Tampereen yliopisto

Sharin Lehtonen

TSUNAMI, ANIME, AND *MARTIAL ARTS*

A corpus-based lexicological study of Japanese borrowings in a historical context and in six varieties of Present-day English

Faculty of Information Technology and Communication Sciences
Master's Thesis
April 2021

TIIVISTELMÄ

Sharin Lehtonen: *Tsunami*, *anime*, and *martial arts*: A corpus-based lexicological study of Japanese borrowings in a historical context and in six varieties of Present-day English
Pro gradu -tutkielma
Tampereen yliopisto
Englannin kielen ja kirjallisuuden maisteriopinnot
Huhtikuu 2021

Tämä pro gradu -tutkielma on korpuspohjainen leksikologinen tutkimus japanilaisista lainoista englannissa. Tutkimus kartoittaa lainojen jakautumista semanttisiin kenttiin sekä niiden käytön yleisyyttä kuudessa nykyenglannin varieteetissa.

Tutkimuksessa käytetään kahta aineistoa. Ensimmäinen on kerätty Oxford English Dictionary -sanakirjasta hakemalla pääsanoja, joiden etymologian jossakin vaiheessa on ollut japaninkielinen lekseemi. Pääsanat on jaettu 705 lekseemi-merkityspariin, jotka on luokiteltu semanttisiin kenttiin perustuviin pää- ja alakategorioihin. Lekseemi-merkitysparien määriä tarkastellaan kategorioittain kokonaisuudessaan sekä 50 vuoden ajanjaksoittain vuodesta 1550 lähtien ensimmäisen esiintymän vuoden perusteella. Löydökset suhteutetaan historialliseen tilanteeseen kunakin ajanjaksona.

Toinen aineisto on peräisin GloWbE-korpuksesta, josta on haettu ensimmäisen aineiston lekseemi-merkityspareja kuuden varieteetin teksteistä: kahden sisemmän ympyrän varieteetin, amerikan- ja brittienglannin, sekä neljän ulomman ympyrän varieteetin, singaporen-, filippiinin-, malesian- ja hongkonginenglannin. Kaikki osumat on tarkasteltu ja relevanttien sanaesiintymien määrät otettu talteen varieteeteittain ja kategorioittain.

Tutkimus osoittaa, että japanilaiset lainat jakautuvat semantiikan perusteella 25, joista 13:lla on yhteensä 51 alakategoriaa. Pääkategoriat YHTEISKUNTA, TAIDE JA KÄSITYÖ, RUOKA JA JUOMA, URHEILU JA KAMPPAILULAJIT, LUONTO sekä TRADITIO, MYTOLOGIA JA USKONTO sisältävät eniten lekseemi-merkityspareja ja käsittävät yhdessä 66.6% kaikista kategorioissa olevista pareista. Suurimmat frekvenssit korpusteksteissä ovat lekseemi-merkityspareilla, jotka kuuluvat pääkategorioihin RUOKA JA JUOMA, VIIHDE JA VAPAA-AIKA, LUONTO, URHEILU JA KAMPPAILULAJIT sekä YHTEISKUNTA. Varieteettien välillä on vaihtelua normalisoiduissa frekvensseissä. Singaporenenglannilla luku on 271 esiintymää miljoonan sanan tekstissä. Tämä on merkittävästi suurempi kuin muilla varieteeteilla. Muiden ulomman ympyrän varieteettien luvut ovat hyvin lähellä toisiaan, sillä ne ovat välillä 142–150 per miljoona sanaa. Normalisoitu frekvenssi on matalin sisemmän ympyrän varieteeteilla, joista amerikanenglannin frekvenssi on 65 ja brittienglannin 48 esiintymää miljoonan sanan tekstissä.

Avainsanat: englannin kieli, lainasana, leksikologia, semantiikka, korpuslingvistiikka

Table of contents

1 Introduction

When there is a need for a word referring to a new concept or a desire to avoid the connotations existing words have, there are a number of ways for a speaker to introduce new lexemes and meanings to the language. Aside from using word-formation techniques to create a new word, one may adopt a lexeme from another language. Lexical borrowings have been a way of increasing the lexicon of English ever since the inception and technically even before it, as influences from other languages such as Latin and Celtic were already being transferred into Common Germanic (Millward & Hayes 2012: 73). In the Old English period, lexical items were borrowed from Old Norse in addition to the two aforementioned languages, and the Norman conquest of 1066 brought French lexemes into Middle English (ibid.: 120–123, 145–146). More recently, in the Modern and Present-Day English periods, Italian, Spanish, Dutch, German, Russian, Hindi, and many more have become donor languages (ibid.: 324–328) due to colonialism, travel, and globalisation. One of these languages is Japanese, from which the first borrowings were adopted in the 16th century.

Previous research on Japanese borrowings in English has been conducted from a variety of perspectives. There are for example studies into the history (Tsuchihashi 1996) and morphology of Japanese borrowings (Cannon 1984), their naturalization (Doi 2013) as well as use in magazines (Carman 1991) and treatment in dictionaries (Yamamoto 2020). Analysis of lexical semantics has been included in some studies, but in many cases the focus has been in some other field of linguistics. There are a few studies such as Schultz's (2017) that are primarily lexicological, but they do not accord much attention to the historical context at the time of the adoption or to actual language use. To my knowledge, there has not been extensive lexicological research done into the Japanese borrowings by utilizing corpora despite the trend towards corpus-based studies in linguistics.

The purpose of this thesis is to explore the lexical semantics of Japanese borrowings in English from both a diachronic and a synchronic perspective. The research questions to be answered are

1. What is the distribution of Japanese borrowings in English into semantic fields during different 50-year time periods?
2. What is the distribution of Japanese borrowings used in Present-day English texts of different varieties into semantic fields?

The borrowings studied will be gathered from *The Oxford English Dictionary* (henceforth the *OED*). Their meanings and the years of first attestation will be examined to answer the first question. The borrowings from the *OED* will then be searched in the Corpus of Global Web-based English (henceforth GloWbE) in the texts of six geographical varieties of English, including two Inner Circle varieties, American English and British English, and four Outer Circle varieties, Singaporean English, Philippine English, Malaysian English, and Hong Kong English (henceforth AmE, BrE, SgE, PhE, MyE, and HKE respectively). The number of borrowings and their relevant tokens per semantic category and variety will be calculated. The data will be examined by both quantitative and qualitative methods: semantic fields will be assigned by inspecting the meanings of the borrowings, which requires qualitative analysis, while the distributions of the borrowings and tokens will be presented numerically. Any further analysis based on these will also be qualitative. The study will take a different approach from previous research, where borrowings have been treated as bearing all of the meanings that have developed since the first attestation, and will instead introduce the concept of lexeme-sense pairs, which was assessed to be suitable for a corpus-based lexicological examination of borrowings.

The structure of the study is such that Section 2 discusses the theoretical background and explains the terminology used. This is followed by Section 3, which focuses on the data and methods. Sections 4 and 5 examine the results drawn from the data so that the former section answers the first research question and the latter the second question. Section 6 closes with some concluding remarks.

2 Theoretical background

This section examines the central concepts related to the study. Section 2.1 defines the term *lexical borrowing*, describes its classification, and discusses borrowings in the history of the English language. Section 2.2 discusses the field of lexical semantics and introduces studies where the semantic fields of Japanese borrowings have been examined, and Section 2.3 recounts the language contact situations between Japan and the Inner Circle and Outer Circle regions studied in a historical context as well as gives an outline of the Englishes in the Outer Circle regions.

2.1 Lexical borrowing

2.1.1 Terminology

Lexical borrowing is the process where one language influences the lexis of another language resulting in the latter acquiring "a new word form or word meaning, or both," from the former (Durkin 2014: 8). The language from which a lexeme or a meaning is borrowed is called the *donor language* (also called the *borrowing*, *receiving, or model language*) and the language that borrows is the *recipient language* (also called the *source or replica language*) (Durkin 2014: 8; Haspelmath 2009: 37). The products of the process are also called *lexical borrowings*, or simply *borrowings*. The metaphor of borrowing is, of course, not accurate as the borrowed lexeme is not returned to the donor language. The desire to avoid the inaccurate metaphor have led to some authors adopting terms such as *transfer*, *transference*, and *copying* instead (Haspelmath 2009: 37). Despite the inaccurate metaphor, this study will use the term *borrowing*, as it is well-established in linguistics. The exception

is references to the field which researches borrowings, which will be called *loanword studies*, another well-established term.

Borrowings may be divided into multiple types based on different criteria. One of these is the demarcation based on who introduces them to the recipient language. Borrowings by native speakers are called *adoptions* and by non-native speakers *impositions* (Haspelmath 2009: 36). This study will not differentiate between these types and will use the verb *adopt* for all borrowings. Another type of division is between *material* and *structural borrowing* (ibid.: 39). *Loanwords* are a type of material borrowing where a word form from another language is adopted together with the meaning of the lexeme (Durkin 2014: 8). Examples of loanwords are *genius*, *ketchup*, and *dance*, which were borrowed from Latin, Chinese, and Old French respectively (*OED* s.v. *genius n.*, *ketchup n.*, and *dance n.*). The meaning of a loanword does not have to exactly correspond to that of the lexeme in the donor language and can simply be similar to it (Durkin 2014: 8). *Calques* are structural borrowings which transfer the meaning but not the form of a lexeme in the donor language, which acts as a model for the form in the recipient language (ibid.: 9). The components of the borrowed lexeme are translated and replaced with components with similar meanings from the recipient language, which is the reason *calques* are also called *loan translations* (ibid.: 9). *Semantic loans* (also called *loan meaning extensions*) are also structural borrowings that differ from calques in that, in semantic loans, the meaning of a lexeme in the donor language is applied to an already existing lexeme in the recipient language (ibid.: 9; Haspelmath 2009: 39). In other words, calques create new lexemes, while semantic loans add meanings to existing lexemes. For instance, *swan-song* is a calque from the German *schwanen(ge)sang* or *schwanenlied* (*OED* s.v. *swan n.*) and (*spiritual*) *father* in the sense '[a] male religious teacher, counsellor, or leader […]' is a semantic loan from post-classical Latin *pater spiritualis*. *Hybrid borrowings*, or simply *hybrids* or *loan blends*, are a type of material borrowing defined as lexemes which "involve the transfer of part of the foreign model and the reproduction of the rest" (Winford 2003: 43).

Borrowings are adapted to the phonology and morphology of the recipient language (Winford 2010: 173). Japanese borrowings in English show adaption not only in these areas but also in orthography because the writing systems of the languages differ, and the Japanese script is mostly unreadable to English-speakers who have not learned the language. Present-day Japanese is written using Chinese characters, *kanji*, two phonetic based kana syllabaries *hiragana* and *katakana*, and the Latin alphabet, or *rōmaji* (Igarashi 2007: 15; Iwasaki 2013: 20). All of these are used in combination, so that kanji are used to write "content words of native or Sino-Japanese origins", hiragana is for some content words of native origin as well as function words and inflectional endings, and katakana is similarly used for some content words, loanwords from non-Sino-Japanese languages, onomatopoeia and for emphatic and stylistic effects (Igarashi 2007: 15; Hadamitzky & Spahn 2011: 8). The use of *rōmaji* is restricted to signs and to conveying emphasis (Igarashi 2007: 15). Japanese may be transliterated in different romanization systems. Two of the most commonly used ones are the Kunrei-shiki, which is favored by the Japanese government and represents all kana syllables in a uniform fashion, and the Hepburn system, where "consonant sounds are spelled as in English, and the vowel sounds as in Italian" (Hadamitzky & Spahn 2011: 10). Japanese borrowings, of course, do not need to follow these systems specifically as they are for representing the Japanese language in Latin letters, and borrowings that were adopted before such standards were developed often show different renditions. Some of the orthographic changes that even recent borrowings may undergo are the addition of a hyphen, as in *shabu-shabu* and the omission of one *a* or *u* in loanwords that have the sequence *aa*, *ou* or *uu* (represented as *ā*, *ō* and *ū* in the Hepburn system) signifying a long vowel (Yamamoto 2020: 37–38). These differ from the aforementioned romanization systems, where hyphens are not used, and long vowels are indicated by a circumflex accent or a macron (Hadamitzky & Spahn 2011: 11).

Borrowings comprise a significant percentage of lexemes in English. There are estimates that approximately "65% or more of English vocabulary is of foreign (non-Germanic) origin" (Winford 2003: 59), although the results vary. In Haspelmath and Tadmor's (2009) comparative study of loanwords in 41 languages, the percentage of loanwords in English was 41%, which placed the language among high borrowers. Durkin (2014) states that 32% of the 92,500 main entries in the ranges of A to ALZ and M to R in the *OED* are loanwords. Below is the figure presented in Durkin's study, which shows the 25 most prolific donor languages (2014: 25). As can be seen, the portion of loanwords of Latin or French origin is by far the largest. Next come other Indo-European languages, Greek, German, Italian, Spanish, Dutch, and early Scandinavian. Japanese is the tenth most prolific language in general and, notably, the most prolific non-Indo-European language, being the source of approximately 160 headwords.

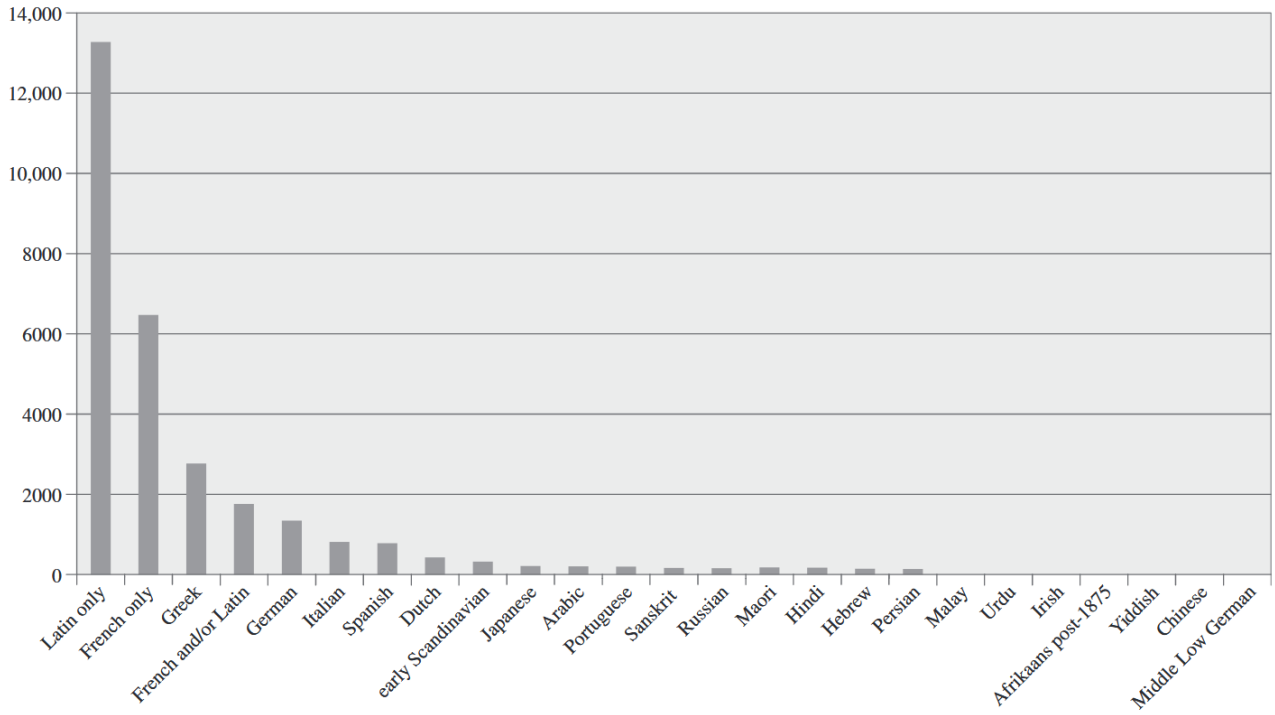


Figure 1. Totals of the loanwords from the 25 most prolific donor languages in *OED3* (Durkin 2014: 25).

The types of words usually borrowed are nouns and adjectives while verbs and especially closed-class function words are more resistant to borrowing (Winford 2003: 51). This is because

nouns and adjectives “form less tightly knit subsystems of the grammar than functional morphemes do” (ibid.). All languages have means to express both new and familiar concepts, so the question of why lexical items are borrowed from other languages in the first place may be asked. Borrowings may be divided into cultural borrowings and core borrowings, the former term referring to borrowings that designate a new concept and the latter to those that have the same meaning as an existing native word (Haspelmath & Tadmor 2009: 46). The adoption of cultural borrowings is frequently due to their convenience and efficiency in bilingual situations: it is easier to use a word that is already known to other speakers than to invent a new one (ibid.: 47). Core borrowings are, however, duplicates of existing lexemes by definition, so it may seem counterproductive if the same logic is followed. The adoption of core borrowings is connected to social and attitudinal factors, which affect borrowability in addition to the linguistic factors discussed before (ibid.: 35). If the donor language has prestige, speakers may be inclined to borrow words from the language to associate themselves with that prestige (ibid.: 48). Other reasons for borrowing include the original word becoming unavailable due to word taboo for example (ibid.).

2.1.2 Borrowings in the history of English

As observed in the introduction, borrowings have always been present in the English language. The Old English period spans from AD 450 to 1100, beginning from Germanic tribes settling in England and ending some years after the Norman Conquest of 1066 when English had gone through significant language change and lost most inflections (Millward & Hayes 2012: 17–18). There were three languages that influenced English the most in this period: Celtic, Latin, and Old Norse. Celtic had been spoken throughout Europe before the Germanic settlement in the British Isles and, as such, borrowings from the Celtic languages were adopted into Germanic on the continent (ibid.: 120). In

the Old English period, the borrowings from Celtic were mostly restricted to place names and name elements, such as *Thames* and *Cornwall*, and terms for landscape, such as *glen* 'valley' and *loch* 'lake' (ibid.: 121; Gelderen 2006: 92). The Irish missionaries who brought Christianity to England also imposed Celtic words into the language, for example *cross*, but the number of borrowings was overall scarce in comparison to the other two languages (ibid.). Latin was similar to Celtic in that its effect on English started from contact with Germanic tribes in continental Europe and is speculated to be the origin of several hundred words during the time (Gelderen 2006: 93). The effect of Latin on Old English began in 597 when Pope Gregory sent missionaries to England (ibid.). Latin influence is apparent in place names and "commercial, military, religious, and cultural terms" such as *street*, *priest*, and *school* (ibid.: 94). Latin also provided calques to Old English. For example, the Latin *unicornis* 'unicorn' was adopted as *anhorn* 'one horn' (ibid.). Over 400 words from Latin have been ascertained to have been in Old English before the Norman Conquest (Fennell 2001: 89). Old Norse was spoken by people from present-day Sweden, Norway, and Denmark who arrived in Britain in the 8th century (Gelderen 2006: 95). The people settled and "intermarried with the indigenous English" (Fennell 2006: 90), which led to a close contact between Old Norse and English. The number of borrowings from Old Norse is estimated to be approximately one thousand (Minkova 2004: 779), although there have been difficulties in assessing the origin of some because the languages have many identical lexemes (Fennell 2001: 91). The influence of Old Norse on the lexicon was deep, as the language affected not only place names and common everyday words, such as *egg*, *sister*, and *sky*, that did not fill a linguistic gap but also pronouns, prepositions, the infinitive marker, and verb *to be* (ibid.).

The Middle English period starts from 1100 and continues until 1500 when the Renaissance began and the printing press was implemented (Fennell 2001: 2). The aftermath of the Norman Conquest of 1066 was that French nobles were appointed into positions in military and church (ibid.: 106). As French became the official language and was held in prestige, it brought along borrowings that represented "almost every aspect of civilization" (Millward & Hayes 2012: 195). Millward and

Hayes list a dozen semantic fields, including food and eating, fashion, arts, music, literature, government, and the church, which received borrowings from French (ibid.). The language was also the source of more general lexemes, such as *age*, *cry*, *flower*, *pay*, and *please* (ibid.: 196). The impact of French on the Middle English vocabulary was large but did not extend to grammatical words, as Old Norse had (ibid.). Latin and Old Norse continued to contribute to English lexicon (ibid.: 192). Trade also brought other languages such as Dutch, German, Italian, and Spanish into contact with Middle English, although their influence was minimal in comparison to French (Gelderen 2006: 102).

The Early Modern English period from 1500 to 1800 saw an improvement in the status of the language, as it started to be accepted as a medium of serious writing after the inkhorn debates (Gelderen 2006: 175). English lacked words that became central in the Renaissance, so terms were taken directly from Classical languages or through Romance languages (ibid.: 176). Latin borrowings in this period were "scientific, technical, artistic, philosophical, educational, and literary terms" (Millward & Hayes 2012: 278). Affixes were also borrowed to form new Latinate words that had not existed in the original languages (ibid.). For example, *cortical* is formed from *cortex* 'bark' and the affixes *-ic* and *-al*, all of a Latin origin, within English (ibid.). Other donor languages included French and Celtic, as in the previous periods, Dutch, which was the source of lexemes related to seafaring and painting, Italian, from which words for trade and architecture as well as art were imported, Spanish and Portuguese, which brought terms for "the exotic products and life forms" from the Far East and the New World due to exploration and colonization (ibid.: 279–80). Non-Indo-European languages include Amerindian languages, which became sources for borrowings because of the English settling in North America (ibid.: 281). Asian languages such as Hindi, Malay, Chinese, and Japanese were the sources of *guru, rattan, ketchup,* and *soy* for example (ibid.). Borrowings were also taken from Near and Middle Eastern languages, from which *jackal*, *shawl*, and *sheik* were borrowed, and to a smaller extent from African languages (ibid. 281–2).

The Modern or Present-day English period starts from 1800, which is close to when the American colonies gained independence (Fennell 2001: 2). Especially developments in science and technology have brought an influx of borrowings (Millward & Hayes 2012: 323). Latinate words formed in English from Classical elements continue to be popular (ibid.: 325). A growing number of borrowings have been taken from languages other than Latin, and there are numerous languages that act as sources for borrowings because of a variety of factors such as "global communications and free exchange of products, information, and research", "diplomacy and wars" (ibid.: 325–6), immigration, and the invention of the World Wide Web. Thanks to the wide spread of English in this period, the selection of donor languages is on a "near-global scale" (Durkin 2014: 399): in addition to all of the languages mentioned before, for instance Russian, Urdu, Arabic, and Yiddish have become donor languages as well (Fennell 2001: 176). It has, however, been noted that borrowing has become less common as word-formation techniques such as blending and compounding have become more popular (Fennell 2001: 177).

2.2 Lexical semantics and the semantic fields of borrowings

Lexical semantics is the area of linguistics which studies word meanings or, more specifically, the meanings of lexemes, which are "groupings of one or more word forms, which are individuated by their roots and/or derivational affixes" (Cruse 2000: 88). The field differs from pragmatics in that the meanings studied are out of context, that is, not part of a specific utterance or text. Lexical semantics focuses on the meanings of content words rather than function words, which belong to the field of grammatical semantics (ibid.: 90). Meanings are divided into two broad types, conceptual meaning and associative meaning (Yule 2010: 113). Conceptual meaning (also known as the *denotative* or *cognitive meaning*) involves denotation, which is the relationship between the lexeme and "the class of objects, properties, etc., to which the expression correctly applies" (Lyons 2009: 207). It is the type

of meaning defined in dictionaries (Yule 2010: 113). Associative meaning encompasses the additional associations or connotations attached to the lexeme (ibid.). This study uses the term *sense* synonymously to *conceptual meaning*, as this is how the word is used in the *OED* as well.

