテン コール で (7%)	with a curtain call
レール	*noun/noun*	9.49	カーテン レール (15%)	curtain rail
越し	*suffix*	8.06	カーテン 越し に (12%)	over the curtain
締める	*o verb*	8.01	カーテン を 閉め て (4%)	close the curtain
仕切る	*de verb*	7.8	カーテン で 仕切ら れ た (5%)	it was partitioned with a curtain
ウォール	*noun/noun*	7.52	カーテン ウォール (16%)	curtain wall
隙間	*no pronom*	7.52	カーテン の 隙間 から (19%)	from a gap in the curtain
緑	*pronom no*	7.5	緑 の カーテン (15%)	green curtain
開ける	*o verb*	7.44	カーテン を 開け て (4%)	open the curtain
レース	*pronom no*	6.44	レース の カーテン (20%)	curtain of lace

[particle] List Middle Word 4: *purojyekuto* 'project'				
マネージメント	*noun/noun*	8.32	プロジェクト マネジメント (17%)	project management
マネージャー	*noun/noun*	8.19	プロジェクト マネージャー (8%)	project manager
チーム	*noun/noun*	7.36	プロジェクト チーム (18%)	project team
リーダー	*noun/noun*	7.12	プロジェクト リーダー (12%)	project leader
立ち上げる	*o verb*	7.1	プロジェクト を 立ち上げ (19%)	launch the project
管理	*noun/noun*	6.35	プロジェクト 管理 (20%)	project management
一環	*no pronom*	6.25	プロジェクト の 一環 と し て (4%)	as part of the project
進める	*o verb*	6.23	プロジェクト を 進め て (13%)	advance the project
一大	*prefix*	6.07	一大 プロジェクト (21%)	one big project
携わる	*ni verb*	5.83	プロジェクト に 携わっ て (7%)	working on the project
[particle] List Middle Word 5: *sukeeru* 'scale'				
壮大	*modifier_Ana*	9.09	壮大 な スケール で (10%)	on a grand scale
メリット	*noun/noun*	7.56	スケール メリット を (9%)	scale merit
ダウン	*noun/noun*	6.3	スケール ダウン し (24%)	scale down
アウト	*noun/noun*	6.21	スケール アウト (18%)	scale out
アップ	*noun/noun*	6.1	スケール アップ し (21%)	scale up
フィギュア	*noun/noun*	6.07	6 スケール フィギュア (23%)	6-scale figure
感	*noun/noun*	6.01	スケール 感 (18%)	scale sensitivity
雄大	*modifier_Ana*	5.81	雄大 な スケール (24%)	majestic scale
モデル	*noun/noun*	5.64	スケール モデル (12%)	scale model
空前	*pronom no*	5.29	空前 の スケール で (14%)	on an unprecedented scale

The data in Table 6.15 shows a continuation of the increase in non-loanword collocates, with 34 of the 50 collocates (68%) being non-loanword collocates and 16 (32%) being loanword collocates (with none excluded). It is clear that just as with the loanwords from the [noun/noun] list, there is gradation of behaviour with the loanwords from different parts of the [particle] list. This is shown by the gradual decrease in green text and increase in red text. That means that some of the loanwords, such as *sukeeru* 'scale', are showing a greater tendency to collocate with loanwords than what was the case in Table 6.13. This can largely be expected because the degree of difference between the [particle] gramrel and the second most preferred gramrel of these loanwords is less than with the loanwords in Table 6.13, meaning that there is a greater tendency for these loanwords to occur frequently in a [noun/noun] gramrel. This is also shown in the Gramrel column by the quite large number of [noun/noun] gramrels, and also in the LCM column by the fact that they are overall less revealing than in Table 6.13, such as 'scale up/down/out'. The two tables below present the last set of collocates in this analysis, which are for the loanwords with an almost even frequency between the [particle] gramrel and their second most preferred gramrel.

Table 6.16 The bottom-five English loanwords from the [particle] list.

	Katakana	Romaji	English	Gramrel 1 Freq.	Gramrel 2 Freq.	Difference
1	シャッター	*shattaa*	shutter	44,218	43,566	0.015
2	ストック	*sutokku*	stock	30,558	30,141	0.014
3	パニック	*panikku*	panic	44,254	43,745	0.012
4	ライオン	*raion*	lion	45,778	45,634	0.003
5	バッグ	*baggu*	bag	218,958	218,399	0.003

Table 6.17 The most salient collocates and Longest-Commonest Match for the bottom-five English loanwords on the [particle] list.

Salient Collocates	Gramrel	logDice	Longest-Commonest Match	English Translation
[particle] List Bottom Word 1: *shattaa* 'shutter'				
スピード	*noun/noun*	9.02	シャッター スピード (21%)	shutter speed
チャンス	*noun/noun*	8.50	シャッター チャンス を (8%)	shutter opportunity
速度	*noun/noun*	8.40	シャッター 速度 (19%)	shutter speed
押す	*o verb*	8.16	シャッター を 押し (21%)	press the shutter
締まる	*ga verb*	7.56	シャッター が 閉まっ て (17%)	the shutter is closed
切る	*o verb*	7.41	シャッター を 切っ (17%)	turn off the shutter
アイランド	*noun/noun*	7.30	「 シャッター アイランド 」 (3%)	"Shutter Island"
ボタン	*noun/noun*	7.11	シャッター ボタン を (12%)	shutter button
下ろす	*o verb*	6.65	シャッター を 下ろし (21%)	lower the shutter
押し	*noun/noun*	6.40	シャッター 半 押し で (7%)	by pressing the shutter halfway
[particle] List Bottom Word 2: *sutokku* 'stock'				
リサーチ	*noun/noun*	8.25	ストック リサーチ - ドイツ フランクフルト (0%)	Stock Research - Germany Frankfurt
抄録	*noun/noun*	8.02	流通 : ストック ブックス 抄録 : (6%)	distribution: stock books abstract:
オプション	*noun/noun*	7.83	ストック オプション (19%)	stock options
ブック	*noun/noun*	7.64	流通 : ストック ブックス 抄録 : (2%)	distribution: stock books abstract:
フォト	*noun/noun*	7.53	ストック フォト (17%)	stock photo
~~ウェル~~	~~*noun/noun*~~	~~5.71~~	~~ディーン・ストック ウェル (6%)~~	~~dean stockwell~~
ウォーキング	*noun/noun*	4.77	ノルディック ・ ストック ウォーキング (8%)	nordic stock walking
カー	*noun/noun*	4.56	ストック カー レース (6%)	stock car race
無くなる	*ga verb*	4.39	ストック が なくなっ (17%)	stock is gone
溜まる	*ga verb*	4.37	-	-
[particle] List Bottom Word 3: *panikku* 'panic'				
発作	*noun/noun*	9.53	パニック 発作 (18%)	panic attack
障害	*noun/noun*	9.15	パニック 障害 (17%)	panic disorder
陥る	*ni verb*	9.06	パニック に 陥っ (21%)	fall into panic
起こす	*o verb*	7.38	パニック を 起こし (24%)	cause panic
症候	*noun/noun*	6.7	パニック 症候 群 (13%)	panic syndrome group
陥れる	*ni verb*	6.62	を パニック に 陥れ (16%)	I panicked
状態	*noun/noun*	6.46	パニック 状態 に (10%)	panic
寸前	*noun/noun*	6.03	パニック 寸前 (19%)	on the verge of panic
引き起こす	*o verb*	5.75	パニック を 引き起こし (14%)	cause panic
フェース	*noun/noun*	5.56	パニック フェイス (14%)	panic face

[particle] List Bottom Word 4: *raion* 'lion'				
キング	*noun/noun*	8.36	ライオン キング (19%)	Lion King
ハート	*noun/noun*	7.14	ライオン ハート (7%)	Lion Heart
クラブ	*noun/noun*	7.11	ライオンズ クラブ (19%)	Lions Club
マンション	*noun/noun*	6.27	ライオンズ マンション (11%)	Lions Mansion
ナイター	*noun/noun*	6.25	文化 放送 ライオンズ ナイター (5%)	Cultural Broadcast Lions Night
鬣	*no pronom*	6	ライオン の たてがみ (16%)	lion mane
魔女	*coord*	5.95	ライオン と 魔女 」 (24%)	Lion and the Witch"
丸	*suffix*	5.75	ライオン 丸 (13%)	Lion Maru
虎	*coord*	5.38	ライオン や トラ (13%)	lions and tigers
檻	*no pronom*	5.06	ライオン の 檻 に (6%)	in the cage of the lion
[particle] List Bottom Word 5: *baggu* 'bag'				
バッグ	*noun/noun*	8.47	-	(bag bag)
ショルダー	*noun/noun*	7.7	バッグ ショルダー バッグ (5%)	bag shoulder bag
トート	*noun/noun*	7.26	バッグ トート バッグ (4%)	bag tote bag
レディー	*noun/noun*	6.63	バッグ レディース (12%)	bag ladies
チャーム	*noun/noun*	6.6	バッグ チャーム (13%)	bag charm
鞄	*noun/noun*	6.4	-	(bag leather)
斜め	*noun/noun*	6.32	バッグ 斜め がけ (14%)	bag diagonal cliff
COACH	*noun/noun*	6.32	コーチ バッグ COACH (7%)	COACH bag COACH
取り出す	*kara verb*	6.23	-	-
バック	*noun/noun*	6.16	バッグ バック (9%)	bag back

This last table of collocation data shows that 26 of the 50 collocates analysed (52%) are non-loanword collocates and 20 (40%) are loanword collocates, with 4 (8%) being excluded. This data therefore exhibits a very similar general pattern of behaviour to the loanwords from the bottom of the [noun/noun] list, in that there is a further variation in their overall collocational behaviour. Here there is an increase of red text (i.e. loanword collocates) compared to Table 6.15, along with an increase in the [noun/noun] gramrels, meaning that such loanwords exhibit a more restricted range of grammatical behaviour. This is again shown by the shorter LCMs compared to Tables 6.15 and 6.13, with many of them restricted to two-word compounds, such as 'bag charm' and 'stock photo'. Overall, however, there is a general balance between the amount of red and green text, representing a balance between loanword and non-loanword collocations, and also the amount of [noun/noun] gramrels compared with other relationships, such as [*o* verb] and [*ga* verb]. To get a clearer idea of the general trends which have been described here and for all the tables of collocates above, Table 6.18 below summarises the most important features of all six tables of collocates.

Table 6.18 The collocate-types of the most salient collocates of the 30 loanwords.

		Type of Collocate						*Total*
		Loanword	*Total*	**Non-Loanword**	*Total*	Excluded	*Total*	
[noun/noun] List	**Top-5**	44	*98*	2	*40*	4	*12*	50
	Middle-5	30		15		5		50
	Bottom-5	24		23		3		50
[particle] List	**Bottom-5**	20	*43*	26	*100*	4	*7*	50
	Middle-5	16		34		0		50
	Top-5	7		40		3		50
	Total		*142*		*139*		*19*	*300*

Table 6.18 shows the overall data of the 300 collocates for the sample of 30 loanwords described in the analysis above. This covers the top-10 most salient collocates of the three sets of five loanwords from the top, middle, and bottom of the [noun/noun] and [particle] lists. To make the table more readable, the order of the three sets of collocates of the [particle] list have been reordered so as to begin with the bottom-five loanwords on the list. This better represents the gradual variation in frequencies of the loanwords: from the loanwords with the highest degree of difference between the [noun/noun] gramrel and the second most preferred gramrel at the top of the table, through to the loanwords with the highest degree of difference between the [particle] gramrel and the second most preferred gramrel at the bottom of the table. Ordering the table in this way shows a regular pattern of a simultaneous decrease in the number of loanword collocates and increase in the number of non-loanword collocates as one scans down the table from the loanwords most strongly preferring the [noun/noun] gramrel to the loanwords most strongly preferring the [particle] gramrel.

The main conclusion to read from this table is that there is an almost perfect matching of the two-way collocational behaviour of the loanwords with their two-way grammatical behaviour. This can be stated as follows: the greater the extent to which a loanword prefers the [noun/noun] grammatical relationship, the greater the tendency for the loanword to collocate with other loanwords than native and Sino-Japanese words; and thus conversely, the greater the extent to which a loanword prefers the [particle] grammatical relationship, the greater the tendency for the loanword to collocate with

native and Sino-Japanese words than other loanwords. The almost perfect match of collocational behaviour with grammatical behaviour is shown by the total number of collocate types matching with each grammatical behaviour type. For the 15 words on the [noun/noun] list, 71.0% (98 of 138) of their most salient collocates are other loanwords (12 collocation candidates were excluded). In an almost perfect mirror image of this, 69.9% (100 of 143) of the most salient collocates of the 15 words on the [particle] list are non-loanwords (with 7 collocates excluded).

Having uncovered this behavioural information, the question remains as to what can account for this different grammatical and collocational behaviour. Returning to Onysko and Winter-Froemel's (2011) theory of catachrestic and non-catachrestic innovations in a language discussed above, it does seem to be the case that these categories match well with the findings in this chapter and the previous one. Onysko and Winter-Froemel (2011) classify catachrestic innovations as the unmarked form of expression conveying I-implicatures (informativeness): they are words which convey the normal way of communicating about something and as such are interpreted in a stereotypical way (p. 1555). They classify non-catachrestic innovations as the marked form of expression conveying M-implicatures (manner): they are words which convey pragmatic markedness and as such are interpreted in a non-stereotypical way (p. 1555). The behaviour of the English loanwords in the present study which most frequently participate in the [particle] gramrel seems well accounted for by this theoretical catagorisation of a catachrestic innovation. The analyses in Sections 6.4 and 6.5 above showed these loanwords overall to be collocating more strongly with native and Sino-Japanese words, which are the more frequent normal way of communicating in Japanese. Furthermore, their Longest Commonest Matches in the tables of collocates shows them to be participating in a very wide range of grammatical relationships, from being the subject and object of a sentence, to a part of noun phrases and compounds, which was shown in the analysis in Chapter Five of the non-loanwords to be the normal, unmarked range of behaviour of nouns in a language.

Similarly, the behaviour of the English loanwords which most frequently participate in the [noun/noun] gramrel seems well accounted for by this theoretical categorisation of a non-catachrestic innovation. Sections 6.4 and 6.5 showed these loanwords overall to be collocating more strongly with other loanwords, and they were shown to participate in a restricted range of grammatical relationships, in some cases almost exclusively restricted to the attributive (i.e. left-hand) part of a compound noun. Comparing this behaviour with that of the non-loanwords in Chapter Five, in which

only eight of the 130 non-loanwords participated most frequently in this relationship, this behavior of the loanwords represents a marked style of communication.

As to the methodology of categorising loanwords as catachrestic or non-catachrestic innovations, discussed in Onysko and Winter-Froemel's (2011) work, it was discussed above how this categorisation is based on the notion of an equivalent expression in the native language. More precisely, they state in their article that the distinction between these two types is "crucially dependent on an answer to the question of whether an innovation introduces a new concept into a language or whether the concept is already expressed in the language by a semantic near-equivalent" (2011, p. 1555). Therefore, the next section will focus on analysing which of the 30 English loanwords in the sample used in this chapter have an equivalent expression in the native and/or Sino-Japanese vocabulary.