Lexemes that are semantically related in a language system belong to the same semantic field (also called *lexical field* and *lexical configuration*) (Lyons 2009: 268; Lipka 1992: 152). Haspelmath and Tadmor (2009) give a short account of the most common semantic fields the loanwords of 41 languages belong to. The table below from their study shows that the semantic fields of Religion and belief, Clothing and grooming, and The house have the highest percentage of borrowings. When it comes to English specifically, Durkin states that the semantic categories that are particularly represented by lexemes of foreign origin are "food and drink", "flora and fauna", "imported goods" and "aspects of local material and social culture", noting that most of the loanwords refer to new objects or concepts encountered in contact with the donor language (2014: 398). The type of contact with the donor language affects what sorts of borrowings are adopted. For example, borrowings from German most often relate to the natural sciences while those from Hebrew are connected to religious and cultural terms (ibid.: 385). Durkin states that Japanese loanwords include "food and drink", "aspects of Japanese material culture", "terms relating to Japanese culture and society", "terms connected with warfare, martial arts, combat, or violence", "names of (ornamental) fish", and "an interjection bidding farewell" (ibid.: 397) but does not delve further into the number of borrowings in each category.

Semantic field	Loanwords as % of total
Religion and belief	41.2%
Clothing and grooming	38.6%
The house	37.2%
Law	34.3%
Social and political relations	31.0%
Agriculture and vegetation	30.0%
Food and drink	29.3%
Warfare and hunting	27.9%
Possession	27.1%
Animals	25.5%
Cognition	24.2%
Basic actions and technology	23.8%
Time	23.2%
Speech and language	22.3%
Quantity	20.5%
Emotions and values	19.9%
The physical world	19.8%
Motion	17.3%
Kinship	15.0%
The body	14.2%
Spatial relations	14.0%
Sense perception	11.0%
All words	**24.2%**

Table 2. Borrowing by semantic field (Haspelmath & Tadmor 2009: 64).

As noted in the introduction, the semantic fields of Japanese borrowings in English have been explored to a limited extent in loanword studies. Appendix 1 contains a list of the semantic categories used in three studies on Japanese borrowings. Tsuchihashi's article (1997) focuses on the history of Japanese loanwords from various dictionaries. She lists ten semantic categories to which the words belong. The categories are broader compared to those used in the two other studies and especially DAILY LIFE is used almost as a catch-all category that includes words such as *tatami, kimono, rickshaw,* and *shiatsu*. The category HISTORY is somewhat questionable, as it is less related to semantics but rather comments on whether the referents are extant at the point in time examined. Many of the words that were classified into this category have referents that were actually extant when the borrowing was adopted. Doi's doctoral thesis (2013) is an in-depth study into the Japanese

loanwords found in the *OED* and Kæmpfer's *the History of Japan*. He uses a division of 36 categories. Schultz's article (2017) is the only one among the three that is explicitly on the lexical semantics of Japanese loanwords. The loanwords under examination are those first attested after 1900 in the *OED*. The loanwords are classified as belonging to eight main categories, which are divided into subcategories and some of these further into even more precise subcategories. Although there are subcategories and the exact wordings vary, the categorization resembles Doi's.

In general, the research so far points to borrowings of Japanese origin coming from nearly all spheres of life. There is, however, variance in the number of borrowings in each category. Schultz's (2017) findings are that the domain of Martial arts has the highest number of borrowings, followed by People and Everyday Life, and The Natural Sciences. Gastronomy is not far behind, as the category has only one borrowing fewer than The Natural Sciences. The top three in Doi's (2013) study are Arts and crafts in first place, Sports and martial arts in second, and Food and drink in third. The categories relating to martial arts and food have relatively high numbers in both studies. The differences are partly due to the studies having different selections of borrowings and different category divisions. Doi divides each of the fields of science into separate categories, which is why there is no equivalent group to The Natural Sciences in the top three. Were the borrowings from the different fields combined, the resulting category would be the largest. Similarly, many of the borrowings which are under the People and Everyday Life category in Schultz's study are in separate categories such as Professions and status and Residents and emigrants in Doi's. Arts and crafts in Doi's study seems to be larger than the corresponding category in Schultz's simply because the borrowings analyzed were not identical, as Doi's Arts and crafts has more borrowings despite not including ones related to literature and music, which are in the category of The Fine Arts and Crafts in the other study.

2.3 Outline of contacts between speakers of English, Japanese, and languages of the Outer Circle regions

2.3.1 Outline of contacts between speakers of English and Japanese

The English first reached Japan in the 17th century. The only Europeans in the country before that were Portuguese and Spanish traders and Jesuit missionaries, the latter having gone there to preach Christianity in 1543 (Meyer 2012: 97). The first known Englishman in the country was Will Adams, who was the second-in-command of the first Dutch ship that arrived in Japan in 1600 (Cullen 2003: 33; Meyer 2012: 98). Just a little over a decade later in 1613, the English East India Company set foot in the country and opened a trading base in Hirado in the same year with the help of Adams (Henshall 2014: 137–8). The enterprise was, however, unprofitable and the English left the country in 1623 (ibid.: 138; Cullen 2003: 35).

What followed a decade afterwards were stricter foreign relation policies in Japan: the presence of Europeans became a concern to the shogunate, which saw the spread of Christianity as a threat (Henshall 2014: 80). This outlook was strengthened by the Shimabara Uprising of 1637–8, which led to the deaths of 35,000 people in Shimabara (ibid.: 381). The unrest was blamed on the Christian beliefs of the peasants in the area, as it was "the most Christianized area in Japan" at the time (ibid.). By 1639 all foreigners except the Chinese, Koreans, and Dutch were expelled from the country, and the latter were confined to the artificial island of Dejima in Nagasaki, the leaving of which required shogunal approval (Henshall 2014: 80; Meyer 2012: 106). The isolationist policies enacted in the 1630s also disallowed the Japanese from departing from or returning to the country (ibid.: 107), and the policies would remain until the Opening of Japan in 1854 (ibid.: 112–3; Russell 2014: 289).

The contact between Japanese and English speakers was minimal and restricted in the years of isolation. There were attempts by the Americans to open fur trade with the Japanese in 1791, but the negotiations failed (Meyer 2012: 122). In 1796 and 1797, a British warship visited the coasts (Cullen

2003: 140) and, in 1808, another one named HMS Phaeton entered the harbor in Nagasaki, the crew demanding supplies and leaving after the demands were met (ibid.: 149). In 1837, the British vessel Morrison attempted to open trade but was fired at (Cullen 2003: 158) in accordance with the edict of 1825 that ordered daimyos to drive away foreign ships on coasts by firing without hesitation (Russell 2014: 287). The edict was replaced in 1842 with one that ordered provision of supplies in emergency situations (McNally 2015: 72). Two U.S. Navy ships tried to gain entry in 1846, but they were refused, and, in 1848, another ship was able to retrieve fifteen American castaway sailors from the country (Perez 1998: 83; Meyer 2012: 122). A rare case of a Japanese speaker entering an English-speaking country is that of Nakahama Manjirō, a fisherman from Tosa, who became shipwrecked in 1841 (Henshall 2014: 282). He and his shipmates were rescued by an American whaler and, while the others were dropped off at Hawaii, Manjirō continued aboard the ship to the United States and lived there for ten years before returning to Japan (ibid.).

What put an end to the isolation was the arrival of the Americans in the form of Commodore Matthew Perry and four black ships in the Edo Bay in 1853 (Cullen 2003: 178; Meyer 2012: 122). He had been tasked to deliver a letter from the American president, who requested “peace and friendship, free trade, good treatment of shipwrecked whalers, and provisioning of coal for vessels” (ibid.: 123). Perry informed the Japanese that he would return the following year for a reply (ibid.). On his second visit, he brought nine vessels, the guns of which had more firing power than the batteries around Edo bay (Cullen 2003: 178; Goto-Jones 2009: 16). Under the pressure of the domestic issues the country was facing at the time and the fear that the Americans would resort to using force, Japan signed a treaty with the United States on March 31, 1854 (Meyer 2012: 123–4). The treaty entailed the opening for supplies of Hakodate, south of Edo, and Shimoda, the latter of which was to have a consular residence instituted, “good treatment of shipwrecked sailors, limited trade”, and the most-favored-nation clause (Meyer 2012: 124). The events led to other countries, including Britain, signing similar treaties with Japan shortly afterwards (ibid.). Townsend Harris, who

was appointed the American consul general, concluded further treaties later in the same decade (ibid.: 124–5). These included for example "diplomatic representation in the capitals of both powers", the right of residence at Shimoda, Hakodate, Osaka, and Edo as well as trade in Nagasaki, Kanagawa, Niigata, and Hyogo (ibid.). Extraterritorial rights were also granted for Americans in Japan in criminal and civil cases (ibid.). The treaties were once again imitated by the British among others (ibid.). All of the treaties were unilateral in that they applied to foreigners in Japan and not to the Japanese abroad, which is why they have been called the *unequal treaties* (Perez 1998: 84).

As the period of isolation had come to a close, the contact between Japanese- and English-speakers increased. Emigration from Japan to America began in 1868 and, in 1910, there were 70,000 Japanese by birth residing in the country, many having gone to work in the sugar plantations (Cullen 2003: 234). The unequal treaties were ceded starting from 1894 in return of various rights to foreigners, including the right to reside anywhere in Japan (ibid.: 207). The country was experiencing enlightenment caused by the rapid modernization and industrialization in the Meiji period of 1868–1912, which brought a need for foreign experts who would develop the educational system (Meyer 2012: 151). Protestant missionaries also arrived from the United States and Britain and established schools (ibid.: 153). The nature of the contact between Japan and the English-speaking regions changed through the course of the wars and interwar periods when Japan joined in the First World War as an ally to Britain and as an opponent of both Britain and the United States in the second one (ibid.: 171). The military advancement of Japan involved both of the countries in the form of the attack on Pearl Harbor and the Japanese occupation of areas of the South East Asia, including Hong Kong and the British Malaya in 1941, Singapore in 1942 as well as the Philippines, which had been under American rule, among others (Church 2017: 91, 125). The bombing of Japan brought an end to the war and what followed was the post-war period of economic growth.

The beginning of the post-war period was marked by the occupation of Japan by the American forces from 1945 to 1952 (Meyer 2012: 203). Soon afterwards the economic boom enabled by the

growing steel industry in the 1950s and 1960s (Goto-Jones 2009: 101) made Japan into a world power internationally known for its consumer products as well as popular culture such as games, comics, and cartoons. Tourism and the general migration of people have since spread Japanese concepts world wide.

2.3.2 English in Singapore, the Philippines, Malaysia, and Hong Kong

The South-east Asian regions of Singapore, the Philippines, Malaysia, and Hong Kong all use English in business and government and, in most of them, it is also a second language. This is the reason why these regions are classified as belonging to the Outer Circle as opposed to the Inner Circle, where English is the first language of most residents (Melchers et al. 2019: 7–8). The language was most often introduced as a result of British colonialism. What the regions have in common is that all of them are relatively close to Japan geographically and were briefly under Japanese rule during the Second World War.

The East India Company acquired Penang in 1786, Singapore in 1819 as well as Malacca in 1824 and, in 1826, the three regions were combined into the Straits Settlements (Church 2017: 87). Especially Singapore functioned as an important site for commerce because of its geographical position (ibid.: 135–6). The Federation of Malaya, encompassing the three regions among others, was formed in 1948, but in 1965 Singapore and Brunei separated into independent states (ibid.: 562). Despite the shared history, the status of English is different in Singapore and Malaysia: in the former it is an official language in addition to Malay, Chinese, and Tamil while in the latter it does not have a similarly high status (Hickey 2004: 563, 566). English is the main medium of education in Singapore and is increasingly used in law and administration (ibid.). Based on the census of 2010, it is also spoken in the homes of 17.0% of the Malay, 32.6% of the Chinese and 41.6% of the Indian

population (Singapore Department of Statistics 2011: 25–6). The popularity of the language is largely due to it being a means of interethnic communication (Hickey 2004: 566). In Malaysia, English spread as a means of education but is now spoken less as it lost its status as an official language in 1967 (ibid.: 563; Yamaguchi & Deterding 2016: 7). It is not used as the medium of education anymore, although it is taught as a compulsory subject (ibid.). The two Englishes spoken in the countries share many features such as the particle *lah*, which has various pragmatic functions, such as signifying informality or intimacy. However, SgE is more influenced by Chinese and MyE by Malay (ibid.: 564–5).

The Philippines is an exception among the four regions in that its contact with English was under American rule rather than British. Prior to the Treaty of Paris in 1898, where the ruling of the country was transferred to the United States, the country had been ruled by Spain since 1565 (Hickey 2004: 575). The intention of the Americans was to have the country become independent, and the Republic of Philippines was formed in 1946 (ibid.; Church 2017: 124). English and Filipino are currently the official languages (Hickey 2004: 576). PhE is based on AmE unlike the Englishes of the other regions studied, so it is rhotic, although an alveolar flap is used instead of the retroflex /r/ (ibid.: 577). English-language education was established at the beginning of the 20th century and, similarly to the situation in Singapore, the language has the role of a lingua franca in the country, where approximately 110 indigenous Austronesian languages are spoken (ibid.: 575–6).

English arrived in the Hong Kong Island in 1821 with British merchants, who used the island in opium trade (Hickey 2004: 571). The region was placed under British rule in the Treaty of Nanking in 1848 and remained so until 1997 except for the brief period of Japanese occupation (ibid.). English is the official language in present-day Hong Kong together with Chinese (Hickey 2004: 572). It is used in the “commercial, legal and administrative spheres and in the media”. According to the census of 2018, Cantonese was the primary language of 88.8% of the residents aged 6–65, but 66.9% also estimated that their competence in spoken English was from average to very good (Census and

Statistics Department 2019: 82), so code-switching and code-mixing between Cantonese and English also takes place commonly (Hickey 2004: 572).

3 Data and methods

This section presents the data and methods used in this study. Section 3.1 explains the concept of lexeme-sense pair while Section 3.2 and 3.3 concentrate on the first and second set of data and the methods used for each respectively.

3.1 Lexeme-sense pair

The previous studies on the semantic fields of Japanese borrowings mentioned in Section 2.2 are not corpus-based, which is why the studies have treated borrowings as lexemes, which may have polysemous meanings. This study will instead employ a concept here termed *lexeme-sense pair* (henceforth abbreviated to *LS pair*). An LS pair is formed from the lexeme part and the sense part, which is one distinct meaning of the lexeme. For the purposes of this study, the sense part will correspond to one sense listed in the *OED* under a headword. The reasons for using LS pairs, instead of simply borrowings with polysemous meanings, are that (1) senses develop and are attested at different times and (2) some senses of the same lexeme may be semantically very far from one another.

The first point is related to the first research question about borrowings adopted in different time periods. For example, *ronin* has two meanings listed in the *OED*: '[…] a samurai without a lord or master […]' and 'a Japanese student who has failed, and is studying to retake, a university entrance examination' (s.v. *ronin n.*). The former sense is first attested in 1858 while the latter approximately a hundred years later. If the study followed what other researchers have done and considered only the former as the first attestation, a problem would arise in the number of borrowings in semantic categories per 50-year periods. If the two meanings were combined under one borrowing, and the

borrowing classified into the semantic categories of related to military (based on the first sense) and education (based on the second sense), this would cause a situation where one borrowing would be added to each of these categories in the period of 1850–1899, despite the sense that connects the borrowing into the field of education being attested at a later time period. This would skew the results and give an inaccurate account on the semantic fields represented in each period.

The second point relating to semantic differences between multiple senses of the same lexeme is especially important when examining borrowings in a corpus text. For example, *kamikaze* has the senses '[t]he pilot of a kamikaze aircraft […]' but also many other ones, including '[a] cocktail made with vodka, orange-flavoured liqueur, and lime juice' (*OED* s.v. *kamikaze n.*). It would be strange to claim that the token of *kamikaze* in *I went to a bar and ordered a kamikaze* has a meaning relating to the military, but if there were no division between the senses, such cases would increase the frequency of the field of military.

Thus, it was assessed that the concept of an LS pair would be suitable when researching borrowings with the help of a corpus. This study will use the term *borrowing* to refer to LS pairs, where either the lexeme part or the sense part or both have Japanese origins. For example, the noun *Ainu* is recorded twice in the data, once as '[a] member of (the people descended from) an aboriginal people of northern Japan […]' and once as '[t]he language traditionally spoken by this people' (*OED* s.v. *Ainu n. and adj.*), and these two are considered two LS pairs. A more detailed definition on what were considered Japanese borrowings in this study is presented in Section 3.2.2.

3.2 First set of data and the methods used

This section presents the first set of data, which is the basis for answering the first research question about the distribution of borrowings into semantic fields during different time periods. The data set was collected from the *Oxford English Dictionary Online* introduced in Section 3.2.1, followed by Section 3.2.2, which is about the data set itself, and Section 33.13.2.3, which explains the methods used for the analysis of the data set.

3.2.1 The *Oxford English Dictionary*

The *Oxford English Dictionary* is a historical dictionary of English, which has its origins in a proposition in 1857 and completion of the first edition in 1928. Besides the printed volumes, the dictionary has been made available electronically from 2000 as the *OED Online*, which this study utilizes for the first set of data, and the compilers are in the process of updating all entries for the third edition. The *OED* includes words from a wide range of regional varieties and registers spanning from slang to formal.

The entries in the *OED Online* consist of the headword, pronunciation, spelling variants, frequency, etymology, senses, quotations, and phrases, compounds, and derivatives built round the headword. The headword includes the lemma, part of speech and a numbering if there are homographs. Pronunciations are transcribed for British and American English and can be listened to by playing sound files. Spelling variants show estimates of the time period a spelling was in use. Frequencies are expressed through Frequency Bands which indicate frequency per million words in current use: Band 1 is for “extremely rare words unlikely ever to appear in modern text”, Band 2 for less than 0.0099, Band 3 for those between 0.01 and 0.099, Band 4 for those between 0.1 and 0.99, Band 5 for those

between 1 and 9.9, Band 6 for those between 10 and 99, Band 7 for those between 100 and 999, and Band 8 for more than 1000 occurrences per million words. Etymological information includes origin, that is the language from which the headword was borrowed or inherited, the direct etymon and full etymology of the headword. The entry is divided based on the different senses of the headword. Each of the senses is followed by citations containing the headword, ordered by the date and title of the publication. The first of the citations is the first attested use in writing. The final sections are reserved for phrases, compounds and derivatives containing the headword.

The *OED Online* was chosen as the primary source because the dictionary contains a large number of Japanese borrowings of varying topics thanks to the policy followed since the inception of the dictionary of including all English words, even borrowings from other languages, and their etymologies (Ogilvie 2012: 27–9). Following the policy, the *OED* provides detailed etymologies for headwords, which serve a central role in the first set of data. The dictionary also contains other information necessary for the study, specifically spelling variants, senses, and first attested uses. Another major factor was the advanced search function allowing for restriction by etymology, which was considered crucial for acquiring data. All in all, the dictionary is well suited for loanword studies, containing essential information for research.

3.2.2 First set of data

The first set of data was acquired from the *OED Online*. It was necessary to define what are considered Japanese borrowings before conducting searches in the *OED*. In addition to lexemes that originate in Japanese and have been directly borrowed from that language to English, this study also includes lexemes which originate in another language and have Japanese at some point of their etymology before the transfer to English. In practice, this means the inclusion of such lexemes as

yakitori, which was formed in Japanese and directly transferred to English (*OED* s.v. *yakitori n.*), *anime*, which is originally from the English noun *animation* and entered English from Japanese (*OED* s.v. *anime n.*), and *mebos*, which was transferred from Japanese to Dutch to South African Dutch and finally English (*OED* s.v. *mebos n.*). English derivatives of Japanese borrowings are excluded, so if a lexeme includes a Japanese base that was already attested in English before, it is here considered an English formation. However, lexemes, such as *akebia*[1], that combine a Japanese base and an affix from another language were included. Another type of lexeme that was included was those which are formed from an English affix, root, or base and a Japanese root or base that has not been attested in English separately before, for example *andosol*[2]. Lexemes derived from Japanese proper names are also considered to be borrowings in this study. The etymology of borrowings will be further discussed later in this section.

The first set of data was obtained through the advanced search function of the *OED Online*, where the string *Jap** was searched in the *Etymology* section of all entries. This yielded a list of 736 entries. Although the *OED Online* provides the possibility of searching the language of origin of headwords, this was not used because it would only retrieve words that have been directly borrowed from Japanese and would ignore multiple types of words that are considered relevant to this study. These are lexemes derived from proper names, which are not listed as originating in any language in the *Origin* section, and semantic loans, as, by definition, the lexemes have existed in English before gaining a new meaning corresponding to the meaning of a Japanese lexeme. The reason for choosing *Jap** as a search term, instead of simply *Japanese*, is that earlier research such as Doi's (2013) has commented that the CD-ROM version of *OED2* marks some etymologies with *Japanese* or *Jap.*, so the practice in this study is a measure against the exclusion of the latter type. A closer examination of each of the 736 entries that were retrieved, however, showed that the *OED Online* marks all of the

[1] The etymology of *akebia* is "< scientific Latin *Akebia*, genus name […] < Japanese *akebi*, in the same sense […] + scientific Latin -a […]" (*OED* s.v. *akebia* n.).
[2] The etymology of *andosol* is "< Japanese *ando*, scientific term for a specific type of volcanic soil (not dated in dictionaries of Japanese; < *an-* dark + *-do* soil (both < Middle Chinese)) + *-sol* comb. form" (*OED* s.v. *andosol* n.).

borrowings with the full word *Japanese* even in the entries that were already present in the second edition. Because the search term was *Jap**, the entry list predictably consists of all lexemes that have a word with the same sequence of letters in the etymology section. Thus, the list also had lexemes that were not relevant from the point of view of this study. Nevertheless, all the headwords from the entry list were initially recorded in the data regardless of relevancy. Each entry was examined and, if the etymology established a connection to the Japanese language, certain items of information from the entry were recorded: senses of the headword, the year of the first attested use of a specific sense, time period of the first attested use, etymology, obsoleteness, spelling variants, *OED* edition, and a possible reason for discarding.

The senses of the headwords are important as they are the basis for the semantic analysis of Japanese borrowings. As detailed in Section 3.1, all senses of the headword were documented separately as LS pairs. The senses were taken from the regular Sense section of an entry as well as from Draft additions and sometimes from the Derivatives section and Compounds section. The Derivatives section most often lists lexemes whose root or base is the headword to which an English affix has been added. These are not of interest to this study. Occasionally, however, compounds that were formed within Japanese are found in the Derivatives section. The latter type, to which for example *sobaya* and *sushiya*[3] belong, was included in the data as these are borrowings as far as English is concerned. The LS pairs taken from the Compounds section were mostly calques, semantic loans, or hybrids; only two of them, *mikoshi* and *makuuchi*, are loanwords that were listed under Compounds. Loanwords are generally analyzed in the recipient language as simplices even if the lexeme in the donor language has a complex form (Haspelmath 2009: 37), which is the reason why there are so few loanwords in the Compounds sections.

The years of the first attested use were recorded per LS pair so that they reflected the year a particular sense was first attested. Some semantic loans do not have their own Sense sections and are

[3] The borrowing *sobaya* '[…] a shop or restaurant which serves soba' (*OED* s.v. *soba n.*) is from the Japanese compound *soba* + *ya* 'house' and *sushiya* '[…] a shop which serves sushi' from *sushi* + *ya* (*OED* s.v. *sushi n.*).

included in the same section as an earlier sense of the headword, which means that the first attested use listed is of the non-Japanese meaning. For example, one of the senses of *busgirl* has the specifying note "[o]riginally with reference to women employed as bus conductors in London during the First World War (1914–18). Now usually in Japanese or Korean contexts" (*OED* s.v. *busgirl n.*). In this situation, the first attested use was taken from the first quotation in a Japanese context, so that in this case the year would be 1961 because the source *Japan & Zen* was inferred to indicate a Japanese context while the earlier years were not. If the year of the first attested use was in brackets, which signifies a non-domestic usage, or if the word was explicitly said to be part of the Japanese lexicon in the quotation, it was ignored, and the year listed next was taken. An example of the latter type is found in the entry for *saké*, where the first attested use of the word occurs in the sentence *Their ordinary drink is a kind of Beer (which they call Saque) made of Rice* (*OED* s.v. *sake n.2*). There are also senses whose years of the first attested use are not certain and are marked with *a* (meaning 'around') before the year, as with the first example of *Shinto* '[a]n adherent of Shinto beliefs' from *a1832* (*OED* s.v. *Shinto n.*), or with a span of years, which is exemplified by *orihon,* the first attestation of which is dated 1882–4 (*OED* s.v. *orihon n.*). In the former case, the year after the marker was considered the year of the first attested use and, in the latter, the earliest year was taken. With this data, it does not matter which of the years from the span of years was chosen, as it does not affect the time periods of the first attested use, that is to say, the initial and terminal years of all recorded senses happened to belong to the same 50-year time period. There was one LS pair that lacked quotations entirely, *dairi-sama* (*OED* s.v. *dairi n.*), which I marked *N/A* (meaning *not available*) in the year of the first attested use. The time periods of the first attested use are 50-year periods starting from 1550, which includes the year a Japanese borrowing was first attested in English. In the case of *dairi-sama*, the time period was also marked as *N/A*.

Etymology was an important factor in separating relevant cases from irrelevant ones as it tells whether the lexeme fulfils the definition of borrowing followed here. It also assists in determining

the type of borrowing, which will be discussed in the latter part of this section. Etymologies were previously discussed at an earlier point, where it is mentioned that the relevant cases may have Japanese at any point of their etymologies: there may be other languages that have preceded Japanese or followed it before the lexeme was borrowed into English. The reason for including lexemes which originate in Japanese but have reached English through another language is that there are cases where the *OED Online* lists Japanese as being the direct donor language although this is strictly speaking not the case. For example, Engelbert Kæmpfer's *The history of Japan* was translated from the German manuscript into English by Johann Gaspar Scheuchzer before its post-humous publishing (Kæmpfer 1999: 7–8). However, the lexemes recorded from the work are not marked *German < Japanese* in the etymology section. Similar cases exist with lexemes that have first been attested earlier than 1853–4, when the isolationist policies of Japan began to be lifted. The opposite, that is lexemes that have been directly transferred from Japanese to English but are not originally Japanese words, are also considered Japanese borrowings in this study. Often such lexemes contain elements from Middle Chinese, which was the donor language for many Japanese words. Because of the extensive borrowing of Middle Chinese compounds with the same elements, the elements have become transparent and analyzable to Japanese speakers and can be used to form new compounds within Japanese, which results in a situation similar to neo-classical compounds in English (Haspelmath 2009: 37–8). Some complexities also arise in the use of the second edition of the *OED*, which does not specify the complete etymology in some entries and, instead, only lists *Japanese*, making it difficult to interpret whether this means that the lexemes originate in Japanese, were directly borrowed from the language, or both. Further examination of the lexemes from the second edition reveals that they are likely directly borrowed from Japanese and include both words originating in Japanese and those having their origins in another language. This problem of incomplete etymologies has fortunately been solved in the third edition, which lists the full etymology of the headword. There are other studies, such as Stubb's research (1998) into German loanwords, which similarly uses a

broad definition of borrowings in terms of etymology, so this classification is not unheard of. Thus, the etymology of the included lexemes was any that had a Japanese etymon at some point. In addition to lexemes that have confirmed etymologies, those lexemes whose origins are unknown or uncertain but are theorized to have a connection to Japanese were also included in the data at this stage.

Notes on the obsoleteness of senses were also recorded. I marked the senses in the data with either *in use* or *obs*. (meaning 'obsolete') depending on whether the headword is still used in that particular sense in Present-day English. Obsolete lexemes are relevant to this study and are included in the data. Furthermore, notes on obsoleteness may be useful with the second set of data as they may account for certain senses or headwords being absent from the corpus. Spelling variants, which are included in the first set of data, will similarly be used in the acquisition of the second set of data: the variants will aid in the creation of search terms entered into the corpus program. The *OED* edition where the information was acquired from was documented, as *OED* is undergoing revisions and some entries relevant to the study may be updated at the time of the writing of this thesis. The reasons for discarding a headword or a sense were also listed, although, at this stage, only words with no etymological connection to Japanese were marked as excluded.

After all 736 entries were examined and the information discussed above was documented, additional criteria for exclusion were applied and the data re-examined with the help of the notes. The reasons for exclusion were:

1. no etymological connection to Japanese,
2. formation within English,
3. uncertain or unknown etymology,
4. proprietary name, and
5. first attested use after 1999.

During the first examination of all 736 entries, the headwords which had the sequence *Jap** in the etymology section but no etymological connection to the Japanese language were already marked

for exclusion. Many of these headwords had the country name *Japan* as their root, such as *Japanesque*, or mention *Japanese* in the etymology section, for example *Goanese*[4]. Additionally, there were cases which had nothing to do with Japan and had another word that coincided with the search term such as derivatives of *jape* or *Japheth*.

The entries that were discarded next were those that had been formed within English, although they have an etymological connection to Japanese. As noted earlier, lexemes which have a previously attested Japanese borrowing as the base and an English affix are not considered borrowings in this study. The same applies to other lexemes formed through different types of word formation techniques in English, which in practice were clipping and conversion. Example of headwords for these three types of lexemes formed within English include *Mikadoism* and *Nipponian* for suffixation, *nunchuck* and *rickshaw* for clipping, and the noun *reap* for conversion.

Some of the entries are listed as having an uncertain or unknown etymology which was speculated as possibly being of Japanese origin. Two such examples are *shaka* (*OED* s.v. *shaka int. and n.*) and *Roshambo* (*OED* s.v. *Roshambo n. and (int.)*). Due to the uncertainty, it was decided that such cases were not included in the data as there is a possibility that the etymologies of the lexemes do not involve Japanese and would thus not represent Japanese borrowings in English.

There were a few proprietary names amongst the data, for example *Betamax* and *Pac-Man*. These were also omitted, with the exception of lexemes that were not originally proprietary names but have later gained the status. In practice, there was only one such case, *Taka-diastase* '[a] preparation containing a variety of enzymes which is obtained after the treatment of rice or bran with the mould Aspergillus oryzæ; now a proprietary name' (*OED* s.v. *Taka-diastase n.*).

Finally, senses that have first been attested after the year 1999 were not included. This is because the last time period of the first attested use is 1950–1999 and the next would be 2000–2049, which will presumably gain more Japanese borrowings than the ones currently listed in the *OED*

[4] *Japanese* is used as an example in the etymology of *Goanese* "< the name of *Goa* […] + -ese suffix, with insertion of -n- for euphony, after e.g. *Japanese*" (*OED* s.v. *Goanese n. and adj.*)

Online. Fortunately, the decision to discard these senses does not skew the results much as there were only two such cases, *Sudoku* and the latter sense of *bukkake*.

After all items that fulfilled any of the five criteria were removed from the data, the number of relevant entries was 623 with 705 relevant senses. There were 113 entries from which all senses were discarded and one entry from which one sense (but not the other) was discarded. Of these, 80 were removed because of the lack of connection to Japanese, 17 because they were formed in English, five because of an uncertain or unknown etymology, ten for being proprietary names, and two for being first attested after the year 1999. The discarded entries are listed in Appendix 2.

The final phase in the arrangement of data was to assess the types of borrowings that the remaining data represented. This was done by inspecting the etymologies of all borrowings and assigning each LS pair a type of borrowing. Different senses of the same lexeme have the same etymology and, because of this, represent the same type of borrowing. Four types of borrowing were identified: loanwords, calques, semantic loans, and hybrids. The exact types of hybrids found in the data are presented below accompanied by examples:

- translated element (to Latin) combined with a Latinate root (*acupuncture*[5]),
- clippings of a loanword combined with an English suffix (*kainic* and *ibotenic*[6]),
- loanwords combined with an English suffix or root (*Ramanas rose* and *andosol*),
- loanwords adopted to French and combined with a French suffix, the suffix being adapted to English (*shikimic*[7]),

[5] *acupuncture* is formed from the "classical Latin *acū*, ablative of *acus* needle […] + post-classical Latin *puncture* […], with the first element being after Japanese *hari* needle" (*OED* s.v. *acupuncture n.*).
[6] The etymology of *kainic* is "< Japanese *kain(in* (< *kainin-sō*, the name of the alga Digenea simplex) + -IC suffix" and of *ibotenic* "< Japanese *iboten(gutake*, the name of the mushroom (*Amanita strobiliformis*) from which the acid was first isolated + -IC suffix" (*OED* s.v. *kainic adj.* and *ibotenic adj.*). The Japanese lexemes were truncated before adding the suffix *-ic*.
[7] The borrowing *shikimic* was adopted from the French *shikimique*, which originates in the Japanese *shikimi* (*OED* s.v. *shikimic adj.*).

- loanwords adopted to German and combined with a German suffix, the suffix being adapted to English (*sudoite*[8]),
- loanwords combined with a Latin suffix (*Nandina* and *akebia*), and
- loanwords combined with an English suffix or root as a partial calque (*miso soup* and *ishikawaite*[9]).

The number of senses which had loanwords as headwords was 642, which makes loanwords the most common type of borrowing. The number of lexemes in the other types of borrowing was considerably lower, which can be seen from Table 3. The exact breakdown of hybrids is presented in Table 4.

	Loanwords	Calques	Semantic loans	Hybrids
Number of LS pairs	642	31	11	21

Table 3. The distribution of types of borrowing per LS pair.

	Translated element to Latin + Latinate element	Clipped loanword + English element	Loanword + French element adapted to English	Loanword + German element adapted to English	Loanword + Latin element	Loanword + English element	Loanword + English element as a partial calque
Number of LS pairs	1	2	1	1	2	6	8

Table 4. The distribution of types of hybrids per LS pair.

[8] The English borrowing *sudoit* is from the German *sudoit*, which is formed from *sudo* (from the name of Toshio Sudo, Japanese mineralogist and crystallographer) + *-it*-suffix (*OED* s.v. *sudoite n.*). The suffix has been adapted to the English suffix *-ite*.

[9] The etymology of *ishikawaite* is "< Japanese *ishikawaishi* […], < *Ishikawa* , the name of a district in Honshu, Japan + *ishi* stone, mineral" (*OED* s.v. *ishikawaite n.*). The first part, *ishikawa*, remains the same as in the Japanese etymon while the second part, *ishi*, was translated as the suffix *-ite,* so the lexeme is a partial calque.

3.2.3 Methods used in the analysis of the first set of data

The methods used in the analysis of the first set of data are both qualitative and quantitative, first involving the establishment of categories that correspond to semantic fields and evaluating the meanings of lexemes, which requires qualitative assessment, and then the counting of frequencies of LS pairs in the categories. Furthermore, the frequencies serve as a basis for further qualitative analysis of the semantic fields of Japanese borrowings in time.

After all the data had been acquired and arranged according to the criteria, categories representing the semantic fields that the Japanese borrowings belong to were constructed by inspecting the senses of all LS pairs. There are two types of categories, main and subcategories. The main categories relate to wide semantic fields, while the subcategories depict more specific fields within a main category. Previous loanword studies such as Doi's (2013), Schultz's (2012, 2017), and Tsuchihashi's (1997), were consulted as an aid to the classification. The categorization is, however, my own and may differ from others due to differences in judgement. There are 25 main categories, of which 13 had at least two subcategories and 12 had none. The number of subcategories is 51 in total. It should be noted that the analysis of subcategories in Section 4 will also include the 12 main categories without subcategories. All the categories are presented below, the subcategories being indented under the main categories they belong to.

1 Architecture and interior design
- 1.1 Architecture
- 1.2 Buildings and property
- 1.3 Interior design and furniture

2 Arts and crafts
- 2.1 Arrangement[10]
- 2.2 Literature
- 2.3 Music
- 2.4 Painting, pictures, and printing
- 2.5 Paper folding and packaging

[10] *Arrangement* refers here to the arrangement of objects such as flowers and stones.

- 2.6 Theater, puppetry, and dance
- 2.7 Gemstones
- 2.8 Pottery and lacquerware
- 2.9 Other crafts
- 3 Attributes and qualities
- 4 Business and finance
  - 4.1 Business and working culture
  - 4.2 Finance
- 5 Cooking and beverages
  - 5.1 Beverages
  - 5.2 Food and cooking
- 6 Currency and units of measurement
  - 6.1 Currency
  - 6.2 Units of measurement
- 7 Education
- 8 Electronics and technology
- 9 Entertainment and leisure activities
  - 9.1 Comics and cartoons
  - 9.2 Leisure activities
  - 9.3 Games
  - 9.4 Travelling and tourism
- 10 Fashion
  - 10.1 Accessories
  - 10.2 Clothes
  - 10.3 Hairstyles
  - 10.4 Footwear
- 11 Government, administration, and politics
- 12 Greetings, interjections, and gestures
- 13 Historical periods
- 14 Manufacturing and materials
- 15 Military and bushido[11]
- 16 Nature
  - 16.1 Environment
  - 16.2 Fauna
  - 16.3 Flora
  - 16.4 Fungi
- 17 Science and medicine
  - 17.1 Alternative medicine
  - 17.2 Biology
  - 17.3 Chemistry and biochemistry
  - 17.4 Geology and mineralogy
  - 17.5 Medicine
  - 17.6 Physics
  - 17.7 Psychology
  - 17.8 Parapsychology
- 18 Seafaring, vehicles, and transportation

[11] The *OED* defines *bushido* as "[i]n feudal Japan, the ethical code of the Samurai or military knighthood" (*OED* s.v. *bushido n.*).

19 Society
- 19.1 Groups and organizations
- 19.2 Professionals and hobbyists
- 19.3 Ethnicity and residency
- 19.4 Status and rank
- 19.5 Titles, forms of address, and names
- 19.6 Other types of people

20 Sports and martial arts
- 20.1 Martial arts
- 20.2 Sports

21 Tradition, mythology, and religion
- 21.1 Tradition
- 21.2 Mythology
- 21.3 Religion

22 Values and concepts

23 Weaponry

24 Writing, language, and bookbinding
- 24.1 Bookbinding
- 24.2 Writing and language

25 Other objects

Once the categories had been established, each sense was examined to determine the categories under which the LS pair belongs. Many LS pairs represent multiple semantic fields, which is why the pairs were marked under all of them. There are two cases in which LS pairs were considered to belong to multiple categories: (1) a single sense in the *OED Online* includes multiple meanings which are different enough to be under different semantic fields, or (2) a single referent is related to multiple distinctive fields. An example of the first type that is related to the way the dictionary has been compiled is *Wagyu* '[a] breed of cattle of Japanese origin, from which is obtained tender marbled beef typically containing a high percentage of unsaturated fat; an animal of this type. Also: the beef obtained from such cattle […]' (*OED* s.v. *Wagyu n.*), which places the LS pair under FAUNA and FOOD AND COOKING. The second type that arises from the referent is exemplified by *karateka* 'an exponent or devotee of karate, a karate expert' (*OED* s.v. *karate n.*), which is placed under the subcategories of PROFESSIONALS AND HOBBYISTS and MARTIAL ARTS. Other research, such as Doi's (2013) and Schultz's (2017), has used similar practices, although by comparison this study has been more liberal in placing the LS pairs under multiple categories.

Information on the existence of the referents was also gathered concurrently with the semantic examination. The *OED Online* marks lexemes whose referents are no longer extant as *Hist*., meaning 'historical'. There are also cases where this marking is not present, but the definition of the lexeme suggests that the referent is historical. For example, *ronin* '[i]n feudal Japan: a samurai without a lord or master, esp. one who lives as a mercenary warrior; a wandering samurai, an outlaw' (*OED* s.v. *ronin n.*) is not labelled with *Hist.*, but the era during which the referent existed is mentioned. Again, sources other than the *OED Online* were used to assist in determining the existence status of the referents.

The data was rechecked once after all of the senses had been examined and the frequencies of the LS pairs in each main and subcategory and types of borrowing were counted per time period. Especially the chronological distribution of the borrowings into main and subcategories will be crucial in answering the first research question (*What is the distribution of Japanese borrowings in English in different semantic fields during different time periods?*). The number of LS pairs in the subcategories were counted and added together to produce the number of LS pairs in the main category. An exception to this were pairs that belonged under multiple subcategories of the same main category, which were counted only once in the main category so as not to skew the results. For example, *kami* '[…] (a title for) a governor, daimyo, or other person of authority or high rank […]' (*OED* s.v. *kami n.2*) belongs to STATUS AND RANK and TITLES, FORMS OF ADDRESS, AND NAMES, both of which are subcategories of the main category SOCIETY. Thus, the LS pair is counted once in the main category of SOCIETY. The chronological distribution of the types of borrowing and the number of LS pairs in the main categories per type of borrowing were calculated to examine how the borrowings were adopted during different times and whether there was a tendency to prefer a certain type of borrowing with different semantic fields.

3.3 Second set of data and the methods used

This section concerns the second set of data acquired from the Corpus of Global Web-based English to answer the second research question of what the distribution of semantic fields of Japanese borrowings is in texts of Present-day English varieties. Section 3.3.1 contains a brief introduction to the corpus and discusses the reasons for selecting specifically GloWbE over other corpora as well as considers the pros and cons in its use. Sections 13.3.2 and 13.3.3 discuss the second set of data and the methods used in the analysis respectively.

3.3.1 Corpus of Global Web-based English

A corpus is a collection of "naturally occurring language texts chosen to characterize a state or variety of language" (Sinclair 1991: 171), that is, it contains a large number of texts with authentic use of language and has been designed to be representative of a specific language or variety (Cheng 2012: 3; Stefanowitsch 2020: 22–3; Szudarski 2018: 6). Corpus-based studies nowadays are conducted overwhelmingly through electronic corpora which have been annotated with additional information such as metadata, textual markup, and linguistic annotation, which are information about the text itself, its formatting, and the grammatical categories of individual words respectively (McEnery & Hardie 2012: 29–30). Corpora are accessed through concordancers or web-based interfaces which allow searches of strings, also known as *nodes,* and their display in context in a line of text, which is called the *key word in context* (KWIC) format. Concordancers also support different types of corpus analysis, for example wordlists, n-gram and keyword analysis (ibid, 2; Szudarski 2018: 19–20).

The corpus used in this study is the Corpus of Global Web-based English, which comprises approximately 1.9 billion words collected from 1.8 million web pages on 340,000 websites across 20 English-speaking countries in December 2012 (GloWbE). The websites in the corpus are divided into

two categories, blogs and general websites. Blogs are informal blogs while general websites include "other web-based materials, such as newspapers, magazines, company websites" of varying registers and 20% informal blogs as it was reported to be impossible to completely separate them from other types of websites in the compilation process (Davies & Fusch 2015: 4). The distribution of texts into these categories is reported as blogs encompassing about 60% and general about 40% of all texts.

The size of GloWbE was a major deciding factor for choosing the corpus. The reason for this is that the first set of data, which is the basis for the second set, contains borrowings many of which are not very frequent according to the frequency bands of *OED*. A vast amount of material is more likely to yield borrowings that are not very commonly in use. Additionally, some of the borrowings are terms related to a special field, such as science and medicine, so a web-based corpus, which contains a wide range of topics, was deemed most suitable for this research. The exact number of websites, web pages, and words per the countries relevant to the study can be seen in the table below. The total number of words in all of the texts is almost one billion, although there is a noticeable difference in the number of words from the two Inner Circle countries and the four Outer Circle ones: the United States and Great Britain both have approximately 390 million words, while Singapore, the Philippines, Malaysia, and Hong Kong each have 40 million.

Country	**General** (may also include blogs)			(Only) **Blogs**			**Total**		
	Web sites	Web pages	**Words**	Web sites	Web pages	**Words**	Web sites	Web pages	**Words**
United States	43,249	168,771	**253,536,242**	48,116	106,385	**133,061,093**	82,260	275,156	**386,809,355**
Great Britain	39,254	232,428	**255,672,390**	35,229	149,413	**131,671,002**	64,351	381,841	**387,615,074**
Singapore	5,775	28,332	**29,229,186**	4,255	17,127	**13,711,412**	8,339	45,459	**42,974,705**
Malaysia	6,225	29,302	**29,026,896**	4,591	16,299	**13,357,745**	8,966	45,601	**42,420,168**
Philippines	6,169	28,391	**29,758,446**	5,979	17,951	**13,457,087**	10,224	46,342	**43,250,093**
Hong Kong	6,720	27,896	**27,906,879**	2,892	16,040	**12,508,796**	8,740	43,936	**40,450,291**
All countries	**107,392**	**515,120**	**625,130,039**	**101,062**	**323,215**	**317,767,135**	**182,880**	**838,335**	**943,519,686**

Table 5. The number of words in each section of GloWbE in the varieties studied.

At the time this thesis was started in 2017, GloWbE was the largest corpus in the BNC corpora, a collection of non-proprietary corpora, but there has since been the inclusion of corpora with higher word counts, namely the News On the Web corpus (henceforth NOW), which contained 11.9 billion words as of February 5, 2021 and continues to grow monthly by 180–200 million words, and the iWeb corpus, which has 14 billion words. These corpora would be excellent tools for the examination of less frequent borrowings; however, they would likely yield such a large frequency for the more common borrowings that determining the meaning of each token manually would not be an ideal approach.

GloWbE allows for search to be restricted to and ordered by country, which was considered essential for this study. There are not many corpora that have texts from as many countries as GloWbE and also support country-specific searches. For example, the aforementioned iWeb corpus does not differentiate between the countries of origin. One corpus that, however, does separate texts by country is the International Corpus of English (henceforth ICE), which is used very commonly for variationist studies. ICE contains texts from 13 countries (The ICE Project 2016), which are Canada, East Africa, Great Britain, Hong Kong, India, Ireland, Jamaica, New Zealand, Nigeria, the Philippines, Singapore, Sri Lanka, and the United States. The advantages of ICE are that the texts have been more carefully curated than those of GloWbE: The speakers and writers of the texts in the ICE corpora "are aged 18 or above, were educated through the medium of English, and were either born in the country in whose corpus they are included, or moved there at an early age and received their education through the medium of English in the country concerned" (ibid.). In the case of GloWbE, the country of origin of the texts was determined by using the *Region* option in the advanced search of Google[12] (Davies & Fusch 2015, 4). While it has been reported that Google's categorization based on region is reliable in identifying the country of origin of a website, it is impossible to determine the ages and language

[12] The parameters that Google uses to determine the country of origin of a website are "1) the IP address for the web server, 2) who links to that website, and 3) who visits the website" (Davies&Fusch 2015, 4).

backgrounds of individual writers. Additionally, ICE corpora contain a set number of specific text types of both spoken and written English (The ICE Project 2016), whereas GloWbE only partially distinguishes between blogs and general websites ignoring greater details of text types. This point is not a great disadvantage as the purpose of the study is to investigate the use of Japanese borrowings in English texts regardless of register or genre. Despite the advantages of ICE, GloWbE was chosen over it because the size of the latter was assessed to be far too small for the studying of Japanese borrowings: ICE has approximately one million words per variety (ibid.), which combined is 150 times less than the word count of GloWbE. ICE also does not include texts from Malaysia, which was considered a further drawback.