6.6 Categorising English Loanwords as Catachrestic or Non-Catachrestic

It was discussed in Chapter Three and also in Section 6.3.1 of this chapter that the traditional method of establishing whether or not a loanword is a non-catachrestic innovation (luxury loan/special-effect-giver etc.) or a catachrestic innovation (necessary loan/lexical-gap filler etc.) is to use a dictionary to look for an equivalent expression. Table 6.19 gives data on an analysis conducted in this way on the 30 loanwords used in the collocational analysis in Section 6.4 above. Two Japanese dictionaries were used in the analysis, the Daijirin general Japanese dictionary (3rd ed.), and Sanseido's Concise Dictionary of Katakana Words (4th ed.). These are the same dictionaries used in the creation of the English loanword list described in Chapter Four. The entries for each word were checked in both dictionaries and if they had equivalent expressions listed, they were recorded in the table. For some of the loanwords, many equivalent expressions were listed so just a sample of these has been recorded in the table. Also, because the main aim of Table 6.19 is to show whether or not the dictionaries list equivalent expressions for the loanwords, transliterations in English alphabet have not been given and the words remain in their original *kanji* and/or *hiragana* script.

Table 6.19 A dictionary analysis of equivalent expressions of the 30 loanwords.

		Katakana	Romanised	English	Daijirin	Concise
[noun/noun] List	**Top-5**	アンド	*ando*	and	および そして… …と	そして… …と
		バイオ	*baio*	bio	生命 生物	生命の 生物の
		ニュー	*nyuu*	new	新しい	新品 新しい
		オート	*ooto*	auto	自動の	X
		フォト	*foto*	photo	写真 映画 光の	写真
	Middle-5	テニス	*tenisu*	tennis	庭球	庭球
		スター	*sutaa*	star	星 花形	星, 恒星 花形 人気者
		ダイヤモンド	*daiyamondo*	diamond	金剛石	金剛石
		セール	*seeru*	sale	売り出し	販売 大売り出し
		コミック	*komikku*	comic	漫画 劇画	X
	Bottom-5	インターネット	*intaanetto*	internet	X	X
		ラジオ	*rajio*	radio	受信装置 放射	受信機
		トレード	*toreedo*	trade	取引 貿易 商売	商業 取引 貿易
		ストップ	*sutoppu*	stop	とまること とめること やめること	止まること 止めること 停止
		ステーション	*suteeshon*	station	駅 停車場 詰め所	駅 停車場 放送局
[particle] List	**Top-5**	シーン	*shiin*	scene	場面 情景 光	光景 情景 場面
		テンポ	*tenpo*	tempo	音楽速度	拍子 楽曲演奏 進行速度 調子
		セクション	*sekushon*	section	部分 仕切り 節 項	区画 仕切り 部分 節 項
		カップル	*kappuru*	couple	夫婦 恋人同士	男女 1 組 夫婦 恋人同士
		コーナー	*koonaa*	corner	隅	曲がり 角
	Middle-5	ヒント	*hinto*	hint	暗示 示唆	暗示 手がかり 間接的示唆
		マスク	*masuku*	mask	面 仮面	面 仮面
		カーテン	*kaaten*	curtain	幕	窓かけ
		プロジェクト	*purojyekuto*	project	事業	投射 企画 計画.

		スケール	*sukeeru*	scale	規模 はかり	目盛り 尺度 縮尺 規模 程度 大きさ
	Bottom-5	シャッター	*shattaa*	shutter	露光装置 よろい戸	よろい戸 露光装置
		ストック	*sutokku*	stock	蓄えた物 在庫品 株券	在庫品 手持ち品 株券
		パニック	*panikku*	panic	不安 驚き 恐怖 恐慌 経済恐慌	恐怖 経済恐慌
		ライオン	*raion*	lion	獅子	獅子
		バッグ	*baggu*	bag	袋 かばん	袋 カバン 手さげ.

If Onysko and Winter-Froemel's (2011) categories of catachrestic and non-catachrestic innovations are to be used to explain the two main types of behaviour of the loanwords in the present study, of those most preferring the [particle] gramrel and those most preferring the [noun/noun] gramrel, then in principle around half of the loanwords in Table 6.19 should have identifiable equivalent expressions and the other half should not. However, the table reveals an interesting finding in this regard. The grey underlined text of *intaanetto* 'internet' highlights that this is the only word which does not have any equivalent expressions listed in either dictionary.[20] Furthermore, there are only two loanwords, *ooto* 'auto' and *komikku* 'comic', which have an equivalent expression listed in only one of the two dictionaries. As such, following Onysko and Winter-Froemel's criteria of classification, on the basis of this dictionary data 29 out of the 30-word sample (96.7%) should be classified as non-catachrestic innovations, and only one out of the 30-word sample (3.3%) should be categorised as a catachrestic innovation. Subsequently, it would appear that the two different types of grammatical behaviour seen in this 30-word sample cannot sufficiently be accounted for by Onysko and Winter-Froemel's two categories of catachrestic and non-catachrestic innovations. Whilst this analysis was only conducted on a sample of 30 loanwords, this conclusion could likely be applied to all the 587 English loanwords, although a larger sample of the loanwords would be needed to confirm this.

Onysko and Winter-Froemel (2011), however, discuss an important qualification of their theory in that a confident categorisation of non-catachrestic and catachrestic

[20] This loanword also appears in the list of 101 Anglicisms in Onysko and Winter-Froemel's (2011) study and is similarly classified in their study as a catachrestic innovation in the German language.

innovations can only be made after judging the appropriateness of the equivalent expressions listed in standard lexical resources such as general or technical dictionaries. This, they suggest, can be done with evidence from usage-based lexical resources such as corpora. They explain how looking for the appropriateness of the equivalent expressions in this way allows the discounting of possible equivalent expressions which "do not sufficiently correspond to the meaning of the anglicism or they are not used as general terms in German" (2011, p. 1557). They do not, however, give specific details of how they qualified or quantified the appropriateness of the equivalent expressions, or how many of the initial equivalent expressions they ended up discounting, beyond the general statement that "we ran searches in the corpora and on German websites to determine the usage of the equivalent German terms or paraphrases" (2011, p. 1557).

In any case, limiting the categorisation to a simple yes/no decision of whether an equivalent term is found in a dictionary or other similar lexical resource of the language brings with it a range of problems. Onysko and Winter-Froemel (2011) acknowledge this in their statement that "while these results confirm that a basic categorization of anglicisms as catachrestic and non-catachrestic innovations is possible, the detailed discussion of selected borrowings stresses the fact that taking decisions on the general pragmatic function of an anglicism is frequently a complex task" (2011, p. 1563). One of the main problems is that the notion of an equivalent expression is a fluid rather than static concept. What was counted as an appropriate equivalent expression in the past may well change over time, if, for example, the loanword gradually becomes the unmarked form of expression or conversely if its usage falls out of favour (Inagawa, 2012) and the equivalent expression comes back into standard usage. Furthermore, decisions as to the appropriateness of an equivalent expression are dependent on the language resources which are consulted, and to properly investigate a range of resources to ensure the appropriateness of one equivalent expression, when a loanword may indeed have many potential equivalent expressions, is a labour-intensive task.

Not only this, but the findings from Table 6.19 and the previous tables of collocates in Section 6.4 show there to be another concern with the idea of basing the categorisation of loanwords as catachrestic or non-catachrestic innovations on the criterion of an appropriate equivalent expression. This concern is that it only allows a binary categorisation of the loanword as either catachrestic or non-catachrestic. This either/or categorisation then necessarily negates the behavioural differences seen with the three sets of loanwords from different parts of the lists in Section 6.4. Indeed, Onysko and Winter-Froemel themselves acknowledge the fact that "the categorization

as catachrestic and non- catachrestic is generally not a strictly categorical either/or decision" (2011, p. 1563).

It has already been discussed that Onysko and Winter-Froemel's (2011) method of categorising a loanword either as catachrestic or as non-catachrestic on the basis of the binary notion of a presence or absence of an equivalent expression is not new but rather a continuation of a long historical tradition in lexical borrowing research. The academic studies of loanwords stretching back throughout the history shown in Table 6.5 have continually stressed the fundamental importance of this presence/absence of what has also been termed a "parallel expression" for the loanword (Matras, 2009, p. 150). This methodology has been convenient for researchers interested in the topic of the functions of loanwords because it fits neatly with the dualistic theoretical frameworks which they have developed around the loanwords. For example, Matras (2009) comments on how Myers-Scotton's (2006) dualistic terminology of cultural and core loans is borne out of her approach of wanting to separately categorise borrowings and codeswitches; something she does, he writes, to "satisfy the internal mechanics of her Matrix Language Frame model" (2009, p. 110). This model is a popular one in studies of bilingualism where a binary distinction is made between the lexemes of a Matrix Language (i.e. the dominant language of the speaker) and the Embedded Language, (the language which contributes the new vocabulary to the Matrix Language).

To give an example of this problem using data from the present study, the loanword *shiin* 'scene' which is at the top of the [particle] list in Table 6.12 has several potential equivalent expressions listed in the two dictionaries, including 場面 *bamen*, 情景 *jyookei*, and 光景 *kookei*. To check the appropriateness of these three potential equivalent expressions, a search in a Japanese-English dictionary shows all three to be defined in English as 'scene', and a check of their frequencies in the jpTenTen11 corpus shows them all to be relatively frequent words in Japanese, with *bamen* having 41.7 occurrences per million, *jyookei* having 3.7 per million, and *kookei* with 17.1 per million. From this data, the loanword *shiin* 'scene' can be judged to have appropriate equivalent expressions and therefore would be categorised in Onysko and Winter-Froemel's framework (2011) as a non-catachrestic loanword functioning in the language as the marked lexical choice conveying pragmatic connotations.

However, empirical data of the grammatical and collocational behaviour of *shiin* suggests that the categorisation is more complicated than that given above. First of all, it has a frequency in the jpTenTen11 corpus of 103.8 per million, which is more than twice that of *bamen* at 41.7. Table 6.13 also shows that its collocates are all non-

loanwords (coloured green in the table) and it participates in a wide variety of grammatical roles in the language, such as being the object of a sentence ('I remember the scene'), and being modified by adjectives ('various scenes') and noun adjuncts ('in every scene', 'that scene'). Furthermore, a check for the loanword *shiin* in the Thesaurus function of the Sketch Engine returns the output shown in Figure 6.3 below. This function of the Sketch Engine returns synonyms of the search word which tend to occur in similar grammatical and collocational contexts (Kilgarriff et al., 2014). The word listed as most similar to *shiin* is 場面 *bamen*, which was introduced above. This is a very important finding because it is empirical evidence from a large corpus of naturally-occurring language data that the loanword *shiin* is used in very similar grammatical and collocational contexts to the native equivalent expression *bamen*. Some of the other synonyms in Figure 6.3 give more details of the behaviour of this loanword. A group of them appear related to movies, such as 映像 *eizoo* 'moving image', 作品 *sakuhin* 'production', 映画 *eiga* 'film', and 台詞 *daishi* 'script'. A few others seem more related to scenery, such as 写真 *shashin* 'photo' and 絵 *e* 'picture'.


シーン Japanese Web 201

Lemma	Score	Freq
場面	0.548	430,040
映像	0.441	875,348
作品	0.385	2,334,948
曲	0.384	2,539,850
映画	0.370	2,144,941
ストーリー	0.363	516,662
写真	0.355	3,065,781
話	0.354	5,596,449
台詞	0.352	425,678
絵	0.350	1,169,469

Figure 6.3 Synonyms of the English loanword *shiin* 'scene'.

Therefore, from an analysis of its naturally-occurring grammatical behaviour in a corpus of the Japanese language, *shiin* seems to be behaving more as a catachrestic than non-catachrestic innovation. Not only is it far more frequent in the corpus than some of its appropriate equivalent expressions, but it collocates frequently with words from all of the Japanese lexical strata, participates in a range of grammatical roles, and appears to have a variety of meanings, such as 'scene of a movie' and 'scene in a picture'. Also, the fact that it participates most frequently in the [particle] gramrel, and only infrequently in the [noun/noun] gramrel with other loanwords (a check of its word

sketch showed [noun/noun] to be its 11th preferred gramrel), is further evidence of it being more of an unmarked rather than marked lexical choice. This empirical data accordingly allows a categorisation of *shiin* 'scene' to be made on the basis of how it is used in the language, rather than the largely abstract notion of whether an equivalent expression can be identified, and furthermore, this categorisation can be one of degree rather than being absolute, in that an examination of the grammatical behaviour of a loanword can show to what degree it is found to be preferring certain grammatical structures over others. As such, the categorisation can be made along the lines of more or less catachrestic or non-catachrestic, rather than simply one or the other.

In the discussion above of how the present study has shown that a categorisation of a loanword as a catachrestic or non-catachrestic innovation is better made on the basis of observed linguistic behaviour than on the presence or absence of an equivalent expression, an important methodological issue is raised. This is that a categorisation should be made on a loanword-by-loanword basis, by a careful examination of different aspects of the loanword's behaviour. Overall, the findings from this chapter and the previous one have shown that whilst the observed differences in the grammatical behaviour of the English loanwords is generally suggestive of different usage associations in the Japanese language, such as being marked lexical choices conveying pragmatic meanings (including conveying images of modernity and being used as euphemistic terms), the loanwords each display individual behavioural characteristics which disallow their orderly placement into defined categories. Whilst Onysko and Winter-Froemel (2011) acknowledge this fact in their statement that "an actual classification is a complex task that depends on contextual evidence, usage frequencies, and close interpretation of the data" (p. 1563), they believe that it is only necessary in a small number of problematic cases. The finding of the present study, however, is that such a close interpretation of the data is best carried out for each individual loanword.

6.7 Chapter Summary

This chapter has analysed the collocational behaviour of a 30-word sub-sample of the English loanwords examined in Chapter Five. In line with a similar analysis carried out by Bordilovskaya (2012, 2016), the aim of this investigation was to explore if the collocational behaviour of the loanwords could help in further understanding the patterns of grammatical behaviour uncovered in Chapter Five. An analysis of the most salient collocates of a sample of 15 loanwords from the [noun/noun] list and from the

[particle] list found a clear pattern of collocational behaviour: the greater the extent to which a loanword prefers the [noun/noun] grammatical relationship, the greater the tendency for the loanword to collocate with other loanwords than native and Sino-Japanese words; and conversely, the greater the extent to which a loanword prefers the [particle] grammatical relationship, the greater the tendency for the loanword to collocate with native and Sino-Japanese words than other loanwords. Onysko and Winter-Froemels (2011) categories of catachrestic and non-catachrestic innovations were found to be able to account for this grammatical and collocational behaviour of the loanwords, with the loanwords from the [noun/noun] list displaying tendencies to be marked lexical choices (non-catachrestic innovations), and the loanwords from the [particle] list displaying tendencies to be the unmarked informational choices (catachrestic innovations). Importantly, however, the findings from this chapter have shown the categorisation of loanwords to be better grounded in an analysis of their grammatical and collocational behaviour than on the notion of an equivalent expression.

7 Conclusion

7.1 Chapter Overview

This chapter summarises the answers reached in Chapters Five and Six to the research questions set out at the end of Chapter Three. After addressing several limitations of the study, it puts forward the overall contributions of the research and discusses the implications of the findings for Japanese-English language contact studies and the field of lexical borrowing research in general. The chapter ends with a brief discussion of several avenues of future research which could further build upon the findings made in this study of the grammatical behaviour of English loanwords in Japanese.