Even if GloWbE was determined to be the best alternative available for the study of Japanese borrowings, there are multiple issues that arise from its use. These issues relate to searching, tagging, the display of results, and accessibility of the source texts. One problem is posed by LS pairs that have more frequently used homographs of the same word class which increase the number of irrelevant tokens. For example, the headword *sun* '[a] Japanese unit of length, equivalent to approximately 1·19 inches (3·03 centimetres)' (*OED* s.v. *sun n.2*) is a homograph of the more common *sun* '[t]he bright celestial object at the centre of the solar system […]' (*OED* s.v. *sun n.1*). A search of *sun* yields approximately 88 thousand tokens across all of the varieties examined. The *OED* lists the first mentioned *sun* with Frequency Band 3, which means that the word occurs "between 0.01 and 0.1 times per million words in typical modern English usage" while the frequency band for the latter is 6, which corresponds to a frequency of 10 to 100 times per million words, so one can expect to find significantly more of the latter. GloWbE uses the CLAWS 7 general-purpose grammatical tagger (Davies & Fusch 2015: 5), which tags tokens by their grammatical features only ignoring semantics, so it is not possible to determine the meaning of a token without inspecting its context manually. It was deemed too time consuming for the scope of this study to examine such

cases with a high frequency of tokens but low number of relevant cases, so a different method utilizing the collocate search function was employed for these cases (see Section 13.3.2).

Another disadvantage is that letters with macrons are displayed in the text as [?] (without the brackets). The question mark is used in the search function of GloWbE as a wildcard of one character, which means that it cannot be used in a query for question marks. There are a handful of LS pairs in the data which are spelled with *ū* or *ō*, the alternative orthographic forms for the long vowels of *u* and *o* in the Revised Hepburn romanization, and are consequently affected by this. For example, the alternative orthographic form of *shoyu, shōyu*, is shown as *sh? yu* in GloWbE. There are altogether seven headwords of Japanese origin in the *OED* which are listed with a macron, *Jōdo*, *jōruri*, *Kōrin*, *Shinshū*, *Shijō*, *shō*, and *yūzen*. Additionally, there are other headwords which could be spelt with a macron, but the *OED* records them without the diacritic such as *shonen* or *ronin*, so it was necessary to include search terms which could retrieve tokens in cases that include the macron. This problem was again circumvented by using the collocate search (see Section 13.3.2).

Additionally, there is a problem that arises from the effect of the passage of time on the accessibility of the original websites. Because the texts were collected in December 2012 and the assembling of the data for this study was done in 2018-2020, there were many sites that had since become inaccessible and would become inaccessible during the compilation of the data. Although the corpus stores all of the texts, it does not include them in the original form. For example, paragraph breaks are indicated by the sign #. These two factors of accessibility and presentation in GloWbE together caused difficulties in determining the sources of tokens (see Section 13.3.3) as it is not always clear whether the writer is the same as that of a previous token. It was possible to avoid the issue to some extent by consulting the Wayback Machine, which is an archive belonging to the Internet Archive and records 475 billion web pages (The Internet Archive). However, there were still cases where a web page was not found in the archive.

In summary, GloWbE was chosen because of its robust yet manageable size and because it allows country-specific search from texts originating from the six countries on whose geographical varieties this study focuses. Other alternatives were considered, but they only possess one of the two aforementioned qualities and so could not be utilized. The corpus poses several challenges mostly in the search of relevant tokens and unavailability of texts in the original form, but collocates and the Internet Archive were employed to solve them.

3.3.2 Collecting tokens from the GloWbE

The second set of data was gathered from GloWbE accessed through the BNC corpora website, so it was necessary to decide on the search terms and settings. A list of search terms was compiled for each LS pair based on the spellings that were recorded from the *OED* for the first set of data as well as different romanization systems, especially versions of the Hepburn romanization. Another point taken into consideration was the plurality of countable nouns since singular and plural forms may be listed separately in GloWbE depending on settings. Furthermore, typos, non-standard segmentation, and non-standard spellings also cause separate listings. The search terms, or nodes, were created to yield non-standard orthographic forms to some degree for example by using the wildcards *?* and *. It is unlikely, however, that all of the possible forms were located for the complete data. Nevertheless, the number of tokens that might not have been added to the data is expected to be remarkably lower than those that were found, and their exclusion is not likely to have a great impact on the results.

Nearly all search terms begin and end in *, which is the wildcard for a string of any length. The reason for this is that non-standard segmentation between preceding or following words affect the spelling of the word or phrase in the *Frequency* tab. Non-standard segmentation may be caused by the tagger or the writer. A very common segmentation error of the tagger involves punctuation. For

example, if the text ends in the relevant token and a full stop, the tagger has interpreted the spelling of the token as [*word.*] (without the brackets) instead of simply [*word*]. Searching with *word* would only yield the latter. Another common error is the use of the slash which the tagger does not recognize as a dividing point and instead tags the string as a whole so that for example strings of the type of *word1/word2* are a separate phrase from *word1* or *word2*. A segmentation error that may happen to writers is the accidental omission of a space after a full stop, for instance. This produces a hit of the type *word1.word2.* Generally, all punctuation marks except the full stop, colon, semicolon, comma, question mark, and exclamation mark written next to a token without spacing is interpreted as being part of the token. Cases where it was not possible to use the wildcard * at the beginning and end of search terms were those that would yield a list so long that it could not be displayed. For example, the list for **ri**, which would normally be used to search *ri* '[a] traditional East Asian unit of length now equal to approx. 3.93 kilometres […] in Japan and Korea […]' (*OED* s.v. *ri n.*), is too long because it would contain all words and phrases where the string *ri* is included. In such cases, the term was searched with * before and after it separately or without the wildcards.

Searches for LS pairs that had no macrons or frequent homographs were conducted through the List display. Default options were used with the exception of the # HITS option, which was increased from the default 100 to 5000 to ensure that as many hits are displayed as possible. As noted before in Section 3.3.1, GloWbE allows for the search to be limited to a part of speech, but this function was not used as there is a possibility of tokens having been mistagged, as it is reported that the CLAWS tagger has "an error-rate of […] 1.5%, with c.3.3% ambiguities unresolved" (UCREL). The limitation would also not have significantly reduced the number of irrelevant tokens based on test searches. A node was then entered into the search bar and the button *Find matching string* was clicked. This produces a list of words and phrases which include the search term. Any word or phrase which could potentially have relevant tokens underwent a preliminary inspection, where the KWIC lines of the varieties studied were examined for Japanese borrowings. The words and phrases that had relevant

tokens were listed and their frequencies were recorded by variety from the *Frequency* tab. The tabs were deselected and selected again based on the list so that only relevant words and phrases were viewed in the *Context* tab, from which the numbers, web pages and contexts were documented in an Excel spreadsheet.

The search for LS pairs that are homographs of a frequently used word or name differs slightly from the above as it involves the use of the *Collocates* function. The function is used to search for words that occur near each other. The node for the LS pair was placed in the *Word/phrase* search bar and a search word that could appear close to and single out tokens that correspond to the LS pair, rather than the homograph, was entered into the *Collocates* search bar (or vice versa). The search terms for the *Collocates* bar were selected by examining the quotations in *OED* entries as well as performing Google searches for related terms. Additionally, *japan** was used for all LS pairs. The collocate was then searched in up to nine tokens before and after the search term in the *Word/phrase* bar. The results were recorded as described in the previous paragraph. As writers may not use any of the related terms within nine words from the LS pair, it is highly likely that all relevant cases have not been included in the data. Fortunately, the number of LS pairs which could only be searched through this method was limited to sixteen, which are *ama*, *belt*, *dan*, *ken* (in three LS pairs), *mat, mon*, *OL*, *pink*, *reap*, *Shin, sun, tan* (in two LS pairs), and *to*. The same information as in the *List* search and the search terms for each token were recorded. The latter is especially important as the same token may appear more than once due to using different words as collocates.

LS pairs which include macrons were searched through either the *List* or *Collocates* function depending on the frequency of tokens and whether or not there were other LS pairs to which the same *List* search could be used. For example, *rōnin*, an alternative spelling of *ronin*, could be searched through the *List* function by entering either *r*, which yields 41,710 tokens, or *nin*, where the corresponding number is 463 tokens. There are, however, no other LS pairs which would use either of these search terms, so it was determined that the *Collocates* function would suit better for the task.

The search was performed by entering *r* into the *Word/phrase* bar and *nin* into the *Collocates* bar. The search was limited to two words after the *r*. This produces one hit in the relevant varieties. The *Collocates* search is generally more time efficient for information collection but poses difficulties, if there are also search terms without macrons. GloWbE presents the tokens of selected words/phrases in the same order as they are found in a webpage and assigns them a running number. This facilitates marking the sources of tokens (see Section 3.3.3), as it is often easy to locate a writer's first use of an LS pair, for example. However, if there are search terms that are searched through both the *List* and the *Collocates* function, it is not possible to display the tokens of both on the same *Context* tab. This causes the numbering not to match when the tokens are moved to the same file, which slows the process of marking the sources.

A search from start to finish can be demonstrated with the LS pair *daimio* '[t]he title of the chief territorial nobles of Japan, vassals of the Mikado […]' (*OED* s.v. *daimio n.*). The search term that was decided on was **daim***. The corresponding headword has no spellings listed in the *OED*. However, the original Japanese word is transliterated as *daimyō* in Hepburn romanization, and one may expect that not all writers include the diacritic, so the form *daimyo* is also possible. Because of the treatment of macrons in the corpus, there may be relevant tokens which are spelled [*daimy?*] (without the brackets). A suitable search term for these three spellings is thus *daim**. This term also coincidentally takes into account the use of plural forms as well as segmentation and typographic errors after the token. There are indeed examples of each for *daimio* as *daimyos*, *daimyo.*, and *daimyo/governors* are all listed as separate hits from *daimyo*. Finally, the * is added to the beginning of the node to yield results with errors before the relevant token. The search produces a list of 211 words and phrases which include the string *daim*. The list was inspected and the tabs next to the relevant varieties and potentially relevant words and phrases were selected. These were specifically *daimyo*, *daimyos*, *daimio*, *daimyo*., *daimyo/governors*, and *daim*, the frequencies of which were recorded in the final data per word or phrase as seen in the *Frequency* tab. The *Context* button was

then clicked to produce a KWIC list in the *Context* tab and the number, variety, web page, and context for each token was recorded.

3.3.3 Arrangement of the second set of data

To answer the second research question about the distribution of semantic fields of Japanese borrowings in texts of Present-day English varieties, the data was to be arranged so that the frequencies of relevant tokens in each variety could be counted. The next phase was the exclusion of irrelevant tokens and marking the senses, sources, and frequencies of relevant tokens.

The data acquired from GloWbE included all tokens of a word or phrase yielded by a search term as long as there was at least one relevant token, so irrelevant tokens had to be separated from the data. Those that were excluded were non-English lexicon, proper nouns, references to the word itself, and LS pairs which are not part of the first set of data. Non-English lexicon was mostly in the form of Japanese written in romaji, which are often indistinguishable from the spellings of Japanese borrowings. For instance, hits for **ki*mon**, which was intended to yield results for *kimono*, include the phrase *kirei na kimono* 'beautiful kimono', which was used as an example in a linguistic description of Japanese adjectives. While *kimono* is also a borrowing, its use in the example phrase is in a Japanese language context and does not reflect usage in English. A similar logic was followed for NPs which included a potentially relevant token and another word of Japanese origin: if the Japanese word was not found in the *OED* and the NP was attested in a Japanese text, the token was interpreted as being part of Japanese lexicon. For example, the search term **udon** (based on the LS pair *udon* '[a] kind of noodle made from wheat flour' (*OED* s.v. *udon n.*)) yields a few tokens of *kitsune udon*, which refers to a type of udon. There is no entry for *kitsune* 'fox' in the *OED* nor is *kitsune udon* listed under the entry for *udon*, so these tokens were discarded. Occasionally there was

a phrase which indicated that the token is used in Japanese. The phrases were for example *called x in Japanese* or *known as x in Japanese* or *Japanese for y*, where x is the token and y a paraphrase. Also related to non-English lexicon, there were cases, such as *Herbal ecstasy pills here to harmful sold as tobacco and ginkgo that use it was a new level was missing*, which did not follow English grammar and were suspected to have undergone machine translation from another language. These were not included in the data.

Japanese borrowings may appear as elements in names, but as these belong to the field of onomastics, tokens which were part of a name were discarded. Some LS pairs had a notable number of tokens which were name elements. One such was *samurai*, where almost 40% of tokens were part of a name. One semantic category which was treated as an exception was names for dishes. For example, in *At the Cove, new dishes have evolved like the Salmon en Papillote, Thai Seafood Stew, Poached Salmon and Soba Noodles, Cajun Style Jambalaya and Flaming Seafood* the relevant token *Soba* appears in the name of a dish. Names of dishes tend to describe the dish and may be indistinguishable from NPs containing common nouns if not capitalized. Judgment becomes even more difficult when the same writer uses both a capitalized and lowercase spelling. It was decided that dish names will be allowed in the data because of the describing factor and difficulty of verifying whether the token is a common noun or a name element.

Tokens not corresponding to a LS pair in the first set of data were discarded. There were tokens where both the lexeme part and the sense part were not recorded in the first set, that is, the tokens were homographs of Japanese borrowings. For example, **sake** was used as a node for *saké* '[a] Japanese fermented liquor made from rice [...]' (*OED* s.v. *saké n.2*), but it also yields results for *sake* '[o]ut of consideration for; on account of one's interest in, or regard for (a person); on (a person's) account' (*OED* s.v. *sake n.1*) typically in *for the sake of* or *for one's sake*. There were also tokens whose sense differed from those in the first set of data, but the lexeme was found. These have been formed through polysemy from a Japanese borrowing, but that specific sense is not recorded in the

OED and the token had to be excluded from the data. For example, *Ronin* has a metaphorical sense that is not found in the entry for *ronin* in the *OED* in *Sherlock embodies independence, profound skill and hope. He is a true Ronin*. As the description 'a samurai without a lord or master, esp. one who lives as a mercenary warrior; a wandering samurai, an outlaw' (s.v. *ronin, n.*) is not valid for Sherlock Holmes and this is the only sense listed in the *OED* for *ronin,* these types of tokens had to be excluded from the data. If only the lexeme part of a token was not identical to that of the spellings in the *OED*, it was still included in the data. Something that should be noted is that attributive noun tokens were accepted into the data as nouns even if the *OED* did not include the *attr.* tag. If the tag was included and it was listed in a separate sense from the noun one, attributive noun tokens were documented under the LS pair with the *attr.* tag.

Because the study places great importance on the senses of lexemes, references to the lexeme itself, rather than the referent, were not considered relevant. For example, *kamikaze* in the sentence *The word " Kamikaze " has, for the most part, been abused by being misused, being a negative rather than a positive* refers to the lexeme, which is emphasized by the use of the phrase *The word* and quotation marks. In this case, there would also be an additional difficulty in determining which of the LS pairs that have *kamikaze* as the lexeme part is being referred to. There are cases such as *tsunami* in *The effect we now call tsunami has been happening since there have been oceans on the planet Earth (billions of years)*, where *tsunami* seems to refer to both the lexeme and the referent. The sentence would not be coherent with the lexeme interpretation only (**The effect we now call the word tsunami* [...]), so these types of tokens were considered relevant.

There was one other type of token that could be considered a borderline case aside from names of dishes. This was tokens referring to entities in role-playing and video games. These tokens are not proper names but refer to a class of characters, items, or locations. The difficulty arises from the fact that often the entity presents additional qualities, some of which only make sense in the context of the game, that the real-life counterpart does not. For example, in [...] *it is a little stupid to have the*

one unit of samurai retainers sit in the level one for when I attack with a single unit [...] the phrase *samurai retainers* refers to a type of character in the video game Total War: Shogun 2. Samurai retainers of Total War: Shogun 2 look outwardly like representations of samurai retainers in the real world and the setting of the game is 16th century Japan, a fitting time period for samurai. However, the samurai retainers in the game have traits such as Attack skill or Charge bonus or spawning "automatically in defense of region capitals", which real-life samurai retainers naturally lacked. This poses the question of whether these tokens should be seen as possessing a meaning that is not recorded in the *OED*, for example 'a class of characters in the video game Total War: Shogun 2', or whether their meanings are to be seen as the same as those with real-life referents. The study leans towards the latter, and these types of tokens were accepted into the data as the referents could still be described with the LS pair that is recorded in the *OED*.

The next step was to mark the senses of tokens corresponding to a LS pair which shared the same lexeme part with another pair. For the most part, this was relatively easy as the contexts were available in the corpus, but there exists a small number of tokens where it was impossible to determine which LS pair was referred to. This was especially the case when the pairs were semantically similar. For example, *kombucha* '[a] type of Japanese tea made with kelp [...]' and *kombucha* '[a]n acidic beverage produced by fermenting sweet tea with a culture of yeast and bacteria, thought to confer various health benefits (more fully kombucha tea) [...]' (*OED* s.v. *kombucha n.*) are difficult to distinguish even when the contexts are apparent, as evidenced by the sentence *Tamara lives in Sacramento with her husband and enjoys running, gardening, and brewing kombucha*. The frequencies of these types of ambiguous cases were documented per variety in the column *Sense ambiguous*.

GloWbE is able to remove duplicate texts to a degree, but "there are still duplicate texts and (more commonly) duplicate portions of text in different pages" (GloWbE) even after using several methods for the removal, so it was necessary to mark identical texts in the relevant tokens. Duplicate

texts may occur within the same site or between different sites of the same or different variety. One common case of the same text being repeated in the same site is found in blogs where the same post may appear by itself in a specific web page, in the main view, where a part of all posts is included, and in an archived version. Another common occurrence is posters in a forum or a comment section quoting other posters. Duplicates in different sites were mostly copied news stories or texts in multiple mirror sites. Any identical texts that were found during the previous stages were preliminarily marked as *Identical text* in the *Source* column, and afterwards all the texts were alphabetically ordered to find duplicates that might have been overlooked. It is likely that all duplicates were not found, but the number should be low enough not to have a great significance for the recorded frequencies. Finally, the tokens that were tagged *Identical text* were inspected to determine which of the identical texts was published first, and the tag was removed from these tokens. Dates from the site itself and the source code of the site were consulted. If the date was not included in any manner and there was no other indication, for example sourcing to another site, it was impossible to determine which of the texts was released first. These types of tokens were marked with *Identical text; cannot determine which was released first*. A special case is posed by collocate searches, where the same text may appear multiple times as the search is conducted with different words and, if the different collocate nodes occur in the same text near the borrowing, the information is retrieved more than once. These cases are thus not the same as those discussed above, as the actual token is not recorded in the corpus twice, but the searches yield a single token multiple times. Because the search terms for each token were recorded, it was possible to discard these types of duplicates without marking them as *Identical text*.

Next came the marking of the frequencies of relevant tokens and websites per word or phrase and variety. The same writer may use the same borrowing multiple times in a single text, which may give the impression that a borrowing is used more widely in the corpus than it actually is, so simply counting all the relevant tokens could skew the results. It was decided that, in addition to all the

relevant tokens, the number of what is here termed as *base occurrences* would also be counted to avoid this problem. A base occurrence is defined as each writer's first relevant token from each text component, which is a part of a text that forms an independent unit within it. In the case of GloWbE a text is all the text in a single web page. For example, the following KWIC lines for the lexeme-sense pair *sushi* contain three relevant tokens.

You've read and heard about	**sushi**	grade tuna and the different
types which are best for	**sushi**	and all those topics.
kind of fish for their	**sushi**	The bad thing is that

All of the tokens are instances of the Japanese borrowing *sushi*, so the number of relevant tokens is three. The texts are from the same website kitchenmaus.blogspot.com, which means that the number of websites is one. All of the texts are from the same web page and text component, but the first and second token are from the same writer. Thus, the number of base occurrences is two, as the second token is disregarded being from the same text component and writer as the first token. If the writer of the first two tokens were to use *sushi* on a different web page, it would be counted as a base occurrence, as the text component would have changed. Similarly, if the writer were the owner of the blog and had used *sushi* in the About me section, which is displayed on each web page of the blog and is written separately from the blog posts, the token would be counted as a base occurrence. The source of each relevant token was examined and was marked by coloring the cell in the Excel sheet if it was the first token in a website, marked as *Base occurrence* if the token was a writer's first relevant token in a text component or *Same source* if the token belonged to the same text component and was used by the same writer as a preceding one.

The frequencies of *Identical text*, *Same source* (all relevant tokens) and *Base occurrence* (first relevant token from a writer in a text component) tags and websites were then counted for each word/phrase per variety. The frequency of the tokens where it was impossible to distinguish which sense was being referred to were marked in the *Sense ambiguous* column and were not included in

the data otherwise. If it was not possible to determine whether the token was from the same writer or text component as a previous one, as in the case of anonymous authorship or when the site was no longer available (see Section 3.3.1), the token was only counted in the frequency of relevant tokens and a note about this was added. If there was difficulty in determining which of two identical texts from different sites was published first, the tokens were only counted in the frequency of relevant tokens and possibly the number of websites and a note was again added. The table below is a summary of how frequencies were counted.

	All relevant tokens	Base occurrences	Websites	Sense ambiguous	Note: Identical text; cannot determine which was published first	Note: Cannot determine if same source
first relevant token in a website	+	+	+	0	0	0
relevant token from the same site but from a different writer or text component from previously recorded	+	+	0	0	0	0
relevant token from the same writer in the same text component as previously recorded	+	0	0	0	0	0
identical text to a previous one	+	0	0	0	0	0
cannot distinguish which sense is used	0	0	0	+	0	0
cannot determine if the writer or text component is the same	+	0	0	0	0	+

cannot determine which of identical texts was published first	+	0	+/0	0	+	0

Table 6. Summary of how frequencies were counted. The x-axis corresponds to the frequencies that were marked and the y-axis to different scenarios met with the tokens. The plus sign indicates that one is added to the frequency and 0 that there is no change in the frequency. The marking +/0 in the case of *cannot determine which of identical texts was published first* means that, if the token is the first relevant token from that website, one will be added to the number of relevant websites, but if it is not the first token, there will be no change.

Ultimately the information that was documented for the second set of data was (1) search terms used in the GloWbE, (2) search term number, (3) word or phrase in GloWbE, (4) spelling of the word or phrase in singular form, as well as (5) frequency of all tokens, (6) frequency of relevant tokens, (7) frequency of base occurrences, and (8) number of websites with relevant tokens across all varieties studied and per variety, (9) number of tokens whose meanings were ambiguous, and (10) notes. Finally, the frequencies of all relevant tokens and base occurrences were used to produce the number of tokens in each main and subcategory per variety. Aside from raw frequencies, it was also necessary to produce versions with normalized frequencies as the word counts of the varieties differ significantly. The purpose of normalized frequency is to allow the comparison of corpora of different sizes by calculating how many times a word would appear per a certain number, typically a million, words in a running text (McEnery & Hardie 2012, 49). The formula used is

$$NF = (\text{number of examples of the word in the whole corpus} \div \text{size of corpus}) \times (\text{base of normalisation}) \text{ (ibid.)},$$

so for example the normalized frequency of the subcategory ARCHITECTURE in AmE texts, which had the raw frequency of 41 tokens, is

$$NF = (41 \div 386{,}809{,}355) \times 1{,}000{,}000 \approx 0.11.$$

The calculations were done similarly for all main and subcategories, which completed the arrangement of data.

4 Semantic fields of Japanese borrowings in English

This section presents the findings drawn from the first set of data, which is the basis to answering the first research question about the distribution of Japanese borrowings in English into semantic fields during different 50-year time periods. Section 4.1 analyzes the overall statistics of Japanese borrowings and their semantic fields in English and discusses the lifespan of the lexemes and referents. Section 4.2 examines the distribution of lexeme-sense pairs in semantic fields into borrowing types, and Section 4.3 how the borrowings are distributed into semantic fields in different 50-year time periods.

4.1 Overall distribution of Japanese borrowings in English

All the 705 LS pairs and their semantic categories are presented in Appendix 3. Most of the LS pairs belong under only one main category or subcategory: 529 LS pairs, which is approximately 75.0%, are associated with one main category, 156 pairs with two, 19 with three, and one pair with four. With subcategories the numbers are slightly different because the LS pairs which belong to multiple subcategories of the same main category were counted only once into the main category, as discussed in Section 3.2.3. The number of LS pairs which were in only one subcategory is 512. There are also 147 pairs that represent two subcategories, 38 that represent three, five that represent four, and three that are in as many as six subcategories. The three LS pairs are *ryu* '[a] style, method, or school of thought relating to a Japanese discipline or art form; (also) an institution or body that teaches such a style […]' (*OED* s.v. *ryu n.*), *ryugi* with the same meaning as *ryu* (*OED* s.v. *ryugi n.*), and *sensei* '[…] (a respectful title or form of address for) a teacher, master, or skilled artist. Also in extended use: a guide or mentor' (*OED* s.v. *sensei n.*).

The distribution of all LS pairs into main categories and subcategories is depicted in Figures 7 and 8 respectively. The largest main categories, with 50 or more LS pairs, are SOCIETY, ARTS AND CRAFTS, COOKING AND BEVERAGES, SPORTS AND MARTIAL ARTS, NATURE, TRADITION, MYTHOLOGY, AND RELIGION, and SCIENCE AND MEDICINE. As many as 66.6% of all LS pairs in the main categories represent one of these categories. All of the large main categories have subcategories, but this does not entirely explain the great size of the categories, as there are five more main categories with subcategories which do not have 50 or more LS pairs. Next, there are main categories that are middle-sized, having fewer than 50 but more than 20 LS pairs. These are ARCHITECTURE AND INTERIOR DESIGN, GOVERNMENT, ADMINISTRATION, AND POLITICS, BUSINESS AND FINANCE, ENTERTAINMENT AND LEISURE ACTIVITIES, CURRENCY AND UNITS OF MEASUREMENT, FASHION, MANUFACTURING AND MATERIALS, and MILITARY AND BUSHIDO. Finally, the small main categories, which have fewer than 20 LS pairs, are VALUES AND CONCEPTS, WRITING, LANGUAGE, AND BOOKBINDING, WEAPONRY, HISTORICAL PERIODS, GREETINGS, INTERJECTIONS, AND GESTURES, ATTRIBUTES AND QUALITIES, SEAFARING, VEHICLES, AND TRANSPORTATION, EDUCATION, ELECTRONICS AND TECHNOLOGY, and OTHER OBJECTS. The only one of these that has subcategories is WRITING, LANGUAGE, AND BOOKBINDING.

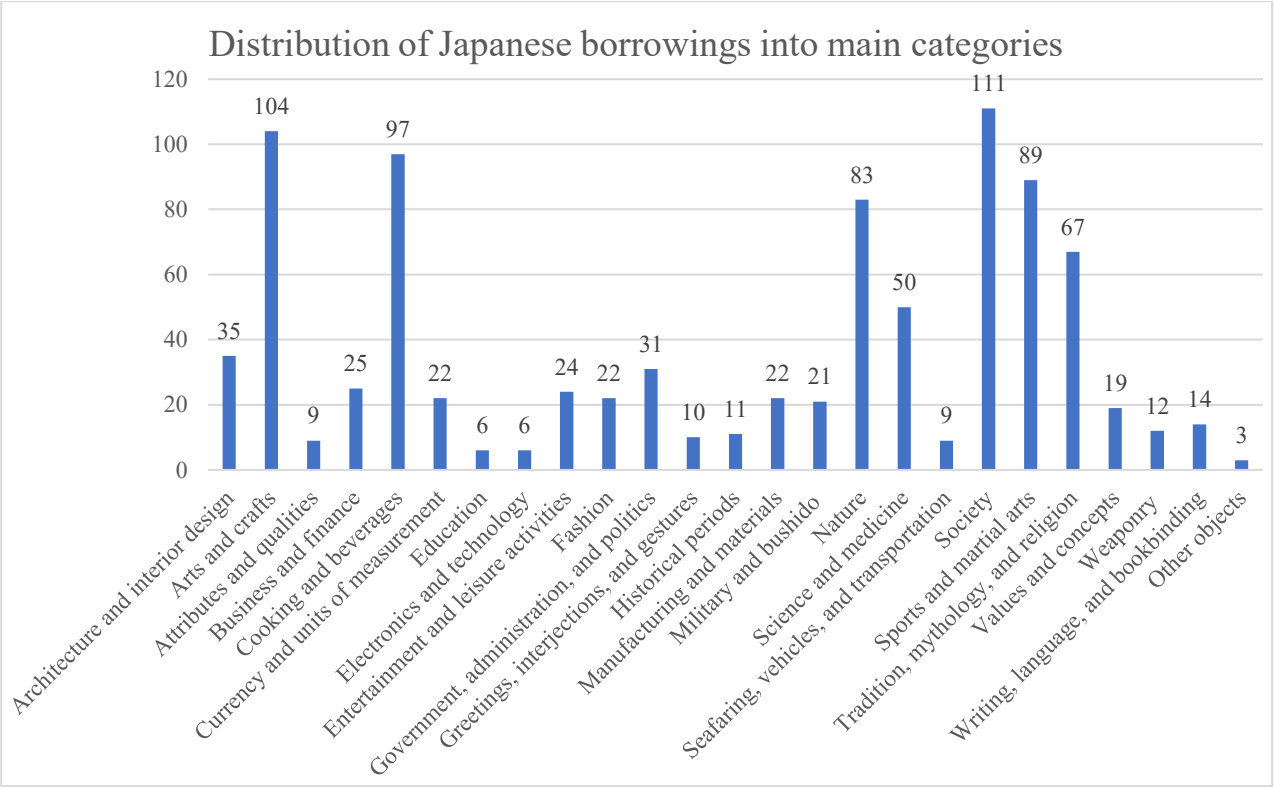


Figure 7. Distribution of Japanese borrowings into main categories.

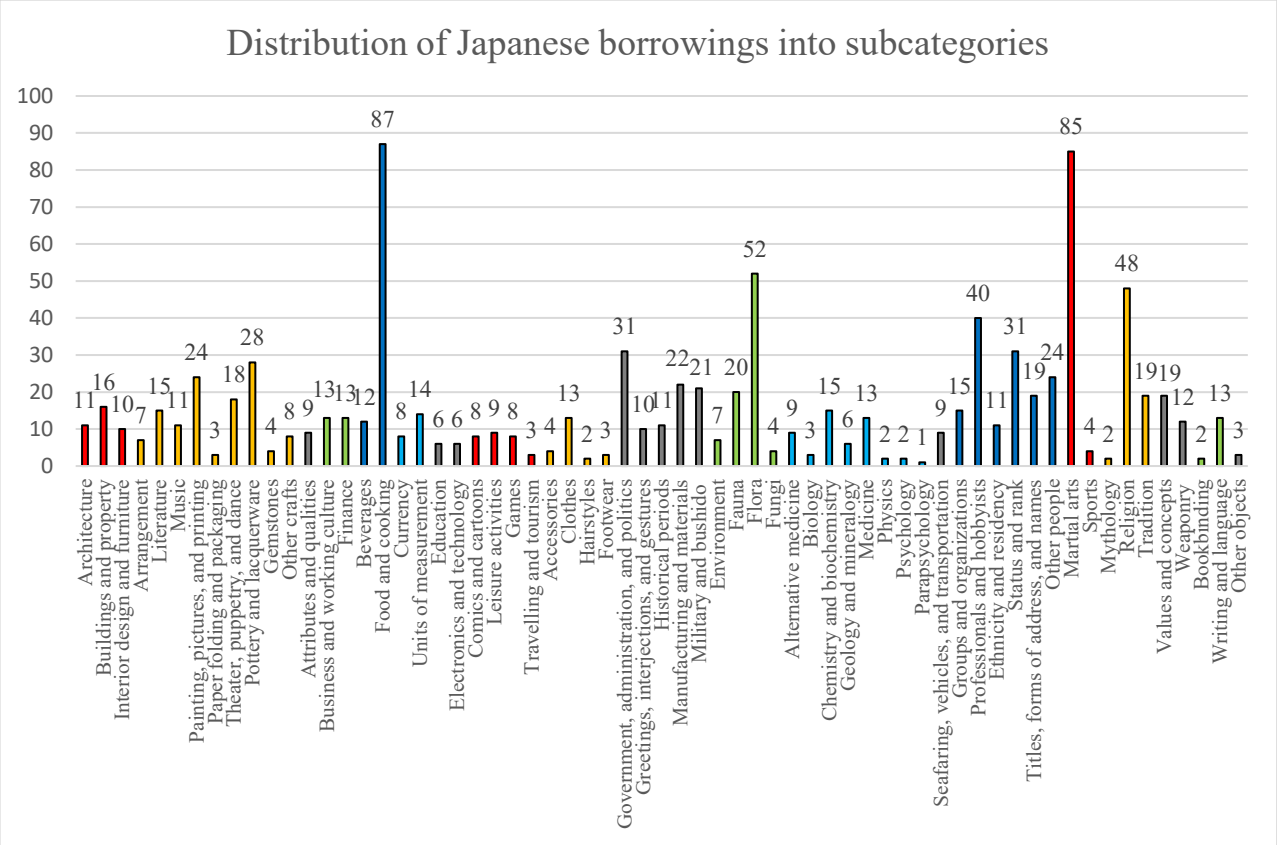


Figure 8. Distribution of Japanese borrowings into subcategories. Adjacent subcategories in the same color belong under the same main category with the exception of the ones in grey, which do not have separate subcategories.

The great majority of the borrowings are in use today, but the data also includes instances that are listed as obsolete in the *OED*, which indicates that all or some of the senses are no longer used in Present-day English. There are four lexemes which have only one sense and this sense is obsolete, so the lexeme should not be found in English anymore. The completely obsolete lexemes are *kami-sama*, *mokum*, *shimose*, and *wacadash*. All four, except *shimose*, are borrowed from a Japanese lexeme that is also the etymon of another lexeme: *kami-sama* and *kami*[13] go back to the Japanese *kami*(*sama)*, *mokum* and *mokume* to *mokume*, and *wacadash* and *wakizashi* to *wakizashi*. It is possible that the second form became more popular and eventually won over the obsolete one at least in the cases of *kami-sama* and *mokum*, where the now obsolete forms have been attested at the same time as the ones that are still in use. The word *wakizashi*, however, is first attested in 1727, 107 years later than the last attestation of *wacadash* (*OED* s.v. *wakizashi n.* and *wacadash n.*), so it seems likely that these two lexemes did not compete with one another. In addition to completely obsolete lexemes, there are two lexemes with multiple senses of which one is obsolete, so that the lexeme can appear in Present-day English in one of the meanings that is still in use. These partially obsolete lexemes are *head drop* in the sense 'a disorder (not identified) characterized by drooping of the head, occurring in parts of Japan, esp. north-eastern Honshu and eastern Shikoku' (*OED* s.v. *head n.1*) and *Yokohama* '[a] type of fine, closely-woven crêpe fabric, typically made of cotton or silk, and originally associated with Yokohama in Japan' (*OED* s.v. *Yokohama n.*). Most of the completely and partially obsolete lexemes and senses were short-lived: the time between the first and last attested uses for *mokum* (*OED* s.v. *mokum n.*) is less than a year, for the obsolete sense of *head drop* four years (*OED* s.v. *head n.1*), for *wacadash* eight years (*OED* s.v. *wacadash n.*), and for *shimose* 15 years (*OED* s.v. *shimose n.*). The obsolete sense of *Yokohama* survived slightly longer, 57 years (*OED* s.v. *Yokohama n.*), and *kami-sama* as long as 244 years (*OED* s.v. *kami-sama n.*). The obsolete lexemes and senses are not

[13]The suffix *-sama* in *kami-sama* is an honorific.

concentrated in a specific semantic category and the main categories coincide with only *head drop* and *shimose*, which belong to the SCIENCE AND MEDICINE category but are not in the same subcategory.

While obsoleteness relates to the lifespan of a lexeme or sense, the tag *historical*, or *Hist*. as it is marked in the *OED*, notes whether the referent of the LS pair still exists, as mentioned in Section 3.2.2. The question of existence is heavily related to the time period examined, and while the *OED* lists the present-day situation, it would be inaccurate to state that the referent of a lexeme is historical when examining Japanese borrowings chronologically. This is because the referent may have been extant at the time of the transfer from Japanese to English. Indeed, some of the LS pairs were first recorded in English during a time when the referent was still extant, for example *Roju* '(A member of) the council of ministers of state in Japan, esp. under the Tokugawa government (1603–1867)' from 1822 or *Minseito* '[i]n Japan: the Constitutional Democratic Party, formed in 1927 and disbanded in 1940' from 1927 (*OED* s.v. *Roju n.* and *Minseito n.*). There are also LS pairs, such as *terakoya* '[i]n the Japanese feudal period, a private elementary school of a kind established orig. in the Buddhist temples' from a text in 1909 (*OED* s.v. *terakoya n.*), that are first attested in English after the referent was no longer in existence. Additionally, some referents no longer serve the same function as previously: this is most often the case with former currencies, which are collector's items today. In these cases, however, the *OED* does not mark the referent as historical.

4.2 Distribution of types of borrowing in the main categories

Table 9 shows the types of borrowing per each main category. There are two categories, COOKING AND BEVERAGES and NATURE, which have LS pairs in all the types of borrowing. Loanwords are the most common type of borrowing in words that have been transferred from Japanese: loanwords are found in all the categories and have the greatest frequency even in cases where there are other types

of borrowings in the category. The main categories which do not have any other types of borrowings than loanwords are ATTRIBUTES AND QUALITIES, EDUCATION, GREETINGS, INTERJECTIONS, AND GESTURES, MILITARY AND BUSHIDO, SEAFARING, VEHICLES, AND TRANSPORTATION, WEAPONRY, and OTHER OBJECTS. All of these are small main categories except MILITARY AND BUSHIDO, which is middle-sized, so the size together with the commonness of loanwords may explain this lack of other types of borrowing to some extent. There are only five main categories where loanwords comprise fewer than 90% of all LS pairs. These are BUSINESS AND FINANCE, ELECTRONICS AND TECHNOLOGY, GOVERNMENT, ADMINISTRATION, AND POLITICS, SCIENCE AND MEDICINE, and VALUES AND CONCEPTS. The overall number of borrowings is relatively low in most of these categories, so statistics in such small numbers may not be very revealing. However, at least with SCIENCE AND MEDICINE and GOVERNMENT, ADMINISTRATION, AND POLITICS, which are large and middle-sized categories respectively, the percentage seems markedly low.

	Number of loanwords	Number of calques	Number of semantic loans	Number of hybrid borrowings	Number of all LS pairs
Architecture and interior design	32	2	1	0	35
Arts and crafts	100	3	1	0	104
Attributes and qualities	9	0	0	0	9
Business and finance	22	3	0	0	25
Cooking and beverages	92	2	1	2	97
Currency and units of measurement	21	0	1	0	22
Education	6	0	0	0	6
Electronics and technology	4	2	0	0	6
Entertainment and leisure activities	22	1	1	0	24
Fashion	20	0	2	0	22
Government, administration, and politics	24	6	1	0	31
Greetings, interjections, and gestures	10	0	0	0	10
Historical periods	10	1	0	0	11
Manufacturing and materials	21	0	1	0	22
Military and bushido	21	0	0	0	21
Nature	76	2	1	4	83
Science and medicine	32	3	0	15	50
Seafaring, vehicles, and transportation	9	0	0	0	9
Society	101	6	4	0	111
Sports and martial arts	85	2	2	0	89
Tradition, mythology, and religion	64	2	1	0	67
Values and concepts	17	1	1	0	19
Weaponry	12	0	0	0	12
Writing, language, and bookbinding	13	0	0	1	14
Other objects	3	0	0	0	3
All categories	826	36	18	22	902

Table 9. Types of borrowing of the LS pairs in the main categories. The number of LS pairs in all categories exceeds 705 because some pairs are in multiple main categories.

Calques have the next greatest frequency for GOVERNMENT, ADMINISTRATION, AND POLITICS as well as BUSINESS AND FINANCE and ELECTRONICS AND TECHNOLOGY. The two last-mentioned are interesting in this context as they represent some of the newest semantic fields in borrowing and they are the only categories where calques are the only type of borrowing found besides loanwords. It is possible that calques are favored in these new fields along with loanwords. In general, calques are spread rather evenly into the 14 main categories with mostly one to three in each, with the exception of the aforementioned GOVERNMENT, ADMINISTRATION, AND POLITICS and SOCIETY, which have six calques each. Some examples are *one-worldism* and *thought crime* in GOVERNMENT, ADMINISTRATION, AND POLITICS and *year man* and *Red Army* in SOCIETY.

Like calques, semantic loans spread across multiple categories: the type of borrowing is found in 13 main categories, of which most have one instance of a semantic loan. Two, FASHION and SPORTS AND MARTIAL ARTS, have two instances each, and one, SOCIETY, has four. Examples from the three categories with most semantic loans are *belt* and *black belt* from both FASHION and SPORTS AND MARTIAL ARTS and *busgirl* from SOCIETY.

Hybrid borrowings differ from the other types of borrowing in that they are found in only four main categories: COOKING AND BEVERAGES, NATURE, SCIENCE AND MEDICINE, and WRITING, LANGUAGE, AND BOOKBINDING. The hybrids are noticeably concentrated in SCIENCE AND MEDICINE, where they comprise of 30% of all borrowings. This is explained by the naming convention in geology and mineralogy, where the names of minerals have the suffix *-ite*. There are seven instances of such mineral names, some of which are *johachidolite*, *todorokite*, and *ningyoite*. Acids also have the suffix *-ic*, as in *kainic*, *ibotenic*, and *okadaic*.

4.3 Chronological distribution of Japanese borrowings

Figure 10 shows how many LS pairs were first attested in each 50-year time period starting from the year 1550. The figure shows that the rate at which borrowings were adopted was not even: the number of LS pairs that entered English in periods before the year 1850 was fewer than 50 pairs per period, at times even fewer than ten, while after 1850 there was a great surge, and the average number of borrowings becomes 200 LS pairs per 50 years. The sporadicity of first attested LS pairs in years before 1855 becomes clear in Figure 11, which illustrates the yearly cumulative sum. There is a noticeable rise in 1613–1619 and a very prominent increase in 1727 when the number of LS pairs more than doubled. After the year 1855 the rate of borrowings became more regular so that at least one new LS pair is attested in almost each year; there are only six years where this was not the case. The rate was especially high from 1870 to the end of the century. These findings on LS pairs are in line with those of other loanword studies on Japanese borrowings, although the exact frequencies differ.

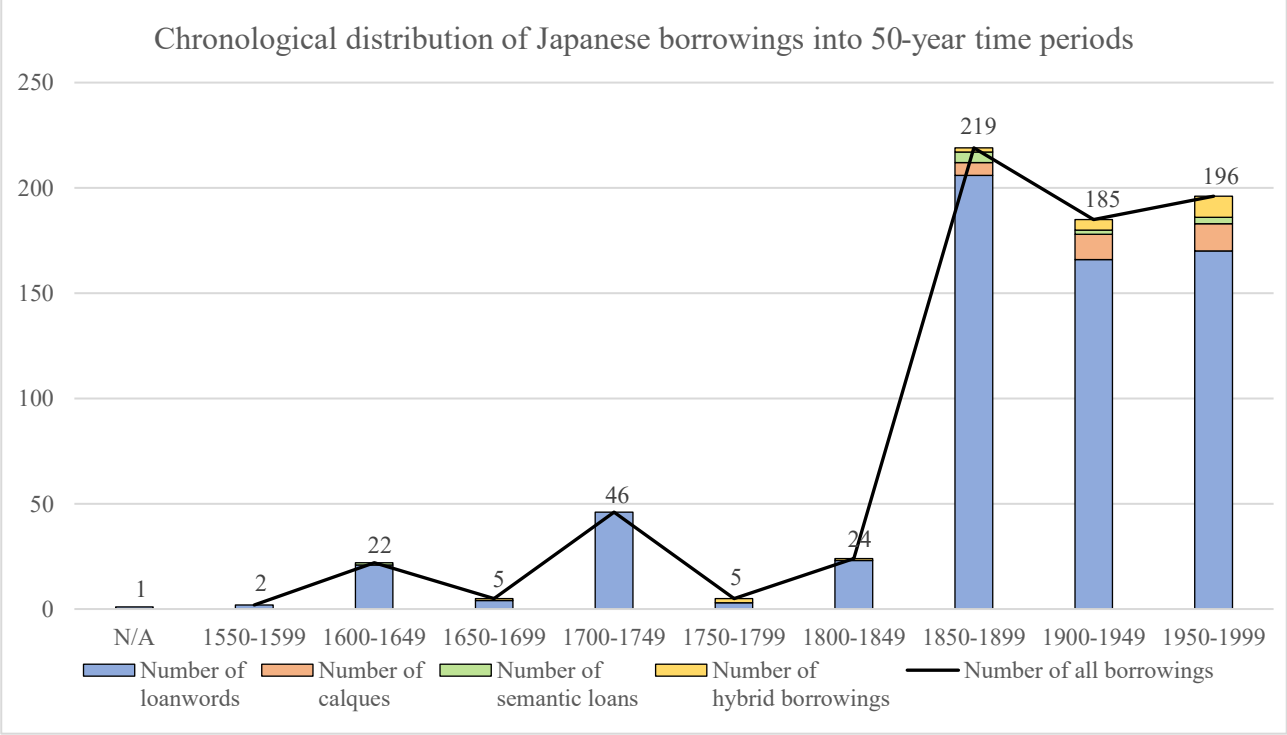


Figure 10. Chronological distribution of Japanese borrowings into 50-year time periods. The stacked bars represent the frequency of each type of borrowing and the line the overall frequency of LS pairs.