7.2 Re-statement of the Aims and Research Questions

The aim of this thesis has been to investigate the grammatical behaviour of frequently-used English loanwords in naturally-occurring Japanese texts, and in doing so to address the lack of attention given to this area of the analysis of English loanwords in Japanese. It was discussed in Chapter Three how not only in this context of English loanwords in Japanese but also in the field of lexical borrowing research in general the empirical investigation of the grammatical behaviour of loanwords has gone largely unexplored. Three interlinking factors were put forward to be causes of this situation. The first was that whilst the tradition of researching loanword phonology has revealed very important insights in the phonological structure of languages, the research has favoured the practice of examining loanwords as single-word units (Zenner & Kristiansen, 2014), and this is unconducive to an analysis of the loanwords' collocational and grammatical relationships in natural contexts of usage. Secondly, the tradition of researching the lexical categories of loanwords, which has been the main practice in borrowability research, has led to similar invaluable insights into how the lexicon of a language is structured but has also limited the idea of loanword grammar to a basic categorisation of their parts-of-speech. Thirdly, the tradition of using dictionaries for loanword data due to their authoritative status and convenient nature for quickly looking up lexical data has meant that corpus-informed resources have often been under-utilised in loanword research.

An outcome of the combination of these three traditions was stated as a lack of empirical research into the grammatical behaviour of loanwords. It was shown that for English loanwords in Japanese the majority of what little has been written previously on the topic has been in the form of unsubstantiated theoretical sketches, such as the statement that loanwords in Japanese "generally follow the morphological and syntactic rules of Japanese grammar" (Stanlaw, 2004, p. 77) and that they "fit into the Japanese syntactical structure as if they were native words, being ascribed particles such as subject and object markers where necessary" (Kay, 1995, p. 72). In light of the theoretical nature of these statements, the primary motivation of the present research has been to gather together the first extensive body of usage-based evidence of English loanwords in Japanese with which to either substantiate or refute such claims.

A further important issue raised in Chapter Three concerned the methodology of the present study. It was explained how the lack of empirical investigation into the grammatical behaviour of loanwords has in large part been an inevitable outcome of a lack of tools which could be used for such a labour-intensive exploration. The methodology used in the present study, of a corpus-based analysis of thousands of examples of the English loanwords in natural Japanese-language usage, has only become widely available in the last few decades (Inagawa, 2012). Furthermore, the word sketches used in this study which summarise tens of thousands of instances of each loanword in the corpus represent one of the more recent corpus-based tools. Because of this, the thesis has also aimed to use some of the most up-to-date corpus analysis tools with which to reconceptualise how the grammatical behaviour of loanwords can be explored. In trying to achieve these aims, the thesis was structured around finding the answers to the following three questions:

1. What patterns of distribution can be observed in the grammatical relationships of English loanwords in Japanese?
2. In what ways are these patterns of distribution similar to and different from patterns of distribution in the grammatical relationships of non-loanwords in Japanese?
3. What factors appear to account for observed differences in patterns of distribution in the grammatical relationships of loanwords and non-loanwords in Japanese?

In seeking the answers to these questions, the aim was to create an extensive body of regular patterns of grammatical behaviour of English loanwords in Japanese (Question 1), which could be used to substantiate or refute previous theoretical claims on the behaviour of the loanwords exhibiting negligible difference to that of other words in the language (Question 2), and to thereby better understand how English loanwords are used in the Japanese language (Question 3). The overall findings which were uncovered in seeking the answers to these questions will be discussed in the next section.

7.3 Principal Findings of the Research

In Chapter Four it was discussed how the corpus-based approach adopted in this study is grounded in the view that lexis and grammar are interdependent: conceptualised together as lexico-grammar rather than separate as lexis and grammar (Halliday, 1985; Hoey, 2005; Sinclair, 1991). Furthermore, the corpus analysis approach was discussed as bringing together the analysis of language structure with language usage, in that the methodology examines naturally-occurring language (usage) in contextual units (structure). This approach was seen as an effective way of countering some assumptions in previous research on loanwords discussed in Chapter Three, of loanwords being treated as single-word units even where their behaviour cannot be adequately accounted for without reference to the natural contexts in which they are embedded, and of relying on the limited data available in dictionaries.

Research Questions 1 and 2 were answered in Chapter Five. The examination of the database of 5870 grammatical relationships (gramrels) extracted from the word sketches of 587 frequently-used English loanwords found two patterns of behaviour which were particularly dominant: 61% of the loanwords were seen to participate most frequently in the [particle] gramrel, and 36.2% in the [noun/noun] gramrel. This was from a total of 57 gramrel types seen across all of the gramrel tokens in the database. These results became properly meaningful when the same analysis was done for the 130 native and Sino-Japanese words (non-loanwords). The database of 1300 gramrels of the non-loanwords was used as baseline comparative data. Whilst some areas of similarity were observed in the behaviour, such as in the number of gramrel types and the dominance of the [particle] gramrel, the loanwords were seen to substantially deviate in one particular aspect of their behaviour. For the non-loanwords, 81.6% of them were found to participate most frequently in the [particle] gramrel, which is similar behaviour

to the loanwords, but only 6.2% were found to participate most frequently in the [noun/noun] gramrel. Summarising the answers to Research Questions 1 and 2 into one sentence: a comparison of the grammatical behaviour of the loanwords and non-loanwords found that the loanwords exhibit a much stronger tendency than the non-loanwords to participate most frequently in the [noun/noun] gramrel.

Returning to the theoretical statements on the grammatical behaviour of English loanwords in Japanese which were discussed in Chapter One and Chapter Three, that loanwords "generally follow the morphological and syntactic rules of Japanese grammar" (Stanlaw, 2004, p. 77) and "fit into the Japanese syntactical structure as if they were native words, being ascribed particles such as subject and object markers where necessary" (Kay, 1995, p. 72), the empirical evidence presented in Chapter Five both partly supports and partly refutes such claims. Many of the loanwords analysed were indeed found to follow morphological and syntactic rules of Japanese and to be integrated into the syntactical structure of Japanese, but it was also found that many of them display a very restricted grammatical behaviour in the way that they occur much more commonly in compound nouns than do the non-loanwords. Overall, the empirical evidence shows that the grammatical behaviour of English loanwords in Japanese is more complex than what has been suggested by the theoretical claims and intuitive statements found in previous studies. The exploration of this issue was in Chapter Six, which is where Research Question 3 is answered.

To answer Research Question 3, the behaviour of the non-loanwords was again used as baseline comparative data. The 8 words participating most frequently in the [noun/noun] gramrel were analysed and their behaviour accounted for, finding that 5 of the 8 words (or 6 when *ichido* 'one time' is included as a collocation of *ichi* 'one') were shown to be numerals. This accounted for their marked grammatical behaviour compared to the other words in the 130-word sample. 2 of the other words were seen to display a more varied grammatical behaviour than the numerals, and their frequent participation in the [noun/noun] gramrel could be explained by their meaning, in that *choo* 'town/city' is very frequently observed in the corpus in compounds such as in the writing of addresses; and *nihon* 'Japan', being a country name, gets compounded with a very large range of words, such as *nihon go* (Japanese language), *nihon kokumin* (Japanese people), and *nihon gakkai* (Japanese academic society). The very large number of loanwords participating most frequently in the [noun/noun] gramrel, however, could not be accounted for in the same way. For this reason, three theories

which had previously been discussed in Chapter Three were turned to as potentially being able to account for the behaviour.

The discussion in Chapter Six, however, showed that two of the three theories were shown to be unable to account for the marked pattern of behaviour of the loanwords. Loveday's (1996) idea of Japanizing and Westernizing patterns was shown to be insufficiently developed as a theory because of the way that he only discussed concrete nouns (such as *doa* 'door' and *raisu* 'rice') and also based his ideas solely on his own intuition of loanword behaviour. Doi's (2014) discussion of the processes of naturalisation of loanwords was similarly shown to be unable to account for the behaviour because of how the theory is based on the extent of integration of the loanwords into the language, and all the loanwords analysed in the present study were selected as already being frequently-used integrated loanwords in the language.

The third theory, on the other hand, that of catachrestic and non-catachrestic innovations in a language, seemed to be highly pertinent to the findings from Chapter Five in its distinction between M-implicatures and I-implicatures, borrowed from Levinson's work on pragmatics and his notion of presumptive meanings (Levinson, 2000). It seemed that the loanwords shown in Chapter Five to be most frequently participating in the [particle] gramrel were what Onysko and Winter-Froemel (2011) had categorised as catachrestic innovations, as words used to "represent the 'normal' way of speaking about the objects or concepts concerned" (p. 1555). This was also suggested by the dominance of this behaviour in the frequently-used non-loanword sample, in that the non-loanword nouns can on the whole be seen to be the normal way of speaking about things. The loanwords most frequently participating in the [noun/noun] gramrel, on the other hand, seem well accounted for by the category of non-catachrestic loanwords, as the marked form of communication which are used for their stylistic effect of being able to express a particular manner of communication. To explore the relevance of this theory in accounting for the behaviour of the loanwords uncovered in Chapter Five, Section 6.4 in Chapter Six then explored the collocational behaviour of a sample of the English loanwords, half of which were seen in Chapter Five to participate most frequently in the [particle] gramrel, and the other half most frequently in the [noun/noun] gramrel.

The analysis in Chapter Six of the most salient collocates of the 30-loanword sample revealed notable patterns of collocational behaviour that matched with their grammatical behaviour and seemed to again very well reflect the distinction between catachrestic and non-catachrestic innovations. It was found that loanwords participating

most frequently in the [noun/noun] gramrel collocated most strongly with other loanwords, and in the vast majority of cases this was most typically in a compound noun relation. This very strong tendency for the loanwords to join together with one another in a single relationship of a compound noun showed a behaviour of the loanwords that was not seen in the behaviour of the 130 non-loanwords, meaning that these loanwords can be seen as non-catachrestic innovations conveying M-implicatures due to their marked behaviour. The loanwords participating most frequently in the [particle] gramrel, on the other hand, collocated most strongly with non-loanwords, and this was in a wide range of grammatical relationships. This behaviour was much more typical of that seen in the non-loanwords, meaning that these loanwords can be seen as catachrestic innovations conveying I-implicatures due to them being the unmarked means of expression. As was noted in Chapter Six, whilst the Longest-Commonest Match given in the tables of collocates gives an important initial impression of the contexts of usage of the loanwords, a more thorough and specialised study of the contexts would be an insightful future direction of the research. This will be returned to in Section 7.6 below.

Table 6.18 in Chapter Six revealed another very important finding from this analysis of the sub-sample of 30 loanwords: that the strength of the collocational and grammatical behaviour of the loanwords was related to the degree of difference in frequencies between the loanwords' first and second most preferred gramrels. Therefore, the loanwords with the greatest degree of difference between [noun/noun] and their second most preferred gramrel had the strongest tendency to collocate with other loanwords, and the loanwords with the greatest degree of difference between [particle] and their second most preferred gramrel had the strongest tendency to collocate with non-loanwords. The importance of this finding was that it suggested the simple binary categorisation into catachrestic and non-catachrestic, as put forward by Onysko and Winter-Froemel (2011), was not sufficient to account for the gradience in grammatical and collocational behaviour of the 30 loanwords analysed in Chapter Six.

A further issue became apparent from the data shown in Table 6.19 in Chapter Six. This was in the way that Onysko and Winter-Froemel's (2011) methodology of categorising catachrestic and non-catachrestic loanwords follows a tradition seen throughout the history of loanword research of seeking to identify an equivalent expression for the loanword in a dictionary or similar lexical resource. If an equivalent expression is found, the loanword is a non-catachrestic innovation. If one cannot be found, it is a catachrestic innovation. Table 6.19 showed, however, that of the 30

loanwords in the sample in the present study, only one did not have any equivalent expressions listed in two major Japanese dictionaries, making 29 of them technically non-catachrestic innovations. Whilst Onysko and Winter-Froemel go on in their article to discuss the need to assess the appropriateness of an equivalent expression in natural language resources to allow a more reliable categorization (p. 1563), such an investigation of all 587 loanwords was beyond the means of the present study. However, even without such a check for appropriateness, the dictionary data alone in Table 6.19 was a strong indication that the theory needed modifying. The modification to the theory suggested at the end of Chapter Six is to base the categorisation on the observed grammatical and collocational behaviour of the loanwords rather than on the notion of an equivalent expression. Not only does this give the categorisation an empirical basis, but also it allows gradience in the extent to which a loanword is classed as catachrestic or non-catachrestic.

7.3.1 *Limitations of the research*

The limitations of this study can be grouped into three interrelated areas: the data, the software, and the researcher. Concerning the data, three main limitations will be discussed. The first is that only 587 English loanwords were analysed in Chapter Five, and only a sub-set of 30 were analysed in Chapter Six. It was discussed in Chapter Two that the Japanese language contains tens of thousands of English loanwords, so the samples analysed in this study represents just a small fraction of all the English loanwords in Japanese. To get a much richer picture of the grammatical and collocational behaviour of English loanwords in Japanese, it would therefore be insightful to examine a larger sample, in the region of thousands rather than the hundreds analysed in this study.

However, the methodology adopted in the present study, which was a mix of a qualitative analysis of the quantitative data output of the word sketches, necessarily limited the amount of word sketches which could be manually analysed. Whilst the software is able to produce a word sketch in seconds, the manual analysis and combining together of the word sketches is much more labour-intensive. For this reason, the number of loanwords analysed was kept to an amount considered satisfactory for achieving the aims of the study. Furthermore, one of the main motivations of this research was to address the issue of many previous studies having

only analysed one or a small handful of loanwords. In this way, whilst future research could benefit from increasing the number of loanwords analysed, the present study stands as the most extensive analysis of the grammatical behaviour of English loanwords in Japanese yet conducted. It will be interesting, then, to see if the continued development of the Sketch Engine software allows in the future an automated way of achieving what was done manually in this research. This is certainly a possibility because even over the course of the present study, the function of creating multi-word word sketches has significantly improved (i.e. word sketches looking at the grammatical and collocational behaviour of phrases and multi-word expressions). Further developments in this area could therefore possibly see the ability to automatically collate and summarise the data in tens or hundreds of word sketches.

A further limitation of the data relates to the nature of the corpus from which the word sketches were derived. As discussed in Chapter Four, the jpTenTen11 corpus is a mega-corpus of web language and this brings with it issues such as sufficiently accounting for the representative of the corpus. It is difficult, for example, to summarise the different genres of web-based language beyond the software offering lists of websites and Top-Level Domains. Even with these lists, web pages can be very eclectic in the language they contain, such as a web page of a news article also containing advertising, links to other pages, reader comments, and embedded feeds from other websites (such as Twitter and Facebook). The Sketch Engine software does, however, make considerable efforts to clean up the 'noise' of web-based language in its raw form in the TenTen corpora (i.e removing web links, repeated text, etc.), and these steps were discussed in Chapter Four. Nevertheless, difficulties in establishing the genres of language usage in web-based corpora, compared to what is possible with more traditional corpora, is a limitation of web corpora that needs to be acknowledged. For this reason, the results of the grammatical behaviour of the English loanwords in the present study were not discussed in terms of their situational context of usage, which needs to be acknowledged as a limitation in the richness of the data discussion. The way to overcome this issue is to conduct a similar analysis on corpora which have been compiled to ensure a more representative balance between genres and registers. An issue with this, however, is that this will likely decrease the size of the corpus and thereby decrease the number of occurrences of loanwords, a problem which was discussed in relation to the corpus-based studies carried out by Inagawa (2010), Mogi (2012), and Bordilovskaya (2016). In any case, the question of whether or not the findings of the present study match well with other genres of language, such as

loanword usage in newspapers, advertising, and political documents, would be an interesting avenue of future research.