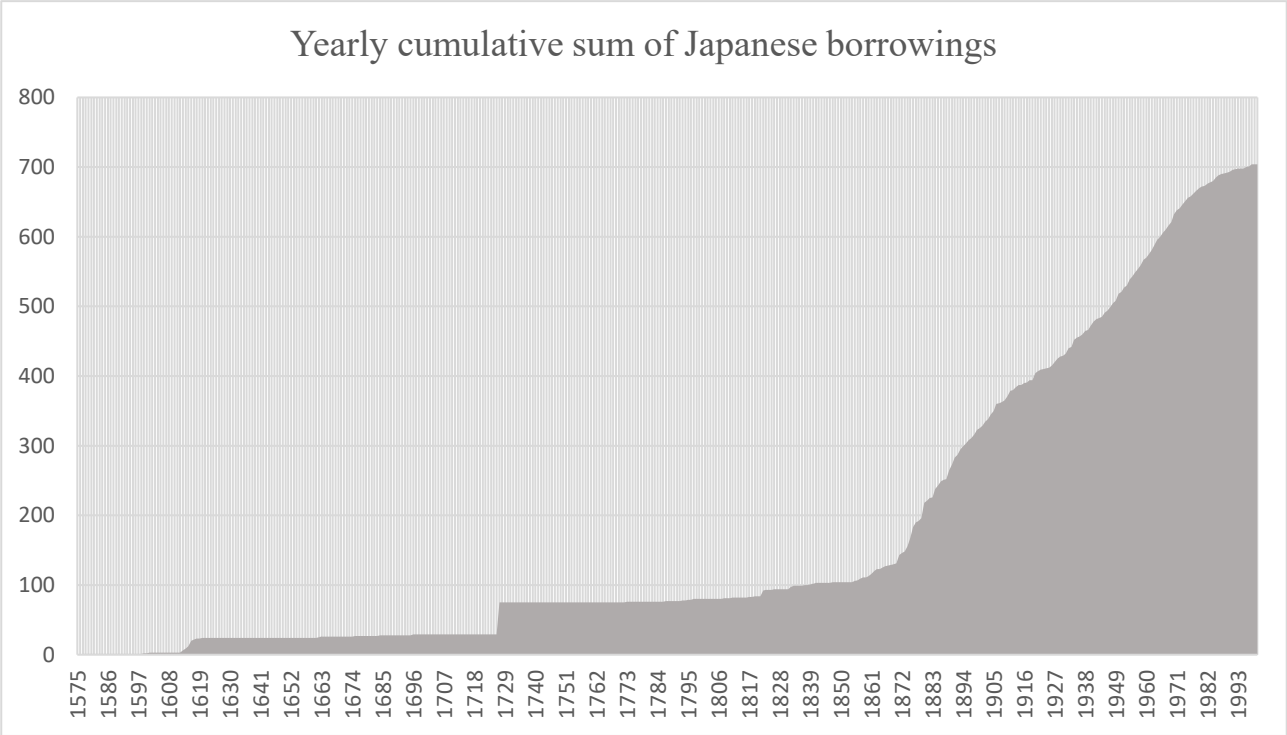


Figure 11. Yearly cumulative sum of Japanese borrowings.

The types of borrowing of the LS pairs first attested in the 50-year time periods are presented in the table below. The most frequent type of borrowing in each period is loanwords, which make up 60–100%[14] of all borrowings in a 50-year period. Loanwords are the only type that is found in all nine periods and during two of them, 1550–1599 and 1700–1749, no other borrowing type was attested. The share of loanwords out of all borrowings has, however, been in decline since the years 1850–1899. Calques have been the second most frequent type of borrowing in all periods they are found in comprising 2.7–6.6% of all borrowings. Contrary to loanwords, calques were not attested before 1850, but their numbers have been increasing in each period after that. The first calque is *rain door*, first attested in 1867. Semantic loans were the earliest type of borrowing other than loanwords that was attested. The LS pair was *mat* '[…] a unit of area equal to approx. 6 feet by 3 feet (1.8 metres by 0.9 metre), corresponding to the traditional size of a tatami' (*OED* s.v. *mat n.1*) from 1613. Semantic loans amount to 1.5–4.6% of all borrowings in the periods they were adopted. Hybrid

[14] If only periods where the number of borrowings is greater than twenty are taken into consideration, the percentage is 87–100%.

borrowings are found in six of the nine periods, which is the second most after loanwords, and they compose 0.9–40% of all borrowings in the periods in which they were first attested. The first attested hybrid is *acupuncture*, which is a calque to Latin combined with a Latinate element. Hybrids follow a similar trend to calques in that their ratio has been on the increase in comparison to other types since 1850.

	Number of loanwords	Number of calques	Number of semantic loans	Number of hybrid borrowings	Number of all LS pairs
N/A	1	0	0	0	1
1550–1599	2	0	0	0	2
1600–1649	21	0	1	0	22
1650–1699	4	0	0	1	5
1700–1749	46	0	0	0	46
1750–1799	3	0	0	2	5
1800–1849	23	0	0	1	24
1850–1899	206	6	5	2	219
1900–1949	166	12	2	5	185
1950–1999	170	13	3	10	196
All	642	31	11	21	705

Table 12. Distribution of types of borrowing of LS pairs first attested in each 50-year time period.

The following sections discuss the distribution of Japanese borrowings into semantic fields in 50-year time periods and explains how the historical situation is reflected in the number of borrowings and which semantic fields are represented in each period. The year 1850 is treated as a dividing point and 4.3.1 examines the periods before 1850 and 4.3.2 the years after and including 1850.

4.3.1 Chronological distribution of Japanese borrowings into semantic fields in time periods before 1850

The first Japanese borrowings appeared in English during 1550–1599, but only two are attested then. One is *Kuge* '[i]n feudal Japan, the name of the nobility attached to the Imperial Court at Kyoto; a court noble', which was first attested in 1577 (*OED* s.v. *Kuge n.*), and the other *bonze* '[a] term applied by Europeans to the Buddhist clergy of Japan, and sometimes of China, and adjacent countries' from 1599, although the LS pair is attested in an undomesticated context in 1553 (*OED* s.v. *bonze n.*). Both of the LS pairs have human referents and belong to the main category of SOCIETY, *Kuge* to the STATUS AND RANK subcategory and *bonze* to the PROFESSIONALS AND HOBBYISTS subcategory. Additionally, *bonze* belongs to the TRADITION, MYTHOLOGY, AND RELIGION main category. During this time period, the Portuguese and the Spanish were the only Europeans in the country. They were traders and Jesuit missionaries who had gone to preach Christianity to the Japanese (Meyer 2012: 97). This is reflected in the borrowing *bonze*, which entered English via French, to which it came from Portuguese (*OED* s.v. *bonze n.*). As the purpose of travelling to Japan was religious for the Portuguese missionaries, it is not unexpected that they would notice the clergy of another religion.

The number of borrowings increased to 22 during the years 1600–1649. The LS pairs represent 13 main categories, of which the largest is COOKING AND BEVERAGES and CURRENCY AND UNITS OF MEASUREMENT. Both of the categories have five LS pairs, for example *miso* and *bento* from COOKING AND BEVERAGES and *oban* and *tatami* from CURRENCY AND UNITS OF MEASUREMENT. The number of subcategories to which the LS pairs belong is 20. All of the LS pairs from the main category of COOKING AND BEVERAGES are in the FOOD AND COOKING subcategory, causing it to be the subcategory with the greatest frequency of pairs. Next come CURRENCY and WEAPONRY with three LS pairs each. FOOD AND COOKING has had LS pairs in each 50-year period after and including this, which makes it the category that has had lexemes and meanings transferred for the longest uninterrupted period. This is understandable as Japanese cuisine differs from its western counterpart, and food is a necessary

commodity that the foreigners visiting Japan most likely partook in. Trade between the Japanese and the Portuguese and the Dutch, who arrived in Japan at the very start of the century (Meyer 2012: 98), is a probable reason for CURRENCY comprising such a large portion of all borrowings. The English also reached the country during this time period. The English East India Company arrived in Japan but stayed only briefly leaving after slightly over ten years (Cullen 2003: 35). Even so, the general increase in the number of borrowings in this period is presumably due to the English having experienced Japan first-hand and giving an account of their experience there.

The years 1650–1699 saw a decrease in the number of borrowings that entered English from Japanese, and there are only five attested LS pairs in five main and subcategories from this period. Despite the low number, two subcategories that did not previously have any LS pairs are represented for the first time: TITLES, FORMS OF ADDRESS, AND NAMES has two borrowings, *dairi* and *kami-sama*, and ALTERNATIVE MEDICINE has similarly two, *acupuncture* and *moxa*, making it also the first subcategory under SCIENCE AND MEDICINE to have LS pairs. Two other subcategories are represented by *dairi* and *kami-sama*, which also belong to BUILDINGS AND PROPERTY and GOVERNMENT, ADMINISTRATION, AND POLITICS respectively. The last borrowing is *soy* in FOOD AND COOKING. The decline in the number of borrowed LS pairs is likely caused by stricter foreign relation policies in Japan implemented in the latter half of the 1630s. The diminished contact between the native population and foreigners is thus reflected in the few borrowings from the subsequent time period.

The period 1700–1749 has the greatest number of first attested Japanese borrowings from the years before 1850, as there are 46 LS pairs in 13 main categories and 23 subcategories. The most LS pairs are found in the main category of TRADITION, MYTHOLOGY, AND RELIGION, with 19 pairs, of which the RELIGION subcategory has the majority of LS pairs, although the other two subcategories are also represented, for the first time. The next largest main categories are NATURE with ten pairs and CURRENCY AND UNITS OF MEASUREMENT with seven. Examples from the three largest main categories include *matsuri* and *Zen* from TRADITION, MYTHOLOGY, AND RELIGION, *adzuki* and *koi*

from NATURE, and *koku* and *momme* from CURRENCY AND UNITS OF MEASUREMENT. The main categories that are represented for the first time are MANUFACTURING AND MATERIALS and WRITING, LANGUAGE, AND BOOKBINDING, which both have two LS pairs: *hinoki* and *sugi* in the former and *kana* and *katakana* in the latter. Additionally, the subcategories of PAINTING, PICTURES, AND PRINTS and GEMSTONES from ARTS AND CRAFTS, and BEVERAGES from COOKING AND BEVERAGES, were also added to the semantic categories of Japanese borrowings in this period. Something that differentiates this time period from the others is that all of the LS pairs originate in the same source, namely Kæmpfer's *The history of Japan*, which describes a wide variety of topics observed by Kæmpfer during his stay in Japan in 1690–1693 (Henshall 2014: 192). He was able to stay in Dejima, as he had joined the Dutch East Asia Company in 1686 (ibid.). As stated in Section 3.2.2, the book was published posthumously in English in 1727. More borrowings were transferred to English because of the work. The continuation of the isolationist policy is, however, likely the reason for the lack of other sources.

The number of borrowings first attested in 1750–1799 is again lower than in the previous 50-year time period, there being five LS pairs in three main categories and four subcategories that entered English. The main categories to which the LS pairs belong are ARTS AND CRAFTS, of which the subcategory MUSIC has *koto*, COOKING AND BEVERAGES, from which both subcategories are represented by *saké* in BEVERAGES and *miso* in FOOD AND COOKING, and NATURE, whose subcategory FLORA has *ginkgo* and *Nandina*. Notably, this is the only time period where the main category of SOCIETY does not have any borrowings. The low number of new borrowings can again be attributed to the isolationist policies continuing. Overall, there was very little contact between English and Japanese speakers.

During the years 1800–1849, the number of borrowings increase once more, and there are 24 borrowings in 12 main categories and 19 subcategories. The largest main category is SOCIETY with seven LS pairs, after one 50-year period with no borrowings, followed by NATURE, TRADITION,

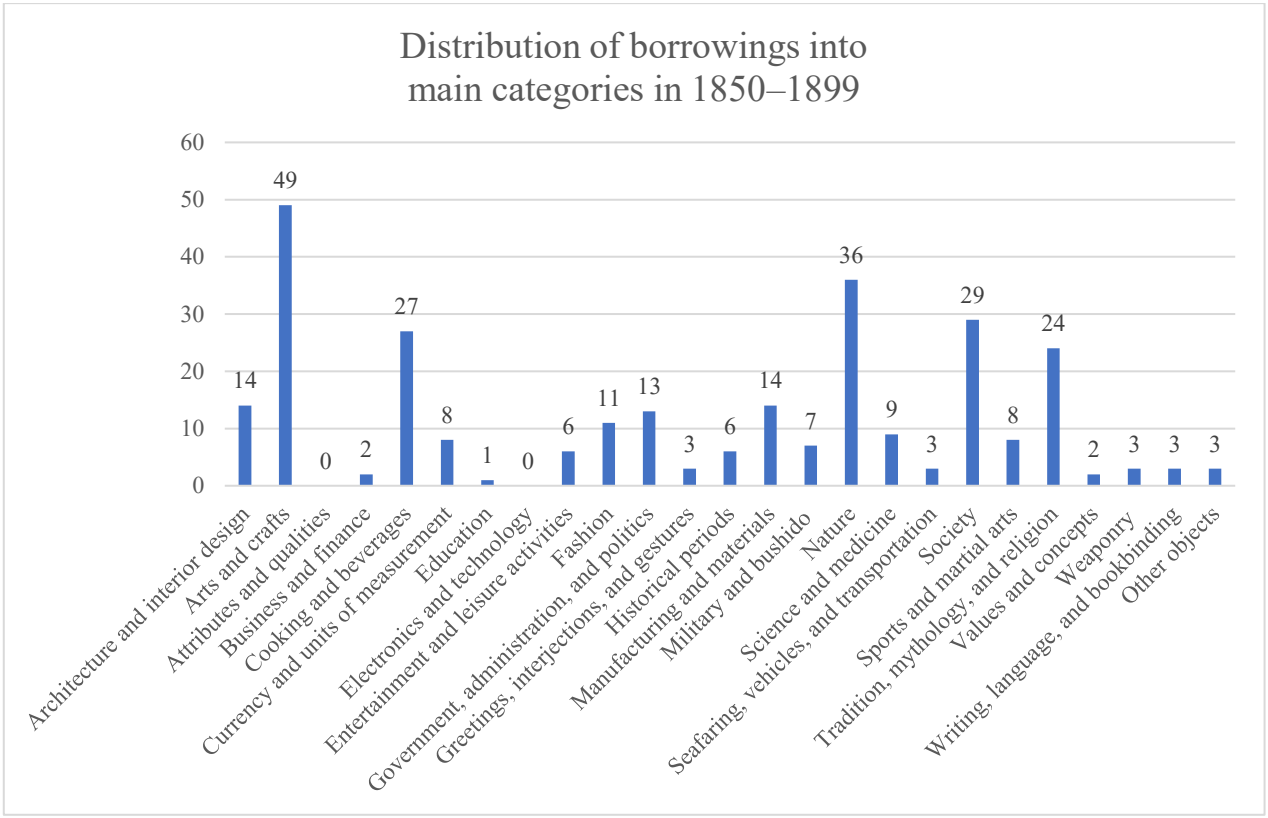
MYTHOLOGY, AND RELIGION, and WRITING, LANGUAGE, AND BOOKBINDING, which all have four pairs. The largest subcategories are FLORA and WRITING AND LANGUAGE, which both are the only ones under their respective main categories that have any LS pairs. Some of these are *ume* and *gobo* in FLORA and *hiragana* and *Ainu* in WRITING AND LANGUAGE. This time period is noteworthy in that it includes the only main category that has had the highest number of borrowings in a single time period before 1850. This category is WRITING, LANGUAGE, AND BOOKBINDING, although the number is matched later in 1900–1949. The subcategories that have LS pairs for the first time are PAPERFOLDING AND PACKAGING from ARTS AND CRAFTS, GAMES from ENTERTAINMENT AND LEISURE ACTIVITIES, and GROUPS AND ORGANIZATIONS and OTHER TYPES OF PEOPLE from SOCIETY. Encounters between English speakers and the Japanese increased during this period, but for the most part they were brief and restricted. The Opening of Japan was near, but during this time period, too, the severe policies still prevented regular contact between English and Japanese speakers and the number of borrowings remained relatively low.

4.3.2 Chronological distribution of Japanese borrowings into semantic fields in time periods after 1850

4.3.2.1 Distribution of Japanese borrowings into semantic fields in 1850–1899

The Opening of Japan in 1854 brought about a remarkable increase in Japanese borrowings. The period of 1850–1899 is the one with the highest number of borrowings entering English in any 50-year period: 219 new borrowings belonging to 23 main categories and 52 subcategories. As can be seen in Figure 13, the range of semantic categories is wide with only ATTRIBUTES AND QUALITIES and ELECTRONICS AND TECHNOLOGY not being represented. The main category with the most borrowings is ARTS AND CRAFTS with 49 LS pairs spanning all subcategories except PAPER-FOLDING AND

PACKAGING. The subcategories that have LS pairs for the first time are ARRANGEMENT, which has for example the LS pairs *nageire* and *rikka* relating to flower arrangement, LITERATURE with pairs such as *haiku* and *tanka* focusing on poetry, THEATER, PUPPETRY, AND DANCE represented by *Kabuki* and *Noh*, and OTHER CRAFTS, examples of which are *mokume* and *shibuichi* from the field of metalwork. The next main category in size is NATURE with 36 LS pairs, which are spread across all subcategories. Until this point, only FAUNA and FLORA had been represented, but there are also LS pairs such as *tsunami* and *shiitake* respectively from ENVIRONMENT and FUNGI in this period. FLORA is still, however, the largest subcategory with 22 pairs accounting for 61% of all pairs in the main category. Some of these are *kombu*, *sakura*, and *soy*. SOCIETY comes third with 29 LS pairs across all subcategories but ETHNICITY AND RESIDENCY. The subcategories with most pairs are PROFESSIONALS AND HOBBYISTS and STATUS AND RANK, which have 11 and 12 borrowings respectively making up 79% of all LS pairs in the main category. An example of the former is *tea man* and of the latter *tycoon* '[t]he title by which the shogun of Japan was described to foreigners'. Notable is that the period has the largest ratio of borrowings with female referents, some of which are *geisha*, *miko* and *mousmé*. All aforementioned main categories have at least four subcategories and, as noted, the LS pairs are spread across almost all of them. This is why the subcategory with the most LS pairs belongs to the next largest main category COOKING AND BEVERAGES, which only has two subcategories. This subcategory is FOOD AND COOKING with 26 LS pairs, for example *sushi*, *soba*, and *tofu*.



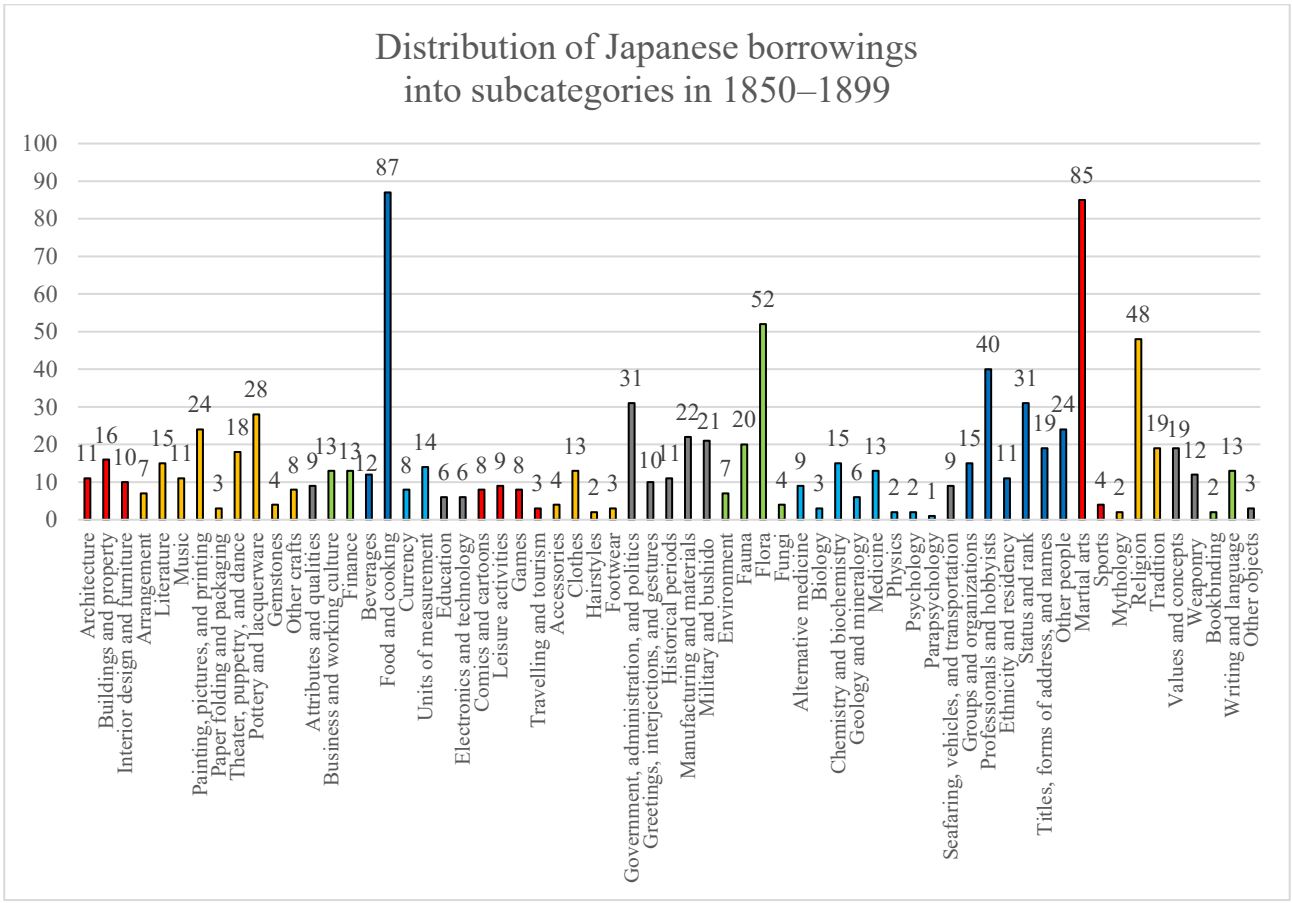


Figure 13. Distribution of the Japanese borrowings first attested in the years 1850–1899 into main and subcategories.

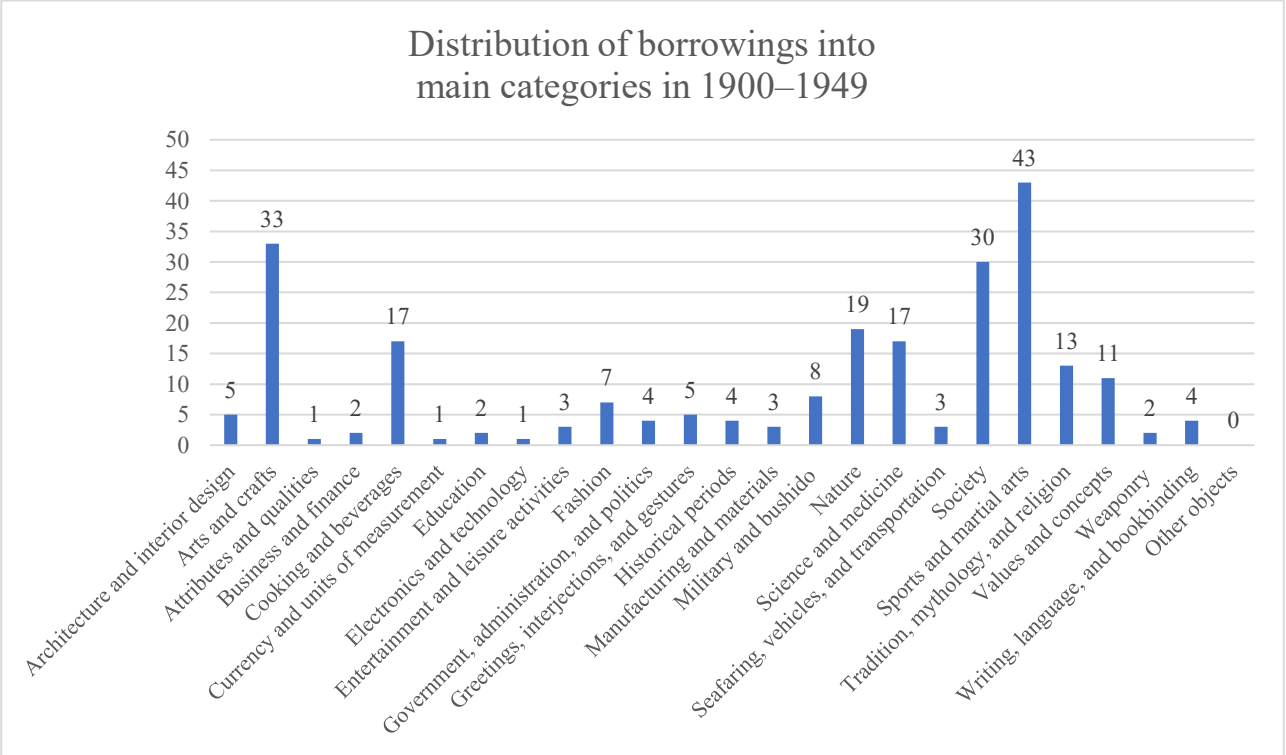
All main categories have more or the same number of LS pairs in 1850–1899 as in the previous periods. Many main categories reach their peak in this period: ARCHITECTURE AND INTERIOR DESIGN with 14 LS pairs, ARTS AND CRAFTS with 49, CURRENCY AND UNITS OF MEASUREMENT with eight, FASHION with 11, GOVERNMENT, ADMINISTRATION, AND POLITICS with 13, HISTORICAL PERIODS with six, MANUFACTURING AND MATERIALS with 14, NATURE with 36, OTHER OBJECTS with three, and TRADITION, MYTHOLOGY, AND RELIGION with 24. SEAFARING, VEHICLES, AND TRANSPORTATION and WEAPONRY also qualify both with three LS pairs although in both cases the number is matched in a later period. New main categories are BUSINESS AND FINANCE with two LS pairs, *inkyo* and *tycoon* '[a]n important or dominant person, esp. in business or politics; a magnate' (*OED* s.v. *tycoon n.*), in the BUSINESS AND WORKING CULTURE subcategory, EDUCATION with the pair *sensei* '[…] (a respectful title or form of address for) a teacher, master, or skilled artist. Also in extended use: a guide or mentor' (*OED* s.v. *sensei n.*), GREETINGS, INTERJECTIONS, AND GESTURES with *sayonara* '[…] used to express good wishes when parting; 'goodbye'' (*OED* s.v. *sayonara int., n., and adj.*), HISTORICAL PERIODS, which has for example *Heian*, *Meiji*, and *reign name*, OTHER OBJECTS with *hibachi*, *soroban*, and *uchiwa*, SPORTS AND MARTIAL ARTS, where half of the LS pairs such as *maegashira* and *yokozuna* relate to sumo wrestling and the other half to other martial arts such as *judo* and *ju-jitsu*, and VALUES AND CONCEPTS with two pairs, *bushido* and *ki*. There are also subcategories that have not been represented before: SCIENCE AND MEDICINE has LS pairs in subcategories other than ALTERNATIVE MEDICINE for the first time. These are CHEMISTRY AND BIOCHEMISTRY, GEOLOGY AND MINERALOGY, and MEDICINE, examples of which include *shikimic*, *shakudo*, and *kakke* respectively. The other subcategories are TRAVELLING AND TOURISM from ENTERTAINMENT AND LEISURE ACTIVITIES, HAIRSTYLE from FASHION, and BOOKBINDING from WRITING, LANGUAGE, AND BOOKBINDING,

The treaties signed in the latter half of the 19th century between Japan and the United States and Japan and the Great Britain starting from that of 1854 enabled citizens from the English-speaking countries to travel to Japan and stay in (Cullen 2003, 180). The increased interaction between English

and Japanese speakers is reflected in a greater amount of literature where borrowings belonging to different semantic fields are found. The Asiatic Society of Japan was founded in 1872 “to collect and publish information on subjects relating to Japan and other Asiatic Countries” (The Asiatic Society of Japan 2021) and the publications of the organization contributed 21 LS pairs across 13 main categories in this time period. Individual writers also use previously unattested LS pairs, but there is no single author responsible for a great majority of the pairs like Kæmpfer in 1700–1749, and the number of first attested pairs tends to be below ten per person. ARTS AND CRAFTS was not represented to any greater extent in the 50-year periods prior to 1850 most likely because English speakers had had restricted access to art in Japan, so it is understandable that the main category would have a great surge in borrowings when moving into the country became less limited. The large number of LS pairs in the NATURE category may reflect a general interest in biology, which may have been amplified by the release of Charles Darwin’s *On the Origin of Species* in 1859. SPORTS AND MARTIAL ARTS is interesting in that the first LS pairs in the main category appear as late as 1850–1899 unlike all the other large main categories, which were first represented by 1650–1699 the latest. Although Japanese martial arts have a long history, this 50-year period is notable in that judo, which became popular in the West, was developed from jujutsu by Dr. Jigorō Kanō during this period (Long 2001: 210–2). The judo was successfully propagated to the West, as the martial art is said to have been the first eastern one to be properly accessible to audiences outside of Japan owing partly to Kanō’s fluency in English and familiarity with Western education (ibid.).

4.3.2.2 Distribution of Japanese borrowings into semantic fields in 1900–1949

The number of borrowings decreases slightly from the previous period in 1900–1949: there are 185 LS pairs first attested during this period. This dip may reflect the general trend in Present-day English towards new vocabulary being formed through blending or combining rather than being borrowed (Fennell 2001: 177). However, the range of semantic fields is wider than in any other 50-year period, as is evidenced by Figure 14, where only OTHER OBJECTS is not represented. The largest main category is SPORTS AND MARTIAL ARTS with 43 LS pairs, all of them in the MARTIAL ARTS subcategory. The pairs include names of martial arts such as *kendo* and *kyudo*, terminology for related attire, rank, and techniques, for example *judogi*, *dan*, and *body drop*, as well as places and people such as *dojo* and *rikishi*. The word *martial arts* itself is also first found in writing in this period. Next in number are ARTS AND CRAFTS and SOCIETY, which have 33 and 30 LS pairs respectively. Some of these are *ikebana* in ARRANGEMENT, *sumi-e* in PAINTING, PICTURES, AND PRINTING, *origami* in PAPER FOLDING AND PACKAGING, *honcho* in STATUS AND RANK, and *nisei* and *sansei* in ETHNICITY AND RESIDENCY.



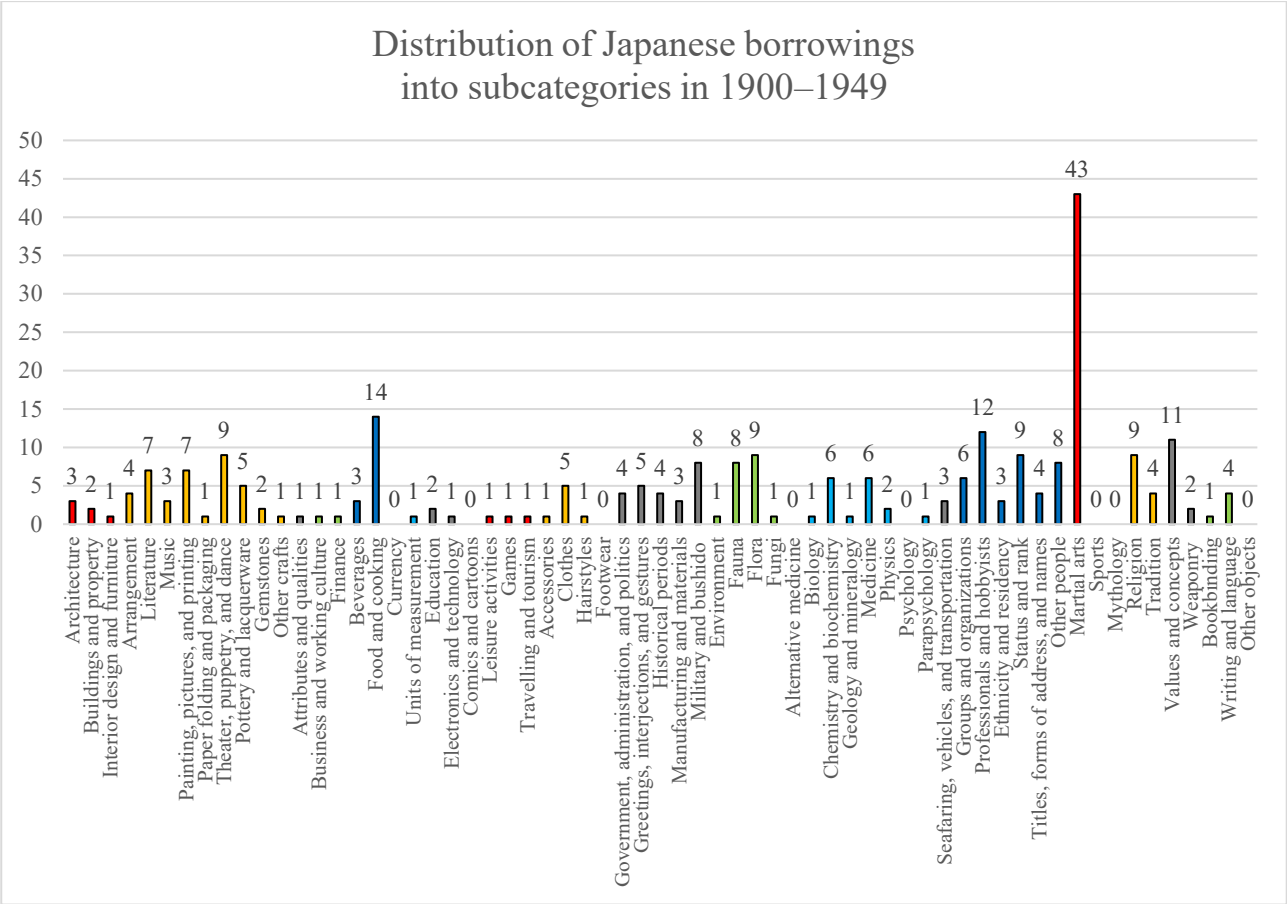


Figure 14. Distribution of the Japanese borrowings first attested in the years 1900–1949 into main and subcategories.

Besides SPORTS AND MARTIAL ARTS, the period of 1900–1949 was the peak for GREETINGS, INTERJECTIONS, AND GESTURES with five LS pairs, MILITARY AND BUSHIDO with eight, and VALUES AND CONCEPTS with 11. Additionally, SEAFARING, VEHICLES, AND TRANSPORTATION reached the same number of LS pairs as in 1850–1899 and WRITING, LANGUAGE, AND BOOKBINDING the same as in 1800–1849, which are the highest figures for each category. The main categories that had LS pairs for the first time are ATTRIBUTES AND QUALITIES and ELECTRONICS AND TECHNOLOGY, which are the last categories that lacked representation. The LS pair in the former is *banzai* '[…] (as if) shouting 'banzai', uproarious, jollificatory […]' and the latter *Yagi*. New subcategories are FINANCE from BUSINESS AND FINANCE with *narikin*, and BIOLOGY, PHYSICS and PARAPSYCHOLOGY from SCIENCE AND MEDICINE, the first area witnessing *aucuba* 'aucuba mosaic […], a mosaic disease which attacks

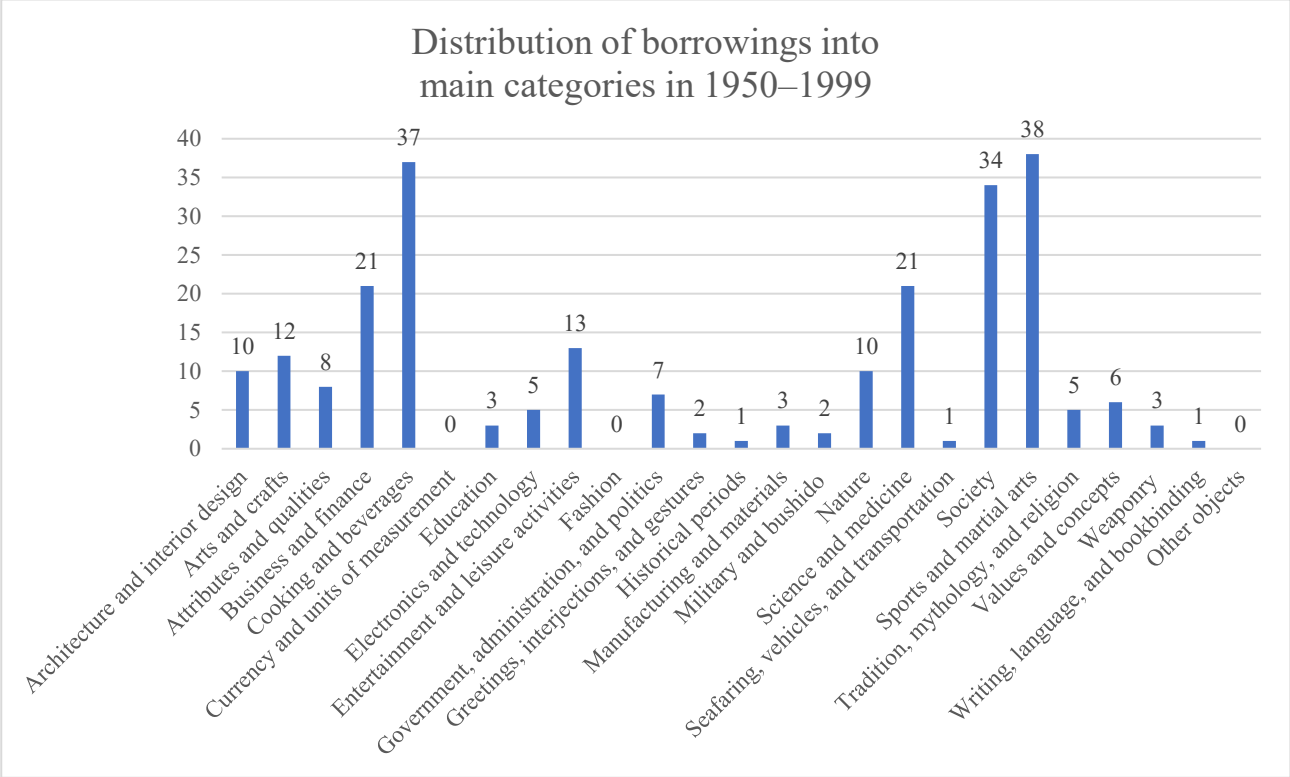
the leaves of solanaceous plants' (*OED* s.v. *aucuba n.*), the second *Kikuchi* and *Yukawa*, and the last *thoughtography*.

Emigration from Japan to America started in the previous period (Cullen 2003: 234). LS pairs such as *nisei* and *sansei* referring to Japanese Americans of a specific generation reflect this social situation. The growth in the popularity of martial arts, especially judo, is apparent as SPORTS AND MARTIAL ARTS is the largest category and most of the LS pairs first attested are names of judo techniques. What facilitated the growing interest in martial arts was President Theodore Roosevelt appointing Yoshiaka Yamashita, Kano's senior student, his judo instructor in 1904 (Long 2001: 213). A major event that shaped the relationship of English-speaking countries and Japan was the Second World War, which had the effect of increasing LS pairs in the category of MILITARY AND BUSHIDO. There are for example three LS pairs where *kamikaze* is in the lexeme part, and although not in the MILITARY AND BUSHIDO main category, *honcho* and *pan-pan* are noted in the *OED* to have been first adopted by Allied prisoners of war and soldiers during or after the war (*OED* s.v. *honcho n.* and *pan-pan n.*). Unlike in the previous 50-year period, there is no organization or author who contributes a large number of new LS pairs and instead writers use few previously unattested pairs, at most around five.

4.3.2.3 Distribution of Japanese borrowings into semantic fields in 1950–1999

The number of borrowings increased slightly to 196 LS pairs in 1950–1999. All main categories except CURRENCY AND UNITS OF MEASUREMENT, FASHION, and OTHER OBJECTS are represented as seen in Figure 15. SPORTS AND MARTIAL ARTS continues to be the main category with the most LS pairs, 38 in total, closely followed by COOKING AND BEVERAGES with 37 LS pairs and SOCIETY with 34 pairs. The number of pairs related to martial arts stays high, but SPORTS is also represented for the

first time having four LS pairs, *kamikaze* '[a] reckless or dangerous way of surfing; spec. a usually deliberate wipeout […]', *keirin*, *sayonara* '[…] designating a hit, esp. a home run, that wins the game', and *Tsukahara*. The subcategory FOOD AND COOKING is once more the largest subcategory with 34 LS pairs, for example *ramen*, *umami*, and *Wagyu*. There are also four LS pairs that denote establishments, *izakaya*, *ryotei*, *sobaya*, and *sushiya*. SOCIETY had more pairs than in any other time period. The subcategory with the most LS pairs is PROFESSIONALS AND HOBBYISTS, where seven of the 15 pairs also belong to the MARTIAL ARTS subcategory.



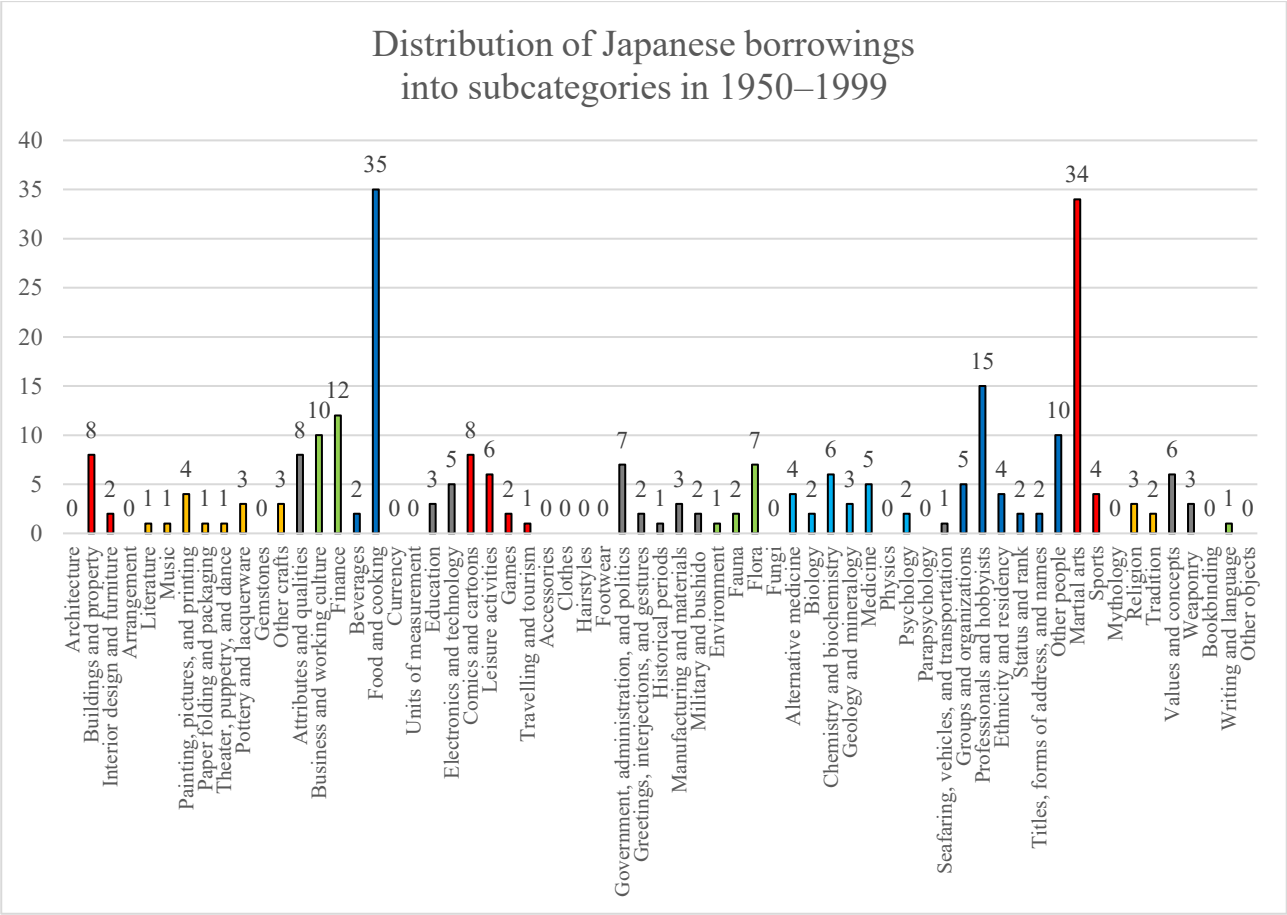


Figure 15. Distribution of the Japanese borrowings first attested in the years 1950–1999 into main and subcategories.

The main categories which peaked in 1950–1999 besides COOKING AND BEVERAGES and SOCIETY are ATTRIBUTES AND QUALITIES with eight LS pairs, BUSINESS AND FINANCE with 21, EDUCATION with three, ELECTRONICS AND TECHNOLOGY with five, ENTERTAINMENT AND LEISURE ACTIVITIES with 13, and SCIENCE AND MEDICINE with 21. WEAPONRY has three LS pairs, which is the same number as in 1600–1649 and 1850–1899. As noted in the previous section, there are no new main categories in this period, but the subcategories which had LS pairs for the first time are the aforementioned SPORTS as well as COMICS AND CARTOONS from ENTERTAINMENT AND LEISURE ACTIVITIES, which has borrowings such as *anime*, *cosplay*, and *manga*, and PSYCHOLOGY from SCIENCE AND MEDICINE, which has *hikikomori* and *Morita*.

SPORTS AND MARTIAL ARTS continues to have the most LS pairs although there are fewer of them than in the previous time period. The practice of martial arts was banned in Japan during the

Allied Occupation until 1952 (Friday 2001: 302). Despite this, judo has been an Olympic sport since 1964. There are 11 LS pairs, all of which belong to the SPORTS AND MARTIAL ARTS category, that are first attested in the writings of E.J. Harrison, but otherwise nearly all of the LS pairs of this time period are attested in the texts of separate authors. BUSINESS AND FINANCE sees a noticeable increase in LS pairs compared to the earlier periods, which may reflect the growth of Japan as an economic power after the war. It is also notable that some of the LS pairs in especially ENTERTAINMENT AND LEISURE ACTIVITIES, which mostly appeared in the latter half of the period, are first attested in Usenet newsgroups showing a precedent to how the World Wide Web would facilitate the diffusion of linguistic innovations later.

5 Japanese borrowings in Present-day English texts

The following sections discuss the findings from the corpus data. Section 5.1 outlines the overall distribution of the Japanese borrowings, which were gathered from the *OED*, into semantic fields in Present-day English texts of different varieties. Section 5.2 is dedicated to individual semantic fields based on the main categories that were established in Section 3.2.3. A figure of all relevant tokens and base occurrences per variety in the main categories and their subcategories is presented in Appendix 6. The frequencies in the main categories may not match the sum of those in the subcategories because the tokens that were instances of LS pairs belonging to multiple subcategories of the same main category were counted only once into the frequency of the main category.