The last data limitation of this study to be discussed here is the fact that the sample of non-loanwords only included a sample of 130. The reason for this smaller sample compared to that of the loanwords is the same as why only hundreds and not thousands of loanword word sketches were analysed: because the manual analysis of the word sketches restricted the number which could be investigated within the limits of the present study. Even still, the sample of non-loanwords was not intended to be the same size as the sample of loanwords, but instead was intended to act as baseline comparative data for the main investigation of the loanwords. It was considered that even in a sample of just 130 words, an idea of where the behaviour of the two samples was similar and where it was different would be able to emerge. The emergent patterns would then be explored in more detail, which is what was done in Chapter Six

There were two main software limitations in the present study: the fact that only the Sketch Engine software was used, and the fact that the data is reliant on the quality of the word sketches. The reason for selecting the Sketch Engine corpus analysis tool from the various other tools available, such as AntConc, WordSmith Tools, and LancsBox, was that it was the only tool available at the time of the start of the research (and still at the time when the research was completed) that allowed grammatical and collocational summaries to be produced of Japanese language data. The Sketch Engine was the only tool I found which could automatically summarise thousands of instances of the loanwords in their linguistic contexts into a compact format which I could then collate together manually into a database of grammatical relationships. Ideally, it would have been insightful to compare the grammatical behaviour of the loanwords in the output from several different tools, to see if different algorithms and statistics in different tools resulted in different overall findings. If in the future such a possibility arises, this would be another interesting avenue of further research.

Related to this issue is that the findings in Chapter Five and Chapter Six are reliant on the quality of the Sketch Engine's word sketches. This issue was discussed in Chapter Four in an explanation of the evaluation projects run by the designers of the software to assess the quality of the automated output in the word sketches. These evaluations not only showed word sketches to be of high quality, but also showed that Japanese word sketches specifically performed best from the four languages studied. The same as above, no other options could be found at the time of the research but if in the future other software is able to produce similar summaries of the grammatical

behaviour of Japanese vocabulary, a comparison of the various outputs would be an important addition to the findings presented in this study.

The limitations of the researcher refer to the personal decisions I made in this research. Chapter Three discussed that there has been no previous large-scale, empirical analysis of the grammatical behaviour of English loanwords in Japanese. Therefore, many of the decisions taken in the present study in how to carry out such an analysis were my own. There were, of course, precedents which I learnt from in studies in other areas of research, such as the extensive corpus-based analyses on grammar patterns carried out by Hunston and Francis (2000), but in the present context of the grammar of English loanwords in Japanese, there were no such precedents. My own skills as a researcher, particularly as a corpus linguist, developed considerably over the course of the research. I began by manually highlighting each instance of a loanword I found on a year's worth of front pages of a national Japanese newspaper, and developed my skills over the course of the research into the methodology I have detailed in Chapter Four. Whilst I believe that this methodology is a sound one, and I have justified the decisions I have made, a researcher approaching the same project with a different methodology may well uncover new insights. However, I believe that a different methodology would still uncover similar patterns of behaviour to those uncovered in the present study.

7.4 Contributions and Implications of the Research

This study provides for the first time an empirically-based, large-scale account of the grammatical behaviour of English loanwords in Japanese. It has found that whilst some English loanwords in Japanese behave very much like native and Sino-Japanese words, being appended with particles so that they can take on a large range of grammatical functions, a very large number deviate from the behaviour of the native and Sino-Japanese words and display grammatical specialisation in that they occur very frequently with other loanwords in a compound noun structure. In this way, the research has contributed to a perspective of loanwords being embedded in a wide variety of grammatical relationships, rather than just simply slotted into open lexical gaps in the language. This latter point in particular has important pedagogical implications.

The massive number of English loanwords occurring in general spoken and written Japanese discourse means that they will be encountered even at the elementary stage of learning Japanese as a foreign language. For the learners of Japanese residing in Japan,

these encounters are not only in their language-learning textbooks, but also all around them in the Japanese linguistic landscape. It was discussed in Chapter 2 how virtually every aspect of Japanese society has seen the introduction of English loanwords into its discourse, creatively displayed on signs, posters, product packaging, and fashion items. As such, a learner of Japanese who has knowledge of the English language will in many ways be assisted by these English loanwords integrated into all aspects of the Japanese language. They are likely to be familiar with the general meaning of the loanwords, even if their meanings do not match exactly with the original forms in English, and with some knowledge of the phonological and orthographical adaptations which the loanwords undergo, will be able to recognise and produce the loanwords. This reduces the learning burden associated with the large amount of new vocabulary at the initial stages of language learning. To assist in this reduction of the learning burden of new vocabulary, there are many lexical resources available to learners on these adaptations made to loanwords, such as specialised loanword dictionaries.

These elementary learners will be largely unfamiliar, however, with how the loanwords can be integrated into the grammatical structure of a Japanese sentence or utterance. For this, there are currently no easy-access English-language resources available and they may therefore assume, for example, that there is always freedom in how the loanwords can be used within Japanese sentences and utterances. In other words, they may assume that an English loanword can be slotted into any appropriate gap in Japanese: the same way as I believed that *sutoroberii* 'strawberry' could be slotted into the 'object' gap to form the utterance *sutoroberii o tabeta* 'I ate strawberries' (see Chapter One). The findings in the present study have shown that this is fact not the case. Rather, it has been shown in the research here that a large number of English loanwords have a specialised grammatical usage in Japanese. Some of them almost always appear in noun compounds and almost exclusively with other loanwords. Other loanwords, on the other hand, display a more standard behaviour in the way that they take the same kind of grammatical markers which the native and Sino-Japanese words take. This knowledge can further help in reducing the learning burden of a word, in that not only is the learner assisted in knowledge of the form and meaning of the word, but also in its grammatical structuring with other words. Having a list of frequently-used loanwords ordered by the extent of grammatical specialisation, at one end, and grammatical standardisation, at the other, would be a very useful lexical resource for learners of Japanese as a foreign language. Such a list could be created from expanding the analysis done in Chapter Six.

This study has also suggested that the long-held practice of categorising the functions of loanwords based on whether or not equivalent expressions can be found in the native language, typically by means of searching in a dictionary, is problematic. As an alternative methodology, this study has suggested that the categorisation is better grounded in empirical foundations, such as the observed grammatical behaviour of the loanwords in natural language. This overcomes the complicated issues of deciding in which lexical resources to search for possible equivalent expressions, and deciding how to most effectively establish the appropriateness of equivalent expressions, in that they may fall in and out of use and take on and lose subtle shades of distinctive meaning over time. Basing the functional categorisation instead on the observed grammatical behaviour of the loanwords themselves also allows the categorisation to be made along a continuum rather than into two discrete types.

7.5 Directions of Further Research

The research in the present study has opened up several avenues of possible further work on this topic: some of these would help strengthen the findings of the present study, and others would be interesting new directions built on the foundation of investigating the grammatical behaviour of loanwords. As was discussed above in Section 7.3.1, there were necessary limitations in the number of word sketches which could be manually analysed in the present study. If future technological developments allow it, it would be interesting to explore a way of automating the collation and summary of the data of very large sets of word sketches. This would then allow comparative analyses to be conducted on specific sub-sets of loanwords, such as frequently-used vs rare loanwords, and technical vs general loanwords. Furthermore, it would be informative to do cross-linguistic comparisons of the grammatical behaviour of loanwords, to see if the same types of grammatical distribution are seen across languages.

One area of analysis in particular could further extend the understanding of the grammatical behaviour of the loanwords presented in this study. It was discussed in Section 4.4.4 how the frequency rank of each loanword was recorded in each of the three corpus-derived wordlists, but that this distribution data was not further investigated after combining the lists into a single collection of loanwords. If the frequency distribution of the loanwords was analysed in depth, it could uncover a

possible relationship between the grammatical behaviour of the loanwords and their frequency distribution in the Japanese lexicon. It may be the case, for example, that the more frequent an English loanword is in Japanese, the greater the tendency there is for it to appear in a certain grammatical relationship. Such expanded knowledge on the grammatical behaviour of English loanwords in Japanese would be especially useful in language education, where patterns of behaviour linked with frequency could be made explicit in vocabulary teaching materials.

Related to this, the findings of the present study would benefit from a comparison of other types of English loanwords in Japanese, such as academic or technical loanwords, or loanwords in specific genres, such as fiction writing and newspaper articles, as well as loanwords in other languages, to see if the same catachrestic and non-catachrestic categorisation can be applied to their observed grammatical behaviour. It would be insightful to see if such rarer, technical, and genre-specific loanwords in Japanese and in other languages also displayed differences in their grammatical behaviour, and if so, to create specially ordered lists.

A further area of analysis which would definitely benefit from a more focused analysis is the precise contexts of usage of the loanwords. Whilst an indication of these was given in the Longest Commonest Matches given in the tables of collocates in Chapter Six, they are only initial sketches of exactly what meanings the loanwords take on in their contexts of usage. A more in-depth analysis of the concordance lines of more than just the 30 loanwords analysed in Chapter Six would be an insightful future direction of the research. The aim of the present study, however, has always been to uncover the grammatical behaviour of the loanwords rather than to focus on their specific meanings. There is also the issue that conducting such an in-depth analysis of the concordance lines would be a very labour-intensive task to do manually, especially considering the thousands of instances of the loanwords in the corpus, and this might return the analysis to individual or small samples of loanwords. The Longest Commonest Match is a feature of the Sketch Engine which is an attempt to overcome this problem, and future developments in corpus analysis software may also come up with better ways of automating an analysis of the meanings of the loanwords in their context.

One particularly appealing area of research on the subject of the grammatical behaviour of loanwords is a side-by-side comparison of the behaviour of a loanword with one of its equivalent expressions, if it has one. The problem of judging the appropriateness of equivalent expressions has been discussed above in Section 7.3.1 as

well as in the work of Onysko and Winter-Froemel (2011), but in some cases, an easy pairing can be made between a loanword and a parallel expression already in the native language. Returning to the example in Chapter One of the English loanword *sutoroberii* 'strawberry' in Japanese and its native equivalent expression *ichigo*, there already exists the corpus-based functionality of comparing the grammatical behaviour of the loanwords. Again, this functionality is so far only available in the Sketch Engine and is offered in the sketch diff function.

Figure 7.1 shows an extract of the sketch diff output for a comparison of the grammatical behaviour of *sutoroberii* and *ichigo*, focusing specifically on the [noun/noun] gramrel (i.e. compound noun structure). The red-white-green colouring scale in the sketch diff output is used to compare how strongly the collocates co-occur with one word over the other, with the deepness of the colouring representing the strength of the collocation. The data in Figure 7.1 shows a fairly even split between the words collocating most strongly with *sutoroberii* (coloured in red), those collocating with both (coloured in white), and those collocating with *ichigo* (coloured in red). The words collocating with *sutoroberii* are all other loanwords, and are words which show *sutoroberii* to be typically used as a flavouring or colouring of other products, such as *sutoroberii furapechiino* 'strawberry frappuccino' and *sutoroberii burondo* 'strawberry blonde'. The words collocating more strongly with *ichigo*, on the other hand, are a mix of loanwords and non-loanwords, and more strongly show the meaning of *ichigo* to be the fruit itself, such as *ichigo jamu* 'strawberry jam' and *ichigo daifuku* 'strawberry stuffed rice cake ball'.

noun/noun	17,295	72,585	0.63	0.36
コーンズ	125	0	7.8	--
クオーツ	459	0	7.5	--
フラペチーノ	70	0	6.7	--
リナックス	48	0	6.1	--
フィールド	633	0	6.0	--
グアバ	38	0	5.9	--
ブロンド	77	11	6.1	2.0
ストロベリー	64	36	5.9	3.7
シェイク	83	61	6.2	4.4
ミルフィーユ	98	94	7.0	5.2
ホイップ	77	115	6.1	5.3
ムース	101	476	5.8	7.1
パフェ	173	815	6.5	7.8
ショート	396	1,302	5.5	7.0
ジャム	749	4,538	7.3	9.4
シロップ	65	468	4.5	6.7
タルト	101	905	5.4	7.8
苺	46	821	2.6	6.5
ミルク	191	6,227	4.3	9.1
牛乳	9	887	0.2	6.5
鈴	0	611	--	6.5
ポッキー	0	268	--	6.6
一会	0	364	--	6.8
白書	0	691	--	7.5
大福	0	4,213	--	10.3

Figure 7.1 A sketch diff for *sutoroberii* and *ichigo* showing the [noun/noun] gramrel.

Such a corpus-based analysis of the side-by-side comparison of the grammatical and collocational behaviour of a loanword and its parallel expression in the native language is one of the most sophisticated analyses which can currently be done with corpus data. As such, this function not only offers a very interesting alternative direction of research into the grammatical behaviour of loanwords, but also stands as an example of the rapid development of corpus-analysis software and the future opportunities for linguistic research afforded by the automated analysis of natural language data.

References

Asahara, M., Ono, H., & Miyamoto, E. T. (2016). Reading-time annotations for Balanced Corpus of Contemporary Written Japanese. *Proceedings of the 26th International Conference on Computational Linguistics (COLING-16)*, 684–694.

Backhaus, P. (2005). *Signs of multilingualism in Tokyo – a linguistic landscape approach*. Universitat Duisburg-Essen.

Backhaus, P. (2007). *Linguistic landscapes: A comparative study of urban multilingualism in Tokyo*. Clevedon: Multilingual Matters. R

Backus, A. (2014). A usage-based approach to borrowability. In E. Zenner & G. Kristiansen (Eds.), *New perspectives on lexical borrowing: Onomasiological, methodological, and phraseological innovations* (pp. 19–39). Berlin: De Gruyter Mouton.

Barrs, K. (2011). Creativity with Scripts in the Japanese Language. *Peerspectives Online*, *7*, 14–19.

Barrs, K. (2012a). Implications of cross-linguistic lexical similarities for Japanese learners of English. *神田外語大学言語教育研究 Kanda University Studies in Linguistics and Language Teaching*, *23*, 27–37.

Barrs, K. (2012b). The modern tragedy of English teaching and learning in Japan. *Asian Englishes*, *15*(2), 102–106.

Barrs, K. (2013a). A preliminary analysis of cross-linguistic lexical similarities between English and Japanese. *Working Papers in Language Education and Research*, *1*(1), 11–20.

Barrs, K. (2013b). Categorising English loanwords in Japanese. *Vocabulary Education & Research Bulletin*, *2*(1), 2–3.

Barrs, K. (2015). Errors in the use of English in the Japanese linguistic landscape. *English Today*, *31*(4).

Ben-rafael, E., Shohamy, E., Amara, H. A., & Trumper-Hecht, N. (2006). Linguistic landscape as symbolic construction of the public space: The case of Israel. *International Journal of Multilingualism*, *3*(1), 7–30.

Berns, M. (2005). Expanding on the expanding circle: Where do WE go from here ? *World Englishes*, *24*(1), 85–93.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, England: Cambridge University Press.

Blair, A. D., & Ingram, J. (2003). Learning to predict the phonological structure of English loanwords in Japanese. *Applied Intelligence*, *19*, 101–108.

Bordilovskaya, A. V. (2012). A Study of Loan Color Terms Collocation in Modern Japanese. In N. Miyake, D. Peebles, & R. . Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1362–1367). Austin, TX: Cognitive Science Society.

Bordilovskaya, A. V. (2016). *Collocation tendencies and classification of gairaigo ddjectives in contemporary Japanese: Corpus-based study*. Kobe University.

Brezina, V., & Gablasova, D. (2015). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics*, *36*(1), 1–22.

Brown, J. B. (1995). Is Gairaigo English? *The Internet TESL Journal*, *1*(2).

Butler, Y. G. (2007). Foreign Language Education at Elementary Schools in Japan: Searching for Solutions Amidst Growing Diversification. *Current Issues in Language Planning*, *8*(2), 129–147.

Caesar, T. (2005). English in Japan. *Salmagundi*, *146/147*, 159–169.

Chavez, A. (2009). The problem of loanwords in Japan, and returning them. *The Japan Times*. Retrieved from www.japatimes.co.jp

Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Abingdon, England: Routledge.

Cipris, Z., & Hamano, S. (2002). *Making sense of Japanese grammar: A clear guide through common problems*. Honolulu: University of Hawai'i Press.

Crawford, C. J. (2009). *Adaptation and transmission in Japanese loanword phonology*. Doctoral Dissertation. Cornell University.

Daijirin. (2006) (3rd ed.). Tokyo: Sanseido.

Daulton, F. E. (1999). English loanwords in Japanese – The built-in lexicon. *The Internet TESL Journal*, *5*(1), 1–5.

Daulton, F. E. (2003). List of High-Frequency Baseword Vocabulary for Japanese EFL Students #2. *The Internet TESL Journal*, *9*(3), 1–15.

Daulton, F. E. (2004). The creation and comprehension of English loanwords in the Japanese media. *Journal of Multilingual and Multicultural Development*, *25*(4), 285–296.

Daulton, F. E. (2008). *Japan's Built-in Lexicon of English-based Loanwords*. Clevedon: Multilingual Matters Ltd.

Daulton, F. E. (2009). A defence of gairaigo. *The Ryugoku University International Centre's Annual Research Report*, *18*, 41–45.

Daulton, F. E. (2011). On the origins of gairaigo bias: English learners' attitudes towards English-based loanwords in Japan. *The Language Teacher*, *35*(6), 7–12.

Diez-Arroyo, M. (2016). Vagueness: A Loanword's good friend. The case of "print" in Spanish fashion. *Pragmatics*, *26*(4), 609–629.

Doi, S. (2014). The naturalisation process of the Japanese loanwords found in the Oxford English Dictionary. *English Studies*, *95*(6), 674–699.

Dougill, J. (2008). Japan and English as an alien language. *English Today*, *24*(01), 18–22.

Durkin, P. (2014). *Borrowed words: A history of loanwords in English*. Oxford, England: Oxford University Press.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Erjavec, I. S., Erjavec, T., & Kilgarriff, A. (2008). A web corpus and word sketches for Japanese. *Journal of Natural Language Processing*, *15*(2), 137–159.

Firth, J. . (1957). A synopsis of linguistic theory, 1930-1955. In *Studies in Linguistic Analysis* (pp. 1–32). Oxford: Blackwell.

Francis, G., Hunston, S., & Manning, E. (1996). *Collins COBUILD Grammar patterns 1: Verbs*. London: HarperCollins.

Francis, G., Hunston, S., & Manning, E. (1998). *Collins COBUILD Grammar Patterns 2: Nouns and adjectives*. London: HarperCollins.

Frellesvig, B. (2010). *A History of the Japanese Language*. Cambridge: Cambridge University Press.

Fukada, T. (2010). Elementary schools to get English. *The Japan Times*. Retrieved from www.japantimes.co.jp

Gatto, M. (2014). *Web as corpus*. London: Bloomsbury Academic.

Gorlach, M. (1990). *Studies in the history of the English language*. Heidelberg: Carl Winter.

Gorlach, M. (Ed.). (2001). *A dictionary of European Anglicisms*. Oxford: Oxford University Press.

Gottlieb, N. (2005). *Language and Society in Japan*. Cambridge: Cambridge University Press.

Greavu, A. (2014). A classification of borrowings: Observations from Romanian/English contact. *Diversité et Identité Culturelle En Europe*, *10*(2), 95–104.

Gries, S. T. (2012). Behavioral profiles A fine-grained and quantitative approach in corpus-based lexical semantics. In G. Libben, G. Jarema, & C. Westbury (Eds.), *Methodological and Analytic Frontiers in Lexical Research* (pp. 57–80). Amsterdam: John Benjamins Publishing Company.

Gries, S. T., & Divjak, D. (2009). Behavioral profiles: a corpus-based approach to cognitive semantic analysis. In V. Evans & S. S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 57–75). Amsterdam: John Benjamins Publishing Company.

Halliday, M. A. . (1985). Systemic background. In *On Language and Linguistics*.

Hasegawa, Y. (2015). *Japanese: A linguistic introduction*. Cambridge: Cambridge University Press.

Haspelmath, M. (2008). Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability. *Empirical Approaches to Language Typology*, (ii), 1–15.

Haspelmath, M. (2009). Lexical borrowing: concepts and issues. In M. Haspelmath & U. Tadmor (Eds.), *Loanwords in the world's language: A comparative handbook* (pp. 35–54). Berlin: De Gruyter Mouton.

Haspelmath, M., & Tadmor, U. (2009). The loanword typology project and the WOLD loanword database. In M. Haspelmath & U. Tadmor (Eds.), *Loanwords in the world's languages: A comparative handbook* (pp. 1–34). Berlin: De Gruyter Mouton.

Hattori, S. (1959). *Nihongo no keito*. Tokyo: Iwanami Shoten.

Haugen, E. (1950). The Analysis of Linguistic Borrowing, *26*(2), 210–231.

Haugen, E. (1953). *The Norwegian language in America: A study in bilingual behavior*. Bloomington: Indiana Univ. Press.

Hickey, R. (2010). Language contact: Reconsideration and reassessment. In R. Hickey (Ed.), *The handbook of language contact* (pp. 1–28). Chichester: Wiley-Blackwell.

Hock, H. ., & Joseph, B. . (2009). Lexical Borrowing. In *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics* (2nd ed.). Berlin: Mouton De Gruyter.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Abingdon, England: Routledge.

Hoffer, B. L. (1990). English Loanwords in Japanese: Some Cultural Implications. *Language Sciences*, *12*(1), 1–21.

Hoffer, B. L. (2002). Language borrowing and language diffusion: An overview. *Intercultural Communication Studies*, *XI*(4), 1–37.

Holden, K. (1976). Assimilation rates of borrowings and phonological productivity. *Language*, *52*, 131–147.

Hongo, J. (2013). Abe wants TOEFL to be key exam. *The Japan Times*. Retrieved from www.japantimes.co.jp

Honna, N., & Takeshita, Y. (1998). On Japan's propensity for native speaker english: A change in sight. *Asian Englishes*, *1*(1), 117–134.

Hunston, S. (2002). Corpora in applied linguistics. Cambridge, England: Cambridge University Press.

Hunston, S., & Francis, G. (2000). *Pattern Grammar*. Amsterdam: John Benjamins Publishing Company.

Hyde, B. (2002). Japan's emblematic English. *English Today*, *18*(03), 3–7.

Igarashi, Y. (2007). *The Changing Role of Katakana in the Japanese Writing System: Processing and Pedagogical Dimensions for Native Speakers and Foreign Learners*. University of Victoria.

Inagawa, M. (2010). *A Corpus-Driven Study of Loanwords : Synchronic and Diachronic Change of English-Derived Words in Contemporary Japanese*.

Inagawa, M. (2012). Evolution of the English-derived word 'karā' in contemporary Japanese. *Japanese Studies*, *32*(3), 377–397.

Irwin, M. (2011). *Loanwords in Japanese*. Amsterdam: John Benjamins Publishing Company.

Ito, J., & Mester, A. (2008). The Phonological Lexicon. In *Optimality Theory in Phonology* (pp. 552–568). Oxford: Blackwell Publishing Ltd.

Jakubicek, M., Kilgarriff, A., Vojtech, K., Rychly, P., & Suchomel, V. (2013). The TenTen corpus family. In *7th International Corpus Linguistics Conference* (pp. 125–127). Lancaster, England.

Kachru, B. B. (1978). Toward restructuring code mixing: an Indian perspective. *International Journal of the Sociology of Language*, *16*, 27–46.

Kachru, B. B. (1988). The sacred cows of English. *English Today*, *4*(4), 3–8.

Kachru, B. B. (1990). World Englishes and applied linguistics. *World Englishes*, *9*(1), 3–20.

Kachru, Y., & Nelson, L. C. (2006). *World Englishes in Asian contexts*. Hong Kong: Hong Kong University Press.

Kageyama, T., & Saito, M. (2016). Vocabulary strata and word formation processes. In T. Kageyama & H. Kishimoto (Eds.), *Handbook of Japanese lexicon and word formation* (pp. 11–50). Boston, M.A.: Walter de Gruyter Inc.

Kang, Y. (2013). Loanwords. In *Oxford Bibliographies: Linguistics*. Oxford.

Katayama, M. (1998). *Optimality Theory and Japanese loanword phonology*.

Kawahara, S., & Sano, S. (2013). A corpus-based study of geminate devoicing in Japanese: linguistic factors. *Language Sciences*, *40*, 300–307.

Kay, G. (1995). English loanwords in Japanese. *World Englishes*, *14*(l), 67–76.

Kilgarriff, A. (2005). Putting the corpus into the dictionary. In *Proceedings of the MEANING workshop*. Trento, Italy.

Kilgarriff, A., Baisa, V., Busta, J., Jakubicek, M., Vojtech, K., Michelfeit, J., … Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography ASIALEX*, *1*, 7–36.

Kilgarriff, A., Kovar, V., Krek, S., Srdanovic, I., & Tiberius, C. (2010). A Quantitative Evaluation of Word Sketches. In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the 14th EURALEX International Congress* (pp. 372–379). Fryske Akademy.

Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In *Proceedings of the Eleventh Euralex International Congress* (pp. 105–116). Lorient, France: Information Technology Research Institute.

Kilgarriff, A., & Tugwell, D. (2001). Word sketch: Extraction and display of significant collocations for lexicography. In *Proceedings of ACL workshop on collocation Computational xxtraction, analysis and exploitation* (pp. 32–38).

Kistler, M. (1995). Understanding Katakana English: a consciousness-raising activity. *CELE Journal*, *15*, 1–8.

Koscielecki, M. (2006). Japanized English, its context and socio-historical background. *English Today*, *22*(04), 25–31.

Kuang, C. H. (2009). Common Loanwords Identified in Japanese Music Magazines. *Polyglossia*, *17*, 1–16.

Levinson, S. C. (2000). *Presumptive meanings. The theory of generalized conversational implicature*. London, England: The MIT Press.

Lexical Computing CZ. Sketch Engine Corpus Analysis Software. Available at https://www.sketchengine.eu/

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh, Scotland: Edinburgh University Press.

Loveday, L. (1996). *Language Contact in Japan: A Socio-Linguistic History*. Oxford: Oxford University Press.

Lovins, J. B. (1975). *Loanwords and the phonological structure of Japanese*. Bloomington: Indiana Univ. Linguistics Club.

Maekawa, K. (2007). Design of a balanced corpus of contemporary written Japanese. In *Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2007)* (pp. 55–58).

Maekawa, K., Yamazaki, M., Maruyama, T., Yamaguchi, M., Ogura, H., Kashino, W., Den, Y. (2010). Design, compilation, and preliminary analyses of balanced corpus of contemporary written Japanese. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation* (pp. 1483–1486).

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, *48*, 345–371.

Martin, A. (2004). The katakana effect and teaching English in Japan. *English Today*, *20*(1), 50–55.

Matras, Y. (2009). Language contact, 366. https://doi.org/10.1017/CBO9780511809873

McArthur, T. (1987). The English languages? *English Today*, *3*(3), 9–13.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics*. Cambridge, England: Cambridge University Press.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. *Corpus linguistics: An introduction* (2nd ed.). Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus based language studies*. Abingdon, England: Routledge.

McKay, S. (2002). *Teaching English as an international language*. Oxford: Oxford University Press.

McKenzie, R. M. (2008). The complex and rapidly changing sociolinguistic position of the English language in Japan: a summary of English language contact and use. *Japan Forum*, *20*(2), 267–286.

Meyer, C. F. (2004). *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Meyer, C. F. (2009). *Introducing English Linguistics*. Cambridge: Cambridge University Press.

Mie, A. (2013). LDP takes aim at English education, seeks to boost TOEFL levels. *The Japan Times*.

Miller, R. A. (1967). *The Japanese Language*. Chicago: University of Chicago Press.

Miura, A. (1979). *English loanwords in Japanese: A selection*. Tokyo, Japan: Charles E. Tuttle Company, Inc.

Mogi, T. (2012). Towards the lexicographic description of the grammatical behaviour of Japanese loanwords: A case study. *Acta Linguistica Asiatica*, *2*(2), 21–34.

Morrow, P. R. (1987). The users and uses of English in Japan. *World Englishes*, *6*(1), 49–62.

Myers-Scotton, C. (2006). *Multiple voices: An introduction to bilingualism.* Malden, MA: Blackwell Publishers Ltd.

Namiki, T., & Kageyama, T. (2016). Word structure and headedness. In T. Kageyama & H. Kishimoto (Eds.), *Handbook of Japanese lexicon and word formation* (pp. 201–235). Berlin: De Gruyter Mouton.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, M.A.: Heinle & Heinle Publishers.

Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). Amsterdam: John Benjamins.

Nation, I. S. P., & Newton, J. (1997). Teaching vocabulary. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 238–254). Cambridge: Cambridge University Press.

NINJAL (2014). BCCWJ word lists [data file]. Retrieved from http://pj.ninjal.ac.jp/corpus_center/bccwj/en/freq-list.html

NINJAL. (1962). *Hokoku 21: Gendai zasshi 90-shu no yogoyoji daiichibunsatsu: hyoki oyobi goihyo (Report 21: Vocabulary and Chinese characters in 90 magazines of today, volume I: General description and vocabulary frequency tables)*. Tokyo.

NINJAL. (1964). *Hokoku 25: Gendai zasshi 90-shu no yogoyoji daisanbunsatsu: bunseki (Report 25: Vocabulary and Chinese characters in 90 magazines of today, volume III: Analysis)*. Tokyo.

NINJAL. (2004). *Gairaigo ni kansuru ishiki chosa (A survey on attitudes to loanwords)*. Tokyo.

NINJAL. (2005). *Gendai zasshi no goi chosa: 1994-nen hakko 70-shi (A vocabulary survey of modern magazines: 70 magazines published in 1994)*. Tokyo.

Nishi, H. (2017). Japanese N deshita in discourse: past form of Ndesu. *Acta Linguistica Asiatica*, *7*(1), 41–56.

Nishimura, M. (1995). A functional analysis of Japanese/English code-switching. *Journal of Pragmatics*, *23*(2), 157–181.

Nuttall, M. (2000). English in Japan. *Japan Railway and Transport Review*, *25*, 56–57.

Odlin, T. (1989). *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.

Onysko, A., & Winter-Froemel, E. (2011). Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics*, *43*(6), 1550–1567.

Osaki, T. (2013, June 27). Gifu man, 71, sues NHK for distress over its excess use of foreign words. *The Japan Times*.

Otake, M. P. (2008). Gairaigo - Remodelling language to fit Japanese. *Tokyo Seitoku University Faculty of Humanities Bulletin*, *15*, 87–101.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, *26*, 47–104.

Preston, D. R., & Yamagata, A. (2004). Katakana representation of English loanwords: Mora conservation and variable learner strategies. *Journal of Sociolinguistics*, *8*(3), 359–379.

Quackenbush, H. C. (1977). English Loanwords in Japanese: Why Are They Difficult for English-Speaking Students? *Journal of the Association of Teachers of Japanese*, *12*(2/3), 149–173.

Rebuck, M. (2002). The Function of English Loanwords in Japanese. *NUCB Journal of Language Culture and Communication*, *4*(1), 53–64. Retrieved from

Richards, K. (2003). *Qualitative Inquiry in TESOL*. London: Palgrave Macmillan UK.

Ringbom, H. (2007a). *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters Ltd.

Ringbom, H. (2007b). The importance of cross-linguistic similarities. *The Language Teacher*, *31*(9), 3–5.

Rychly, P. (2008). A Lexicographer-Friendly Association Score. *Proceedings of Recent Advances in Slavonic Natural Language Processing*, 6–9.

Sanseido's concise dictionary of katakana words. (2010). Tokyo: Sanseido.

Schmidt, C. K. (2009). Loanwords in Japanese. In *Loanwords in the world's languages: A comparative handbook*. Berlin: De Gruyter Mouton.

Seargeant, P. (2005). Globalisation and reconfigured English in Japan. *World Englishes*, *24*(3), 309–319.

Seargeant, P. (2008). Ideologies of English in Japan: the perspective of policy and pedagogy. *Language Policy*, *7*(2), 121–142.

Seargeant, P. (2009). *The Idea of English in Japan: Ideology and the Evolution of a Global Language*. Clevedon: Multilingual Matters.

Sharoff, S. (2006). *Creating general purpose corpora using automated search engine queries*. *WaCky! Working papers on the Web as Corpus*. Gedit, Bologna.

Shepherd, J. W. (1995). Loanwords: A pitfall for all students. *The Internet TESL Journal*, *2*(2), 1–6.

Shibatani, M. (1990). *The languages of Japan*. Cambridge: Cambridge University Press.

Shibatani, M., & Kageyama, T. (2016). Introduction to the handbooks of Japanese language and linguistics. In M. Minami (Ed.), *Handbook of Japanese applied linguistics* (pp. vii–xxxiv). Boston, M.A.: Walter de Gruyter Inc.

Silverman, D. (2006). *Interpreting Qualitative Data* (3rd ed.). London: Sage Publications Ltd.

Simon-Maeda, A. (1995). Language awareness: Use/misuse of loan-words in the English language in Japan. *The Internet TESL Journal*, *1*(2), 1–5.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Smith, M. (2015a). Word Categories. In J. R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp. 1–23). Oxford: Oxford University Press.

Smith, M. (2015b). Word Categories. In J. R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp. 1–23). Oxford: Oxford University Press.

Srdanovic, I. (2014). Corpus-based collocation research targeted at Japanese language learners. *Acta Linguistica Asiatica*, *4*(2), 25–36.

Srdanovic, I., Ida, N., Shigemori-Bucar, C., Kilgarriff, A., & Kovar, V. (2011). Japanese Word Sketches: Advances and Problems. *Acta Linguistica Asiatica*, *1*(2), 63–82.

Stanlaw, J. (2004). *Japanese English: Language and Culture Contact*. Hong Kong: Hong Kong University Press.

Stanlaw, J. (2010). Japanese-English Language and Culture Contact: How Much English Is Found in Japanese, How Is It Used, and Why Should Students and Teachers Care? *Journal CAJLE*, *11*, 40–59.

Strevens, P. (1980). *Teaching English as an international language: From practice to principle*. Oxford: Pergamon Press.

Stubbe, R., & Yokomitsu, H. (2012). English loanwords in Japanese and the JACET8000. *Vocabulary Education & Research Bulletin*, *1*(1), 10–11.

Stubbs, M. (1996). *Text and corpus analysis*. Oxford, England: Blackwell Publishers Ltd.

Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishing.

Takashi, K. (1990). A sociolinguistic analysis of English borrowings in Japanese advertising texts. *World Englishes*, *9*(3), 327–341.

Tamaoka, K., & Miyaoka, Y. (2003). The cognitive processing of Japanese loanwords in katakana. *Japanese Psychological Research*, *45*(2), 69–79.

Tanikawa, M. (2013, August 25). Japan's education minister aims to foster global talent. *New York Times*. Retrieved from http://www.nytimes.com/2013/08/26/world/asia/Japans-Education-Minister-Aims-to-Foster-Global-Talents.html?_r=0

Thomas, J. (2017). *Discovering English with Sketch Engine* (2nd ed.). Brno: Versatile.

Thomason, S. (2001). *Language contact: An introduction*. Washington: Georgetown University Press.

Thomason, S., & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. Los Angeles, CA: University of California Press.

Tomoda, T. (2002). Defining the concept of loanword in Japanese: Results from a survey of 860 respondents. *Monash University Linguistics Papers*, *2*(2), 3–14.

Tono, Y., Yamazaki, M., & Maekawa, K. (2013). *A frequency dictionary of Japanese: Core vocabulary for learners*. Abingdon, England: Routledge.

Uffmann, C. (2015). Loanword Adaptation. *The Oxford Handbook of Historical Phonology*, (April 2018), 1–25.

Van Coetsem, F. (1988). *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris Publications.

Van Hout, R., & Muysken, P. (1994). Modeling lexical borrowability. *Language Variation and Change*, *6*, 39–62.

Vos, F. (1963). Dutch influences on the Japanese language. *Lingua*, *12*, 341–388.

Walker, R. (2009). Barrier to communication? The use of katakana English by Japanese students. *Bulletin of Tokyo Women's University*, *60*(1), 77–88.

Webb, J. H. . (1990). *A guide to modern Japanese loanwords*. Tokyo: The Japan Times.

Weinreich, U. (1953). *Languages in contact: Findings and problems*. New York: Mouton.

Whitney, A. D. (1881). On mixture in language. *Transactions of the American Philosophical Association*, *12*, 1–26.

Winford, D. (2003). *An introduction to contact linguistics*. Malden, MA: Blackwell.

Winford, D. (2010). Contact and borrowing. In R. Hickey (Ed.), *The handbook of language contact* (pp. 171–187). Chichester: John Wiley & Sons.

Winter-Froemel, E. (2008). Studying loanwords and loanword integration: Two criteria of conformity. *Newcastle Working Papers in Linguistics*, *14*, 156–176.

Zenner, E., Dirk, S., & Geeraerts, D. (2014). What makes a catchphrase catchy? Possible determinants in the borrowability of English catchphrases in Dutch. In *New perspectives on lexical borrowing: Onomasiological, methodological, and phraseological innovations* (pp. 41–64). Berlin: Walter de Gruyter Inc.

Zenner, E., & Kristiansen, G. (2014). Introduction: Onomasiological, methodological and phraseological perspectives on lexical borrowing. In E. Zenner & G. Kristiansen (Eds.), *New perspectives on lexical borrowing: Onomasiological, methodological, and phraseological innovations* (pp. 1–17). Berlin: De Gruyter Mouton.

Appendices

Appendix 1: MeCab part-of-speech tagset used in the jpTenTen11 corpus

Copied from: https://www.sketchengine.eu/tagset-jp-mecab/

Japanese (original)	English	English (description)
代名詞	Pron	pronoun
副詞	Adv	adverb
助動詞	Aux	auxiliary_verb
助詞-係助詞	P.bind	particle(binding)
助詞-副助詞	P.adv	particle(adverbial)
助詞-接続助詞	P.conj	particle(conjunctive)
助詞-格助詞	P.case	particle(case)
助詞-準体助詞	P.nom	particle(nominal)
助詞-終助詞	P.fin	particle(phrase_final)
動詞-一般	V.g	verb(general)
動詞-非自立可能	V.bnd	verb(bound)
名詞-助動詞語幹	N.aux	noun(auxiliary)
名詞-固有名詞-一般	N.prop.g	noun(proper.general)
名詞-固有名詞-人名-一般	N.prop.n.g	noun(proper.name.general)
名詞-固有名詞-人名-名	N.prop.n.f	noun(proper.name.firstname)
名詞-固有名詞-人名-姓	N.prop.n.s	noun(proper.name.surname)
名詞-固有名詞-地名-一般	N.prop.p.g	noun(proper.place.general)
名詞-固有名詞-地名-国	N.prop.p.c	noun(proper.place.country)
名詞-数詞	N.num	noun(numeral)
名詞-普通名詞-サ変可能	N.c.vs	noun(common.verbal_suru)
名詞-普通名詞-サ変形状詞可能	N.c.vs.ana	noun(common.verbal.adjectival)
名詞-普通名詞-一般	N.c.g	noun(common.general)
名詞-普通名詞-副詞可能	N.c.adv	noun(common.adverbial)
名詞-普通名詞-助数詞可能	N.c.count	noun(common.counter)
名詞-普通名詞-形状詞可能	N.c.ana	noun(common.adjectival)
形容詞-一般	Ai.g	adjective_i(general)
形容詞-非自立可能	Ai.bnd	adjective_i(bound)
形状詞-タリ	Ana.tari	adjectival_noun(tari)
形状詞-一般	Ana.g	adjectival_noun(general)
形状詞-助動詞語幹	Ana.aux	adjectival_noun(auxiliary)
感動詞-フィラー	Interj.fill	interjection(filler)
感動詞-一般	Interj.g	interjection(general)
接尾辞-動詞的	Suff.v	suffix(verbal)
接尾辞-名詞的-サ変可能	Suff.n.vs	suffix(nominal.verbal_suru)
接尾辞-名詞的-一般	Suff.n.g	suffix(nominal.general)
接尾辞-名詞的-副詞可能	Suff.n.adv	suffix(nominal.adverbial)
接尾辞-名詞的-助数詞	Suff.n.count	suffix(nominal.counter)
接尾辞-形容詞的	Suff.ai	suffix(adjective_i)
接尾辞-形状詞的	Suff.ana	suffix(adjectival_noun)
接続詞	Conj	conjunction
接頭辞	Pref	prefix
空白	Ws	whitespace
補助記号-ＡＡ-一般	Supsym.aa.g	supplementary_symbol(ascii_art.general)
補助記号-ＡＡ-顔文字	Supsym.aa.e	supplementary_symbol(ascii_art.emoticon)
補助記号-一般	Supsym.g	supplementary_symbol(general)
補助記号-句点	Supsym.p	supplementary_symbol(period)
補助記号-括弧閉	Supsym.bo	supplementary_symbol(bracketopen)
補助記号-括弧開	Supsym.bc	supplementary_symbol(bracketclose)
補助記号-読点	Supsym.c	supplementary_symbol(comma)
	Supsym.q	supplementary_symbol(quotes)
記号-一般	Sym.g	symbol(general)
記号-文字	Sym.ch	symbol(character)
連体詞	Adn	adnominal
	Unknown	unknown
*	Empty	empty

Appendix 2: Full list of 587 English loanwords

	English Loanword	English Translation
1	ページ	page
2	コメント	comment
3	システム	system
4	サービス	service
5	ブログ	blog
6	テレビ	tv
7	センター	centre
8	ゲーム	game
9	データ	data
10	バック	back
11	ポイント	point
12	アップ	up
13	タイプ	type
14	デザイン	design
15	ホテル	hotel
16	チーム	team
17	イメージ	image
18	グループ	group
19	モデル	model
20	ソフト	soft
21	ニュース	news
22	セット	set
23	イベント	event
24	トップ	top
25	スポーツ	sports
26	メーカー	maker
27	インターネット	internet
28	ビジネス	business
29	メンバー	member
30	シリーズ	series
31	コース	course
32	プログラム	program
33	ライブ	live
34	カメラ	camera
35	クラス	class
36	シーン	scene
37	ドラマ	drama
38	メディア	media
39	ネットワーク	network
40	ドア	door
41	プレー	play
42	スタイル	style
43	タイトル	title
44	ユーザー	user
45	アクセス	access
46	プロジェクト	project
47	スタート	start
48	バランス	balance
49	コスト	cost
50	タイム	time
51	ブランド	brand
52	ビデオ	video
53	リスク	risk
54	テスト	test
55	スーパー	super
56	ガラス	glass
57	ショップ	shop
58	メニュー	menu
59	オープン	open
60	デジタル	digital
61	メッセージ	message
62	ワイン	wine
63	エンジン	engine
64	テーブル	table
65	プレゼント	present
66	フリー	free
67	レビュー	review
68	サポート	support
69	パワー	power
70	ベッド	bed
71	マンション	mansion
72	ルール	rules
73	アルバム	album
74	ランキング	rankings
75	クリスマス	Christmas
76	コンピューター	computer
77	オリジナル	original
78	コーナー	corner
79	パターン	pattern
80	ツアー	tour
81	スピード	speed
82	ストレス	stress
83	カット	cut
84	バッグ	bag
85	コミュニケーション	communication
86	シャツ	shirt
87	マーク	mark
88	ラジオ	radio
89	スター	star
90	プラス	plus
91	モード	mode
92	アイテム	items
93	カップ	cup
94	チャンス	chance
95	ツール	tool
96	ブルー	blue
97	スペース	space
98	コピー	copy
99	ブラック	black
100	コントロール	control
101	ストーリー	story
102	ピアノ	piano
103	ボランティア	volunteer
104	カバー	cover
105	ステージ	stage
106	サイド	side
107	ダイエット	diet
108	ケーキ	cake
109	クリーム	cream
110	ピンク	pink
111	ショー	show
112	サーバー	server
113	レポート	report
114	ノート	note
115	アドバイス	advice
116	シーズン	season
117	プラン	plan
118	ゴルフ	golf
119	パーティー	party
120	プロセス	process

121	バイク	bike
122	ハウス	house
123	ダンス	dance
124	インタビュー	interview
125	シンプル	simple
126	ヒット	hit
127	リーグ	league
128	カレー	curry
129	リアル	real
130	クリア	clear
131	トラブル	trouble
132	ショック	shock
133	ホワイト	white
134	コンサート	concert
135	キャラクター	character
136	ソフトウェア	software
137	テキスト	text
138	ポケット	pocket
139	アドレス	address
140	ウイルス	virus
141	セミナー	seminar
142	エリア	area
143	ギター	guitar
144	グリーン	green
145	マイナス	minus
146	バージョン	version
147	アート	art
148	ボックス	box
149	タイミング	timing
150	ゴール	goal
151	ドライブ	drive
152	ルート	root
153	オイル	oil
154	チケット	ticket
155	ガイド	guide
156	コンテンツ	content
157	ワード	word
158	リング	ring
159	トレーニング	training
160	マシン	machine
161	アウト	out
162	パック	pack
163	タクシー	taxi
164	パンツ	pants
165	ニーズ	needs
166	ファッション	fashion
167	レンズ	lens
168	スープ	soup
169	ワーク	work
170	セックス	sex
171	テープ	tape
172	インストール	installation
173	ダウンロード	download
174	ワールド	world
175	メリット	merit
176	ナンバー	number
177	ハード	hard
178	スキル	skill
179	パート	part
180	マッサージ	massage
181	コラム	column
182	リリース	release
183	オフィス	office
184	ルーム	room

185	ニュー	new
186	オークション	auction
187	サイン	sign
188	ウェブ	web
189	アプリケーション	application
190	リズム	rhythm
191	スクール	school
192	スキー	skiing
193	マスター	master
194	ライフ	life
195	ケーブル	cable
196	カウンター	counter
197	カテゴリー	category
198	チーズ	cheese
199	ステップ	step
200	チャンネル	channel
201	キャンペーン	campaign
202	レッド	red
203	レッスン	lessons
204	ペース	pace
205	アクション	action
206	ポスト	post
207	ゲスト	guest
208	ロボット	robot
209	オプション	options
210	オレンジ	orange
211	パーツ	parts
212	ブック	book
213	キャンプ	camp
214	グッズ	goods
215	トマト	tomato
216	パネル	panel
217	ドライバー	driver
218	ベルト	belt
219	ショッピング	shopping
220	レコード	record
221	ラスト	last
222	メモリー	memory
223	プール	pool
224	スーツ	suit
225	フィルム	film
226	ダブル	double
227	プリント	print
228	タイヤ	tire
229	ジャケット	jacket
230	アイディア	idea
231	ヘッド	head
232	サラダ	salad
233	スイッチ	switch
234	ディスク	disk
235	スケジュール	schedule
236	キッチン	kitchen
237	タオル	towel
238	ロング	long
239	シングル	single
240	パフォーマンス	performance
241	アプローチ	approach
242	フォーム	form
243	マップ	map
244	マニュアル	manual
245	アイス	ice
246	ボード	board
247	サークル	circle
248	チェーン	chain

249	オーナー	owner
250	スタジオ	studio
251	ランド	land
252	ゲット	get
253	フロント	front
254	パッケージ	package
255	タッチ	touch
256	ショート	short
257	モニター	monitor
258	スカート	skirt
259	グローバル	global
260	レシピ	recipe
261	ビタミン	vitamin
262	マーケット	market
263	センス	sense
264	サンプル	sample
265	トーク	talk
266	リサイクル	recycling
267	レンタル	rental
268	ドレス	dress
269	マウス	mouse
270	ジャンプ	jump
271	チャレンジ	challenge
272	エピソード	episode
273	フィールド	field
274	サウンド	sound
275	アピール	appeal
276	ダメージ	damage
277	ランク	rank
278	スタンド	stand
279	シナリオ	scenario
280	テニス	tennis
281	シルバー	silver
282	デート	date
283	テンション	tension
284	ジャズ	jazz
285	ミルク	milk
286	バブル	bubble
287	マーケティング	marketing
288	コーチ	coach
289	シェア	share
290	クール	cool
291	パートナー	partners
292	パーク	park
293	ショット	shot
294	オーバー	over
295	ユニット	unit
296	スポット	spot
297	スペシャル	special
298	ジュース	juice
299	コミュニティー	community
300	ゴールド	gold
301	マガジン	magazine
302	インチ	inch
303	グラフ	graph
304	マイク	microphone
305	ラベル	label
306	スムーズ	smooth
307	バッテリー	battery
308	マネージャー	manager
309	ポスター	poster
310	ビッグ	big
311	キング	king
312	ガソリン	gasoline

313	シャワー	shower
314	カップル	couple
315	クラシック	classic
316	チョコレート	chocolate
317	ボリューム	volume
318	ポジション	position
319	アンド	and
320	ノウハウ	know-how
321	マルチ	multi
322	セール	sale
323	クレジット	credit
324	レーザー	laser
325	テクニック	technique
326	ユーロ	euro
327	コンセプト	concept
328	プロフィール	profile
329	ファミリー	family
330	ヒント	hint
331	クライアント	client
332	グレー	gray
333	エントリー	entry
334	タンク	tank
335	デザイナー	designer
336	トンネル	tunnel
337	アレンジ	arrange
338	オーダー	order
339	レンジ	range
340	ブレーキ	brake
341	ファースト	fast
342	ドラゴン	dragon
343	スクリーン	screen
344	ペーパー	paper
345	キーボード	keyboard
346	ピーク	peak
347	タウン	town
348	コンクリート	concrete
349	デッキ	deck
350	マイクロ	micro
351	コミック	comic
352	ハンドル	handle
353	コンパクト	compact
354	プラスチック	plastic
355	マスク	mask
356	ミサイル	missile
357	ターン	turn
358	ストレート	straight
359	ブラウザ	browser
360	インパクト	impact
361	デスク	desk
362	ベンチ	bench
363	ナイフ	knife
364	エンド	end
365	ホット	hot
366	ドーム	dome
367	ボーイ	boy
368	エラー	error
369	マナー	manners
370	シフト	shift
371	カタログ	catalogue
372	ターゲット	target
373	ラッキー	lucky
374	キャスト	cast
375	クリニック	clinic
376	インテリア	interior

377	インフルエンザ	influenza
378	カロリー	calorie
379	スピーカー	speaker
380	リビング	living
381	リラックス	relax
382	サイクル	cycle
383	フィルター	filter
384	パスワード	password
385	ベンチャー	venture
386	カレンダー	calendar
387	アンテナ	antenna
388	バトル	battle
389	プレス	press
390	フォト	photo
391	フォーラム	forum
392	キット	kit
393	タレント	talent
394	カーテン	curtain
395	リットル	litre
396	ボート	boat
397	ファンド	fund
398	スライド	slide
399	マイル	mile
400	ライバル	rival
401	ギャラリー	gallery
402	リゾート	resort
403	ボーナス	bonus
404	ゾーン	zone
405	マラソン	marathon
406	エアコン	air conditioning
407	オブジェクト	object
408	プレート	plate
409	アクセサリー	accessories
410	ビーチ	beach
411	ミュージック	music
412	キャンセル	cancel
413	ストップ	stop
414	メカニズム	mechanism
415	エレベーター	elevator
416	レモン	lemon
417	オーディオ	audio
418	スマート	smart
419	ドラム	drum
420	パイプ	pipe
421	ビジョン	vision
422	タワー	tower
423	プライベート	private
424	シュート	shoot
425	マジック	magic
426	ポップ	pop
427	ライセンス	license
428	ゲート	gate
429	オート	auto
430	テクノロジー	technology
431	チャット	chat
432	フルーツ	fruit
433	バナナ	banana
434	ストーン	stone
435	ハッピー	happy
436	ローカル	local
437	バックアップ	backup
438	トレード	trade
439	コンビ	combi
440	バイオ	bio

441	フロー	flow
442	アナログ	analogue
443	ラウンド	round
444	ビニール	vinyl
445	ジェット	jet
446	ライター	writer
447	センサー	sensor
448	スコア	score
449	ポート	port
450	リフォーム	renovation
451	テンポ	tempo
452	キャッチ	catch
453	ハンド	hand
454	デザート	dessert
455	セブン	seven
456	プライバシー	privacy
457	アップル	apple
458	シール	seal
459	スケール	scale
460	ネーム	name
461	ガード	guard
462	ダイヤモンド	diamond
463	モーター	motor
464	ラップ	wrap
465	マクロ	macro
466	ステーション	station
467	ヒーロー	hero
468	ホスト	host
469	デバイス	device
470	ソング	song
471	フォロー	follow
472	カウンセリング	counselling
473	パスタ	pasta
474	スタンダード	standard
475	ジュニア	junior
476	ビット	bit
477	メンテナンス	maintenance
478	ハーブ	herb
479	ジャーナリスト	journalist
480	テント	tent
481	ユニーク	unique
482	シンポジウム	symposium
483	マニア	mania
484	ドライ	dry
485	アメリカン	American
486	フロア	floor
487	ストック	stock
488	ドリンク	drink
489	ミッション	mission
490	デパート	depart
491	ドクター	doctor
492	シミュレーション	simulation
493	トータル	total
494	ライオン	lion
495	ロビー	lobby
496	ギャップ	gap
497	マネー	money
498	プラザ	plaza
499	キャベツ	cabbage
500	レイアウト	layout
501	リクエスト	request
502	エース	ace
503	チキン	chicken
504	ガイドライン	guidelines

505	バトン	baton
506	ロープ	rope
507	クロス	cross
508	インク	ink
509	パンフレット	pamphlet
510	インター	inter
511	ミネラル	mineral
512	コンサルタント	consultants
513	ウォーター	water
514	ストリート	street
515	ミックス	mix
516	プロデューサー	producer
517	ベテラン	veteran
518	グラフィック	graphics
519	プライド	pride
520	モダン	modern
521	ガーデン	garden
522	フォント	font
523	チャンピオン	champion
524	スパイ	spy
525	トレンド	trend
526	シャッター	shutter
527	スタジアム	stadium
528	パニック	panic
529	セカンド	second
530	レギュラー	regular
531	エリート	elite
532	コンタクト	contact
533	ストア	store
534	クイズ	quiz
535	カーブ	curve
536	コイン	coin
537	ドラッグ	drag
538	オーケストラ	orchestra
539	マザー	mother
540	パンチ	punch
541	アクセント	accent
542	チャート	chart
543	スリー	three
544	シンボル	symbol
545	セクション	section
546	コンテスト	contest

547	パイロット	pilot
548	アニメーション	animation
549	ムード	mood
550	アカデミー	academy
551	レール	rail
552	フォーマット	format
553	キック	kick
554	ツリー	tree
555	シャープ	sharp
556	トライ	try
557	カメラマン	photographer
558	プレッシャー	pressure
559	メソッド	method
560	スクリプト	script
561	クリーン	clean
562	パスポート	passport
563	ループ	loop
564	ドキュメント	document
565	ハードディスク	hard disk
566	ロマン	roman
567	ミステリー	mystery
568	インターフェース	interface
569	ターミナル	terminal
570	インターナショナル	international
571	クリップ	clip
572	グラウンド	ground
573	ライス	rice
574	ピッチ	pitch
575	フェスティバル	festival
576	ラリー	rally
577	ネクタイ	tie
578	ライフスタイル	lifestyle
579	カルチャー	culture
580	ダイレクト	direct
581	カウンセラー	counsellor
582	ブリッジ	bridge
583	ニュアンス	nuance
584	アマチュア	amateur
585	タイル	tile
586	アナウンサー	announcer
587	フェリー	ferry

Appendix 3: Full list of 130 native and Sino-Japanese nouns

	Non-Loanword	English Translation
1	事	thing
2	物	stuff
3	時	time
4	人	man
5	今	now
6	所	place
7	自分	myself
8	中	during
9	後	rear
10	方	person
11	訳	translation
12	本当	truth
13	為	for
14	場合	case
15	話	story
16	日本	Japan
17	前	before
18	子供	children
19	気	spirit
20	内	inside
21	感じ	feeling
22	一	one
23	そう	so
24	二	two
25	必要	necessary
26	仕事	work
27	余り	remainder
28	皆	all
29	方	person
30	次	next
31	問題	problem
32	目	eye
33	頃	area/page
34	上	up
35	他	other
36	家	house
37	日	day
38	一人	one person
39	人間	human
40	時間	time
41	言葉	word
42	まだ	not yet
43	手	hand
44	意味	meaning
45	形	form
46	三	three
47	最初	first
48	間	while
49	最近	recently
50	友達	friend
51	一緒	together
52	生活	life
53	国	country
54	現在	current
55	気持ち	feeling
56	会社	company
57	実際	in fact
58	先生	teacher
59	二人	two people
60	心	heart
61	金	money
62	顔	face
63	町	town
64	今日	today
65	昔	old days
66	全て	all
67	子	child
68	水	water
69	当時	at that time
70	場所	place
71	声	voice
72	最後	last
73	車	car
74	今度	now
75	体	body
76	男	man
77	女性	female
78	学校	school
79	先	ahead
80	世界	world
81	状態	state
82	相手	partner
83	母	mother
84	以上	the above
85	関係	relationship
86	四	four
87	店	shop
88	頭	head
89	電話	phone
90	本	book
91	夜	night
92	者	person
93	親	parent
94	名前	name
95	家族	family
96	部分	portion
97	一度	one time
98	結果	result
99	状況	situation
100	時代	age
101	今回	this time
102	人生	life
103	通り	street
104	内容	contents
105	経験	experience
106	駅	station
107	映画	movies
108	身	body

109	客	customer
110	質問	question
111	猫	cat
112	向こう	beyond
113	同時	simultaneous
114	六	six
115	場	place
116	大人	adult
117	期待	expectation
118	事実	fact
119	一体	unity

120	島	island
121	増加	increase
122	判断	judgment
123	両親	parents
124	家庭	home
125	船	boat
126	命	life
127	流れ	flow
128	予定	plans
129	馬	horse
130	学生	student

Appendix 4: Full list of loanwords with [particle] as the first most preferred gramrel

	English Loanword	English Translation	Difference in G1 and G2 Frequency
1	シーン	scene	4.57
2	テンポ	tempo	3.92
3	セクション	section	3.90
4	カップル	couple	3.36
5	コーナー	corner	3.27
6	センス	sense	3.03
7	スタイル	style	2.95
8	パターン	pattern	2.93
9	ギャップ	gap	2.90
10	コンセプト	concept	2.80
11	ロビー	lobby	2.78
12	カメラマン	photographer	2.70
13	キャベツ	cabbage	2.65
14	インパクト	impact	2.62
15	マネージャー	manager	2.55
16	トラブル	trouble	2.55
17	ブラウザ	browser	2.54
18	エレベーター	elevator	2.49
19	プール	pool	2.46
20	メニュー	menu	2.42
21	ニュアンス	nuance	2.35
22	デパート	department store	2.34
23	スペース	space	2.32
24	インタビュー	interview	2.32
25	ガイドライン	guidelines	2.29
26	フォーム	form	2.27
27	モード	mode	2.22
28	プロデューサー	producer	2.21
29	スープ	soup	2.15
30	ムード	mood	2.11
31	テンション	tension	2.09
32	メンバー	member	2.05
33	カウンセラー	counselor	2.02
34	エピソード	episode	1.99
35	スカート	skirt	1.97
36	アクセント	accent	1.94
37	タイミング	timing	1.94
38	ルート	root	1.92
39	ドライバー	driver	1.91
40	ステージ	stage	1.89
41	グラフ	graph	1.89
42	レイアウト	layout	1.86
43	ストーリー	story	1.85
44	ターゲット	target	1.83
45	ピッチ	pitch	1.80
46	キーボード	keyboard	1.80
47	アルバム	album	1.78
48	ジャーナリスト	journalist	1.77
49	ツール	tool	1.76
50	オーケストラ	orchestra	1.75
51	シンポジウム	symposium	1.74
52	インターフェース	interface	1.71
53	プライド	pride	1.69
54	ポスター	poster	1.65
55	チーム	team	1.64
56	ゾーン	zone	1.64
57	グラウンド	ground	1.62
58	ベッド	bed	1.61
59	ライフスタイル	lifestyle	1.60
60	アプローチ	approach	1.59

61	ドラマ	drama	1.59
62	ストレス	stress	1.58
63	リズム	rhythm	1.58
64	アナウンサー	announcer	1.58
65	メリット	merit	1.57
66	プロセス	process	1.55
67	イメージ	image	1.55
68	スケジュール	schedule	1.55
69	コンビ	combi	1.55
70	カメラ	camera	1.55
71	テント	tent	1.54
72	エリア	area	1.54
73	リクエスト	request	1.53
74	テープ	tape	1.53
75	プラン	plan	1.53
76	コンサート	concert	1.52
77	クラス	class	1.50
78	パンフレット	pamphlet	1.49
79	パスタ	pasta	1.49
80	シナリオ	scenario	1.48
81	プログラム	program	1.48
82	パーツ	parts	1.47
83	オブジェクト	object	1.44
84	パートナー	partners	1.44
85	スタジアム	stadium	1.43
86	アイテム	items	1.41
87	タクシー	taxi	1.41
88	クロス	cross	1.41
89	コラム	column	1.41
90	ルール	rules	1.41
91	カーブ	curve	1.41
92	ライター	writer	1.40
93	カウンセリング	counseling	1.40
94	コース	course	1.39
95	テクニック	technique	1.38
96	デザイン	design	1.36
97	パフォーマンス	performance	1.35
98	トンネル	tunnel	1.34
99	パンチ	punch	1.33
100	レッスン	lessons	1.31
101	ケーキ	cake	1.28
102	テーブル	table	1.28
103	ソング	song	1.28
104	アプリケーション	application	1.27
105	カウンター	counter	1.26
106	マシン	machine	1.25
107	デバイス	device	1.24
108	スーツ	suit	1.24
109	デッキ	deck	1.23
110	タイプ	type	1.23
111	ジュース	juice	1.21
112	バランス	balance	1.19
113	メソッド	method	1.19
114	ポジション	position	1.19
115	オーナー	owner	1.18
116	シャツ	shirt	1.18
117	アドレス	address	1.18
118	スポット	spot	1.17
119	パスポート	passport	1.15
120	ビジョン	vision	1.15
121	レポート	report	1.14
122	ニーズ	needs	1.14
123	クライアント	client	1.13
124	アレンジ	arrange	1.13

125	スクリプト	script	1.13
126	メカニズム	mechanism	1.13
127	リスク	risk	1.11
128	ナイフ	knife	1.08
129	エラー	error	1.06
130	マイク	microphone	1.05
131	ツリー	tree	1.05
132	ハードディスク	hard disk	1.05
133	デザート	dessert	1.04
134	ファンド	fund	1.02
135	スピーカー	speaker	1.02
136	デート	date	1.02
137	キャラクター	character	1.02
138	コメント	comment	1.02
139	フォント	font	1.01
140	シール	seal	1.01
141	マーク	mark	1.01
142	ユーザー	user	0.99
143	プレゼント	present	0.99
144	マナー	manners	0.98
145	パーティー	party	0.98
146	タオル	towel	0.98
147	テクノロジー	technology	0.98
148	タイル	tile	0.96
149	タイトル	title	0.95
150	シリーズ	series	0.95
151	アイディア	idea	0.94
152	イベント	event	0.94
153	プレッシャー	pressure	0.93
154	タイヤ	tire	0.93
155	ギター	guitar	0.91
156	オプション	options	0.91
157	フィルター	filter	0.91
158	チャート	chart	0.90
159	ドア	door	0.90
160	プレート	plate	0.90
161	ネクタイ	tie	0.89
162	コンテスト	contest	0.88
163	エアコン	air conditioning	0.88
164	ショー	show	0.87
165	オークション	auction	0.87
166	ブリッジ	bridge	0.87
167	モニター	monitor	0.86
168	レンジ	range	0.86
169	ラップ	wrap	0.86
170	コンテンツ	content	0.86
171	レシピ	recipe	0.85
172	スコア	score	0.85
173	パンツ	pants	0.85
174	バイク	bike	0.84
175	ページ	page	0.84
176	チャンス	chance	0.84
177	シュート	shoot	0.82
178	ヒント	hint	0.82
179	マスク	mask	0.82
180	カーテン	curtain	0.81
181	プロジェクト	project	0.81
182	スケール	scale	0.81
183	キック	kick	0.79
184	タレント	talent	0.78
185	コンサルタント	consultants	0.78
186	クリニック	clinic	0.77
187	ワイン	wine	0.76
188	グッズ	goods	0.75

189	ボード	board	0.75
190	システム	system	0.75
191	スキル	skill	0.74
192	レンズ	lens	0.74
193	ショット	shot	0.74
194	セミナー	seminar	0.74
195	マンション	mansion	0.72
196	スイッチ	switch	0.72
197	ピーク	peak	0.71
198	サーバー	server	0.71
199	ショップ	shop	0.71
200	サービス	service	0.70
201	ユニット	unit	0.70
202	ノウハウ	know-how	0.69
203	モデル	model	0.68
204	カレンダー	calendar	0.67
205	グループ	group	0.66
206	ペース	pace	0.66
207	ライバル	rival	0.66
208	フェリー	ferry	0.65
209	メッセージ	message	0.64
210	マーケット	market	0.64
211	メンテナンス	maintenance	0.63
212	ミッション	mission	0.63
213	ボート	boat	0.62
214	バージョン	version	0.61
215	ゲスト	the guests	0.61
216	パート	part	0.59
217	コミュニケーション	communication	0.59
218	ベンチ	bench	0.59
219	データ	data	0.59
220	スピード	speed	0.59
221	ツアー	tour	0.59
222	ブログ	blog	0.59
223	ボックス	box	0.58
224	パスワード	password	0.58
225	サラダ	salad	0.57
226	リビング	living	0.55
227	ショック	shock	0.55
228	センサー	sensor	0.55
229	デザイナー	designer	0.55
230	ポケット	pocket	0.54
231	ジャケット	jacket	0.54
232	パネル	panel	0.52
233	バトン	baton	0.52
234	シャワー	shower	0.51
235	フォーラム	forum	0.50
236	ポイント	point	0.50
237	キャスト	cast	0.48
238	エンジン	engine	0.48
239	マニュアル	manual	0.48
240	トレーニング	training	0.47
241	ドーム	dome	0.47
242	マップ	map	0.47
243	ドライブ	drive	0.47
244	キャンプ	camp	0.46
245	パイロット	pilot	0.46
246	ストア	store	0.46
247	チケット	ticket	0.46
248	インフルエンザ	influenza	0.45
249	テスト	test	0.45
250	クリーム	cream	0.44
251	フィールド	field	0.44
252	パッケージ	package	0.43

253	コーチ	coach	0.42
254	アンテナ	antenna	0.41
255	スクリーン	screen	0.41
256	メディア	media	0.41
257	フォーマット	format	0.40
258	シーズン	season	0.39
259	ドレス	dress	0.39
260	コスト	cost	0.38
261	キット	kit	0.38
262	ギャラリー	gallery	0.38
263	チャンネル	channel	0.38
264	シェア	share	0.37
265	ビタミン	vitamin	0.37
266	フォロー	follow	0.37
267	スタジオ	studio	0.37
268	ダメージ	damage	0.36
269	ゴール	goal	0.36
270	サークル	circle	0.35
271	フロア	floor	0.35
272	コンクリート	concrete	0.34
273	ターミナル	terminal	0.34
274	ループ	loop	0.34
275	フィルム	the film	0.34
276	バックアップ	backup	0.32
277	エントリー	entry	0.32
278	ニュース	news	0.31
279	ロープ	rope	0.31
280	マウス	mouse	0.30
281	カバー	cover	0.30
282	ゲーム	game	0.29
283	パイプ	pipe	0.29
284	ネーム	name	0.29
285	フェスティバル	festival	0.29
286	コントロール	control	0.28
287	スタンド	stand	0.27
288	カタログ	catalog	0.26
289	ジャンプ	jump	0.26
290	ハンドル	handle	0.26
291	サイクル	cycle	0.26
292	ダンス	dance	0.25
293	コンピューター	computer	0.25
294	パック	pack	0.25
295	ボリューム	volume	0.25
296	ホテル	hotel	0.25
297	ヒーロー	hero	0.24
298	ノート	note	0.24
299	ベルト	belt	0.24
300	トーク	talk	0.23
301	リング	ring	0.22
302	キャンペーン	campaign	0.21
303	アドバイス	advice	0.21
304	ルーム	room	0.21
305	トレンド	trend	0.21
306	チョコレート	chocolate	0.19
307	テキスト	text	0.18
308	ドラム	drum	0.18
309	ドリンク	drink	0.18
310	ストーン	stone	0.18
311	サイン	sign	0.17
312	セックス	sex	0.17
313	ラリー	rally	0.16
314	キッチン	kitchen	0.16
315	ミサイル	missile	0.16
316	マイル	mile	0.15

317	プレー	play	0.15
318	サウンド	sound	0.15
319	ワード	word	0.14
320	ピアノ	piano	0.14
321	マッサージ	massage	0.13
322	トマト	tomato	0.13
323	ターン	turn	0.12
324	ライス	rice	0.12
325	カレー	curry	0.11
326	リリース	release	0.11
327	アップ	up	0.10
328	アクセサリー	accessories	0.10
329	ロボット	robot	0.10
330	ブレーキ	brake	0.10
331	ドキュメント	document	0.09
332	ガラス	glass	0.08
333	ピンク	pink	0.08
334	ライブ	live	0.08
335	ソフトウェア	software	0.08
336	センター	center	0.07
337	エース	ace	0.07
338	キャンセル	cancel	0.07
339	オイル	oil	0.07
340	ボランティア	volunteer	0.06
341	カロリー	calorie	0.06
342	ステップ	step	0.06
343	テレビ	tv	0.06
344	ウイルス	virus	0.06
345	アニメーション	animation	0.05
346	マラソン	marathon	0.05
347	コピー	copy	0.04
348	メーカー	maker	0.04
349	シンボル	symbol	0.04
350	ミルク	milk	0.03
351	シミュレーション	simulation	0.03
352	ガード	guard	0.02
353	カテゴリー	category	0.02
354	シャッター	shutter	0.01
355	ストック	stock	0.01
356	パニック	panic	0.01
357	ライオン	lion	0.00
358	バッグ	bag	0.00

Appendix 5: Full list of loanwords with [noun/noun] as the first most preferred gramrel

	English Loanword	English Translation	Difference in G1 and G2 Frequency
1	アンド	and	32.69
2	バイオ	bio	16.26
3	ニュー	new	12.48
4	オート	auto	12.18
5	フォト	photo	11.51
6	マルチ	multi	9.83
7	ダブル	double	9.08
8	ロング	long	8.96
9	クレジット	credit	8.47
10	ファースト	fast	7.87
11	スリー	three	7.13
12	スペシャル	special	6.17
13	ホット	hot	6.04
14	ホワイト	white	5.97
15	ショート	short	5.13
16	ブラック	black	4.66
17	ウェブ	web	4.55
18	セカンド	second	4.44
19	リアル	real	4.24
20	デジタル	digital	3.75
21	レッド	red	3.65
22	レンタル	rental	3.42
23	ハンド	hand	3.41
24	ワールド	world	3.38
25	ローカル	local	3.31
26	アメリカン	american	3.25
27	オープン	open	3.06
28	ジェット	jet	2.97
29	アカデミー	academy	2.79
30	マイクロ	micro	2.72
31	アナログ	analog	2.65
32	ショッピング	shopping	2.61
33	リサイクル	recycling	2.59
34	シルバー	silver	2.51
35	ビニール	vinyl	2.43
36	スライド	slide	2.37
37	ハード	hard	2.36
38	マザー	mother	2.26
39	コンパクト	compact	2.23
40	インター	inter	2.16
41	ベンチャー	venture	2.15
42	ドラゴン	dragon	2.12
43	ゴルフ	golf	2.12
44	ウォーター	water	2.04
45	オーディオ	audio	2.01
46	オリジナル	original	1.94
47	スポーツ	sports	1.90
48	キャッチ	catch	1.86
49	ラッキー	lucky	1.77
50	グリーン	green	1.72
51	アイス	ice	1.72
52	レーザー	laser	1.71
53	セブン	seven	1.65
54	リゾート	resort	1.63
55	レビュー	review	1.62
56	インチ	inch	1.55
57	プレス	press	1.54
58	フリー	free	1.53
59	クリア	clear	1.48

60	シングル	single	1.45
61	アマチュア	amateur	1.44
62	ジュニア	junior	1.44
63	スーパー	super	1.44
64	クラシック	classic	1.43
65	キング	king	1.39
66	マジック	magic	1.38
67	ゴールド	gold	1.36
68	プラザ	plaza	1.34
69	ケーブル	cable	1.26
70	ダウンロード	download	1.22
71	ガーデン	garden	1.19
72	レギュラー	regular	1.17
73	アート	art	1.14
74	ミュージック	music	1.14
75	オーダー	order	1.13
76	ブルー	blue	1.10
77	カルチャー	culture	1.09
78	ヘッド	head	1.08
79	ボーイ	boy	1.07
80	クリスマス	christmas	1.06
81	フロント	front	1.06
82	アクセス	access	1.05
83	プライベート	private	1.04
84	ストリート	street	1.04
85	タウン	town	1.02
86	ファミリー	family	1.01
87	マクロ	macro	1.00
88	スタンダード	standard	0.99
89	サンプル	sample	0.98
90	ビット	bit	0.97
91	トライ	try	0.95
92	ビジネス	business	0.94
93	ライフ	life	0.92
94	ラウンド	round	0.91
95	エリート	elite	0.91
96	マイナス	minus	0.90
97	チキン	chicken	0.88
98	コイン	coin	0.87
99	ミネラル	mineral	0.85
100	アップル	apple	0.82
101	パワー	power	0.82
102	ブック	book	0.80
103	レモン	lemon	0.80
104	オーバー	over	0.78
105	テニス	tennis	0.77
106	スター	star	0.75
107	ダイヤモンド	diamond	0.75
108	セール	sale	0.73
109	コミック	comic	0.72
110	スパイ	spy	0.71
111	ファッション	fashion	0.69
112	インターナショナル	international	0.69
113	プラスチック	plastic	0.68
114	ドクター	doctor	0.68
115	バブル	bubble	0.67
116	トップ	top	0.67
117	ポスト	post	0.66
118	ポップ	pop	0.63
119	ガソリン	gasoline	0.61
120	グラフィック	graphics	0.61
121	ドラッグ	drag	0.60
122	ミックス	mix	0.59
123	モーター	motor	0.57

124	インク	ink	0.55
125	ホスト	host	0.55
126	ヒット	hit	0.52
127	プラス	plus	0.51
128	タッチ	touch	0.51
129	ランク	rank	0.50
130	トータル	total	0.49
131	バック	back	0.47
132	コンタクト	contact	0.47
133	オレンジ	orange	0.45
134	ライセンス	license	0.44
135	マーケティング	marketing	0.44
136	ビデオ	video	0.42
137	エンド	end	0.42
138	バトル	battle	0.38
139	アウト	out	0.37
140	パーク	park	0.36
141	ダイエット	diet	0.35
142	ボーナス	bonus	0.35
143	スキー	skiing	0.34
144	マネー	money	0.33
145	ユーロ	eur	0.32
146	ブランド	brand	0.31
147	ナンバー	number	0.30
148	ストレート	straight	0.30
149	グレー	gray	0.29
150	インテリア	interior	0.29
151	プライバシー	privacy	0.29
152	ミステリー	mystery	0.29
153	プリント	print	0.28
154	シャープ	sharp	0.27
155	マガジン	magazine	0.25
156	シフト	shift	0.23
157	ワーク	work	0.23
158	リラックス	relax	0.22
159	マスター	master	0.22
160	スクール	school	0.21
161	ランキング	rankings	0.21
162	フルーツ	fruit	0.20
163	タイム	time	0.19
164	ジャズ	jazz	0.17
165	デスク	desk	0.17
166	メモリー	memory	0.17
167	アクション	action	0.17
168	カップ	cup	0.16
169	タワー	tower	0.15
170	ネットワーク	network	0.13
171	ソフト	soft	0.12
172	チャット	chat	0.12
173	タンク	tank	0.11
174	チーズ	cheese	0.10
175	ビーチ	beach	0.10
176	ハーブ	herb	0.10
177	フロー	flow	0.10
178	レコード	record	0.09
179	リフォーム	renovation	0.09
180	クイズ	quiz	0.09
181	プロフィール	profile	0.09
182	バッテリー	battery	0.08
183	ペーパー	paper	0.08
184	バナナ	banana	0.08
185	チェーン	chain	0.07
186	ガイド	guide	0.06
187	ラベル	label	0.06

188	ポート	port	0.05
189	レール	rail	0.05
190	ランド	land	0.05
191	ラスト	last	0.05
192	ディスク	disk	0.05
193	ベテラン	veteran	0.04
194	クリップ	clip	0.04
195	ロマン	roman	0.04
196	リーグ	league	0.04
197	セット	set	0.04
198	サポート	support	0.04
199	リットル	liter	0.03
200	カット	cut	0.03
201	コミュニティー	community	0.03
202	ゲート	gate	0.03
203	ハウス	house	0.03
204	チャンピオン	champion	0.03
205	スタート	start	0.03
206	オフィス	office	0.02
207	サイド	side	0.02
208	インターネット	internet	0.02
209	ラジオ	radio	0.02
210	トレード	trade	0.02
211	ストップ	stop	0.01
212	ステーション	station	0.01