5.1 Overall distribution of Japanese borrowings into semantic fields in Present-day English texts

There are all in all 73,881 relevant tokens that are instances of the LS pairs from the *OED* in the main categories. The overall distribution of the tokens across the main categories is depicted in Figure 16. There are five main categories, COOKING AND BEVERAGES, ENTERTAINMENT AND LEISURE ACTIVITIES, NATURE, SPORTS AND MARTIAL ARTS, and SOCIETY, which have over 5,000 relevant tokens, eleven categories, BUSINESS AND FINANCE, SCIENCE AND MEDICINE, CURRENCY AND UNITS OF MEASUREMENT, TRADITION, MYTHOLOGY, AND RELIGION, GOVERNMENT, ADMINISTRATION, AND POLITICS, ARTS AND CRAFTS, MILITARY AND BUSHIDO, VALUES AND CONCEPTS, ARCHITECTURE AND INTERIOR DESIGN, WRITING, LANGUAGE, AND BOOKBINDING, and ATTRIBUTES AND QUALITIES, which have between 1,000 to 5,000 tokens while the remaining nine have fewer than 1,000 tokens. All of the main categories are represented although there are some subcategories which are not. The main category with the most tokens is COOKING AND BEVERAGES, which has 14,523 relevant tokens and

8,239 base occurrences, and the smallest is OTHER OBJECTS, which has eight tokens and six base occurrences.

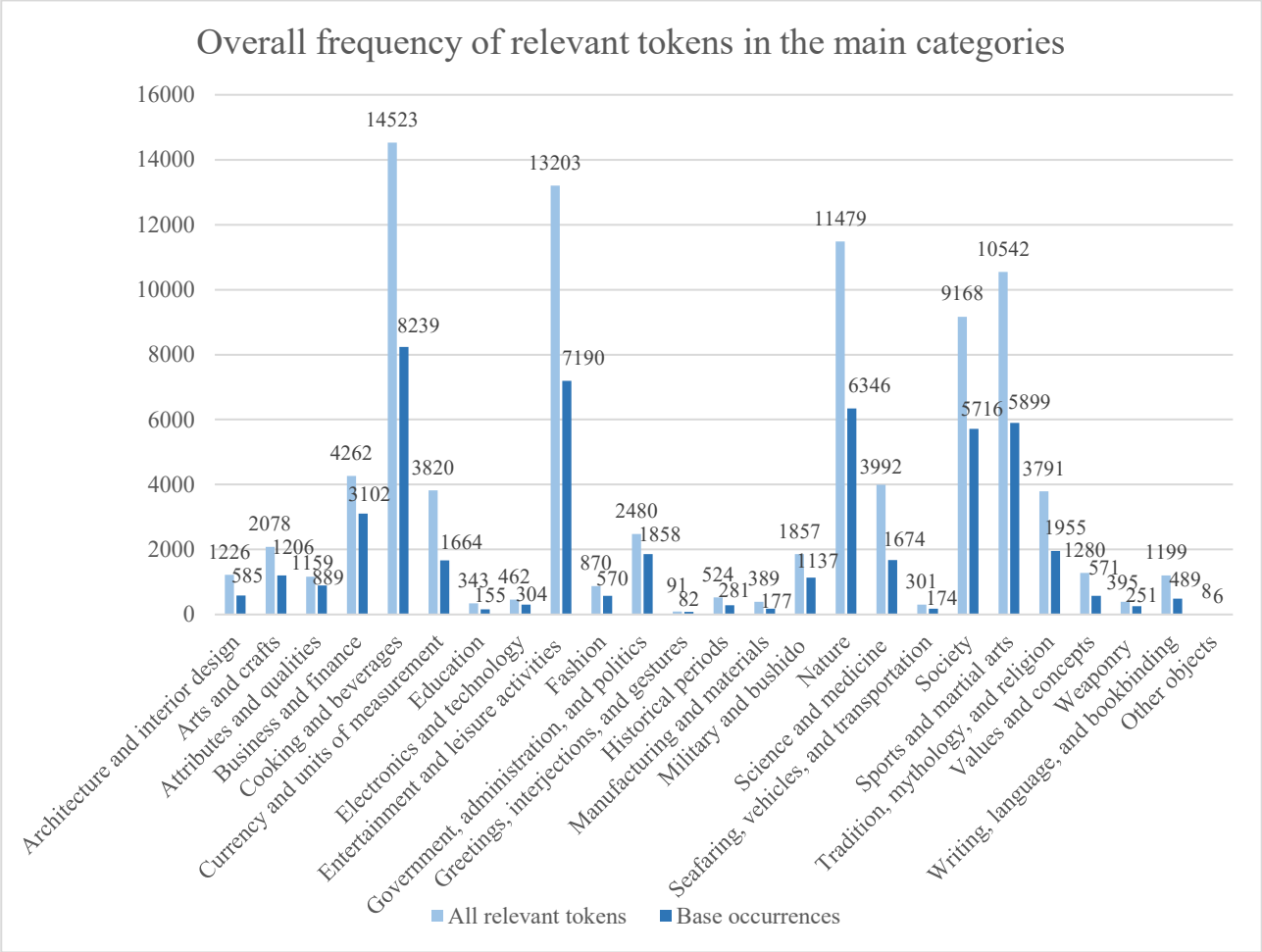


Figure 16. Overall frequency of relevant tokens in the main categories.

If the distribution of the LS pairs from the *OED* into each main category is compared to the frequency of tokens found in the corpus data, it can be noted that the number of LS pairs in a category does not necessarily correlate with the number of tokens, as seen in Figure 17. That is to say, a high number of different LS pairs in a semantic category does not guarantee that it yields a high number of tokens. For example, ARTS AND CRAFTS has the second highest number of different LS pairs, 104, but is eleventh in the frequency of tokens. Even if the fact that not all LS pairs are necessarily used in Present-day English texts is taken into consideration, ARTS AND CRAFTS has fewer tokens than BUSINESS AND FINANCE despite the latter having only 14 different LS pairs with relevant tokens in comparison to the 59 pairs in the former. Conversely, ENTERTAINMENT AND LEISURE ACTIVITIES has

overall 24 LS pairs, of which 21 have relevant tokens, but has the second highest frequency of tokens. There are of course also main categories such as COOKING AND BEVERAGES or SPORTS AND MARTIAL ARTS that have both a high number of borrowings and tokens, but the high frequency cannot always be attributed to a great number of different LS pairs. There are cases where the distribution of tokens is only concentrated to a few LS pairs. These will be discussed in later sections.

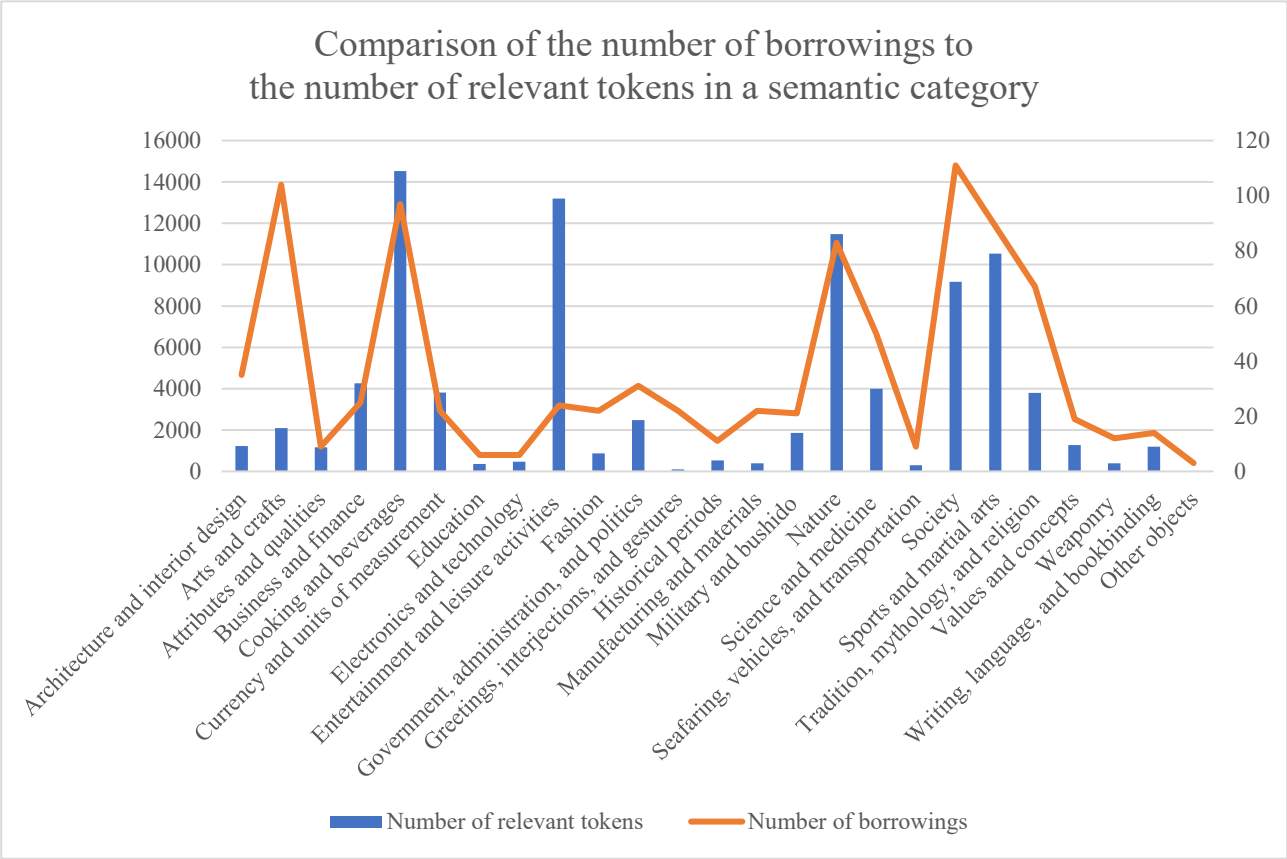


Figure 17. Number of borrowings versus number of relevant tokens in the main categories. The number of borrowings is shown by the orange line, which corresponds to the right y-axis. The number of relevant tokens is illustrated by the blue bars, which relate to the left y-axis.

The number of LS pairs which have at least one relevant token is 463, which is approximately two-thirds of all 705 pairs. This raises the question of whether the period in which the LS pairs are first attested has an influence on whether they are found in Present-day English texts. Below is a figure which illustrates how many LS pairs have relevant tokens and how large a percentage these comprise of the number of all pairs adopted in the same 50-year time period. The number of new LS pairs is so low in the periods 1550–1599, 1650–1699, and 1750–1799 that the percentages of LS pairs

found in Present-day English texts will likely not aid in answering the question, but the remaining periods show that there is no trend of older pairs being in use less than newer ones. To the contrary, of the LS pairs that were first attested in 1600–1649 or 1700–1749, over 70% are still in use. The percentage declines in the periods after the year 1800 and is the lowest in 1900–1949, where it is approximately 59%. The percentage increases again in 1950–1999 to 74%. It is, however, possible that there is a downward trend where newer LS pairs are falling out of use more than older ones, as seen in the periods between 1800–1949. Due to the texts in the corpus having been collected in 2012, the borrowings from 1950–1999 may still be new enough not to have declined significantly, but some of them may fall out of use as time passes in the next 50-year period.

	Number of all LS pairs	**Number of LS pairs with relevant tokens**	**Number of LS pairs without relevant tokens**	**Percentage of LS pairs with relevant tokens from all LS pairs from the period**
N/A	1	0	1	0.0
1550–1599	2	1	1	50.0
1600–1649	22	16	6	72.7
1650–1699	5	3	2	60.0
1700–1749	46	35	11	76.1
1750–1799	5	5	0	100.0
1800–1849	21	14	7	66.7
1850–1899	222	136	86	61.3
1900–1949	185	109	76	58.9
1950–1999	196	144	52	73.5

Table 18. Number of LS pairs with relevant tokens per the 50-year time period the borrowings were first attested in.

The twenty most frequent borrowings can be seen in Table 19 below. The twenty LS pairs represent 12 main categories and together account for 53% of all tokens. The borrowings in the top ten are *quantitative easing* from BUSINESS AND WORKING CULTURE, *yen* from CURRENCY AND UNITS OF MEASUREMENT, *sushi* and *tofu* from COOKING AND BEVERAGES, *anime* and *manga* from ENTERTAINMENT AND LEISURE ACTIVITIES, *tsunami* from NATURE, *acupuncture* from SCIENCE AND

MEDICINE, *martial art* from SPORTS AND MARTIAL ARTS, and *Zen* from TRADITION, MYTHOLOGY, AND RELIGION. The most frequent borrowing is *tsunami*, which has 6,318 relevant tokens comprising 7.1% of all tokens. The frequency is so high largely because of the Tōhoku earthquake and tsunami of 2011, which happened only a year before the compilation of the corpus, and to some extent because of the Indian Ocean earthquake and tsunami of 2004. A corpus compiled prior to these events would likely have a different LS pair as the most frequent pair. The reason why ENTERTAINMENT AND LEISURE ACTIVITIES is the second largest main category despite only having 22 LS pairs is apparent from the figure: four of the pairs, *anime*, *manga*, *karaoke*, and *cosplay*, have high enough frequencies to be in the top twenty. The frequencies of all LS pairs that have relevant tokens is presented in Appendix 4.

	borrowing	**token**	**base**
1	tsunami	6,318	3,659
2	anime	4,721	2,414
3	martial art	4,342	2,453
4	yen	3,680	1,561
5	sushi	3,376	1,650
6	manga	3,250	1,853
7	acupuncture	2,615	1,085
8	tofu	2,115	1,153
9	Zen	2,059	1,001
10	quantitative easing	1,991	1,393
11	karaoke	1,955	1,508
12	tycoon (s2)	1,861	1,423
13	ramen	1,410	728
14	ninja (n. s1)	1,244	724
15	karate	1,181	787
16	samurai (s1a)	1,158	672
17	judo	1,124	627
18	soy (s2)	1,094	663
19	cosplay (s1)	949	387
20	reiki	773	285

Table 19. Twenty most frequent Japanese borrowings overall. The column *token* contains the number of all relevant tokens and *base* the number of base occurrences. The part of speech or the number of the sense (e.g. s1, s2) in the *OED* are presented in parentheses in the case of LS pairs that share the same lexeme part.

The frequencies of LS pairs in all varieties studied can be seen in Table 20. The variety with the highest normalized frequency (henceforth abbreviated to *NF*) is SgE, where it is 271. This is significantly higher than in the other Outer Circle varieties, which nevertheless have relatively high NFs close to one another: PhE has 150, MyE 149, and HKE 142. The Inner Circle varieties have lower NFs compared to those of the Outer Circle varieties with AmE having 65 and BrE 48. Base occurrences make up 56% of all relevant tokens in the texts overall and the percentage varies between

47–62% in the different varieties. A summary of the varieties in the order of NFs in each main category is presented in Appendix 5. It is apparent from the summary that the tokens in SgE are not concentrated in a certain semantic field, as the variety has the highest NF in 15 categories. HKE is the first in five categories, PhE in four and MyE in one. The Inner Circle varieties are in the last place in most cases. A more detailed breakdown of each category is presented in the following sections.

		RF	NF
AmE	token	25,013	64.665
	base	14,746	38.122
BrE	token	18,672	48.172
	base	11,650	30.056
SgE	token	11,661	271.346
	base	5,775	134.381
PhE	token	6,467	149.526
	base	3,065	70.867
MyE	token	6,303	148.585
	base	3,249	76.591
HKE	token	5,757	142.323
	base	3,011	74.437
All	token	73,881	78.304
	base	41,307	43.780

Table 20. Frequencies of Japanese borrowings in each of the varieties studied. *RF* stands for *raw frequency*.

5.2 Distribution of Japanese borrowings into semantic categories in Present-day English texts

5.2.1 Architecture and interior design

ARCHITECTURE AND INTERIOR DESIGN has 1,226 relevant tokens and a NF of 1.3, which is distributed in the subcategories so that approximately 65% of the tokens belong to BUILDINGS AND PROPERTY, 20% to INTERIOR DESIGN AND FURNITURE and 16% to ARCHITECTURE, as seen in Appendix 6. The main category has the NF of 0.5–11.7 in the varieties studied. A closer inspection of the frequencies shows that SgE has the highest raw frequency (henceforth abbreviated to *RF*) of all relevant tokens, which is 503. This is unexpected, as the SgE corpus is almost ten times smaller than the AmE or BrE

corpora. However, the high frequency is mostly because of two LS pairs, *onsen* and *ryokan*, which together account for approximately 74% of all relevant tokens in SgE. Most of these tokens are repeated by the same author in the same text component, which is why the base occurrence in the main category is significantly lower, only 163 instances. Nevertheless, the NF of the tokens of ARCHITECTURE AND INTERIOR DESIGN is highest in SgE at 11.7. The subcategories follow the order BUILDINGS AND PROPERTY, ARCHITECTURE, and INTERIOR DESIGN AND FURNITURE from most to least frequent. The NF of the first is especially high thanks to the aforementioned *onsen* and *ryokan*.

Both HKE and MyE have NFs close to three, the number being 2.9 for HKE and 3.0 for MyE. The order of the subcategories from most to least frequent is the same as with SgE, that is, BUILDINGS AND PROPERTY, ARCHITECTURE, and INTERIOR DESIGN AND FURNITURE for both varieties. BUILDINGS AND PROPERTY is especially interesting as MyE has a higher NF of all relevant tokens, but the situation is the opposite with the NF of the base occurrences, which is nearly twice as high for HKE, as it is for MyE. This means that MyE has a large gap between the relevant tokens and base occurrences. The reason for this is mostly the borrowing *onsen* having the RF of 91 relevant tokens, of which only 17 are base occurrences. HKE has the same number of the base occurrences for *onsen*, but it is also matched by *izakaya*. The latter LS pair has only four base occurrences in MyE, which leads to the variety having a lower number of base occurrences than HKE in this subcategory and subsequently in the main category as well.

AmE, BrE, and PhE all have a NF of less than one, and the lowest NF of 0.5 belongs to BrE. The distribution between the subcategories in AmE is such that INTERIOR DESIGN AND FURNITURE has the most tokens followed by BUILDINGS AND PROPERTY, and finally ARCHITECTURE, whereas for PhE the first two are in reverse. The latter variety has a significantly lower number of tokens than the other Outer Circle varieties, which all have at least three times as many tokens, and the difference between the subcategories is within five tokens of each other. BUILDINGS AND PROPERTY is the most frequent for BrE with the other two subcategories having the same number of tokens. However, over half of

the 81 tokens in BUILDINGS AND PROPERTY in BrE are from the same source and the number of base occurrences is 40, the same as it is in INTERIOR DESIGN AND FURNITURE.

The number of LS pairs which have relevant tokens in at least one variety is 25 out of the 35 pairs from the *OED* in the main category. Six of the 25 are found in all the varieties: *futon* '[…] any thin mattress or low-lying bed deriving from or resembling the Japanese original' (*OED* s.v. *futon n.*), *izakaya*, *onsen*, *ryokan*, *torii*, and the attributive *tatami* (henceforth *tatami (s2)*). The most frequently used LS pairs overall are *ryokan, onsen*, and *futon* in the aforementioned sense (henceforth *futon (draft)*), but as can be seen in Table 21, there is some variance between the varieties and none of them have the exact same ordering of the top three most frequent borrowings. The LS pair *futon (draft)* is the most frequent pair in AmE, BrE, and PhE. MyE and HKE have *onsen* as the most common LS pair, AmE and SgE as the second and BrE as the third while only PhE did not have it in the top three. SgE is the only variety in which *ryokan* is the most frequent LS pair. It is in second place in BrE and the third in PhE and HKE. Indeed, 80% of all the tokens of *ryokan* are from SgE texts. A similar pattern can be seen with *futon (draft)* where 59% of the tokens are from AmE, but the tokens are slightly more evenly spread between varieties in the case of *onsen*. Other LS pairs which are in the top three in more than one variety include *tatami (s2)*, which is found in AmE and MyE, and *izakaya*, which occurs in all of the Outer Circle varieties.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	ryokan	315	futon (draft)	96	futon (draft)	41
2	onsen	314	onsen	28	ryokan	24
3	futon (draft)	164	tatami (s2)	23	onsen	23

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	ryokan	250	futon (draft)	12	onsen	91	onsen	49
2	onsen	120	izakaya	8	tatami (s2)	14	izakaya	23
3	izakaya	36	ryokan	5	izakaya	7	ryokan	12

Table 21. Frequencies of the three most frequent Japanese borrowings in ARCHITECTURE AND INTERIOR DESIGN in each of the varieties studied.

5.2.2 Arts and crafts

ARTS AND CRAFTS has 2,087 relevant tokens and a NF of 2.2. The latter is relatively low considering that the main category has the second highest number of different LS pairs among the categories. The tokens are divided into all of the subcategories, but there is major variance in the frequencies: the subcategory with the lowest RF is GEMSTONES and the lowest number of base occurrences is in OTHER CRAFTS, whereas the highest RF is recorded in PAPER FOLDING AND PACKAGING, which has 571 relevant tokens. The latter is interesting in that it is the smallest subcategory within the main category in terms of the number of different LS pairs and only two of them, *noshi* and *origami*, have relevant tokens. Additionally, *noshi* is used only once in a MyE text, so all the other tokens are for *origami*. Besides PAPER FOLDING AND PACKAGING, the subcategories which have more than 100 tokens are LITERATURE with 551 tokens, PAINTING, PICTURES, AND PRINTING with 418, THEATER, PUPPETRY, AND DANCE with 262, MUSIC with 159, and POTTERY AND LACQUERWARE with 124. ARRANGEMENT, OTHER CRAFTS, and GEMSTONES all have fewer than 100 tokens, the numbers being 63, 17, and 14 for the subcategories respectively.

The variety with the highest NF is PhE, which has the NF of 6.4, followed by SgE with 4.4. Despite the difference in the number of all relevant tokens, the two varieties are very close to one another in terms of NF of base occurrences, as it is 2.9 for both. This is largely because the base occurrences are notably lower than the number of all relevant tokens in LITERATURE, MUSIC, and THEATER, PUPPETRY, AND DANCE in PhE. For example, MUSIC has 41 tokens, but the number of base occurrences is only nine. In comparison, POTTERY AND LACQUERWARE is the only subcategory where the base occurrences are less than 50% of all relevant tokens in SgE. The three subcategories with the highest RFs in PhE are LITERATURE, THEATER, PUPPETRY, AND DANCE, and PAINTING, PICTURES, AND PRINTING, and the variety has no tokens of LS pairs from OTHER CRAFTS and GEMSTONES. The order

of the three subcategories with the highest frequencies is PAINTING, PICTURES, AND PRINTING, PAPER FOLDING AND PACKAGING and LITERATURE for SgE.

MyE and HKE have the next highest NFs of 3.1 and 2.7. The two varieties are close in the NF of the base occurrences with 1.5 for both. Although the frequencies are similar overall, the varieties differ clearly in all the subcategories except in POTTERY AND LACQUERWARE and GEMSTONES. HKE also lacks tokens of LS pairs in ARRANGEMENT and OTHER CRAFTS. AmE and BrE have NFs that are below that of all the texts. The NF is slightly higher for AmE at 2.0, and lowest for BrE at 1.6. Interestingly, BrE is the variety with the lowest NF only in PAINTING, PICTURES, AND PRINTING, which means that the variety simply has a generally low NF across the subcategories.

The number of different LS pairs with relevant tokens is 59 out of the 104 found in the *OED*. Only seven of these are used in texts of all varieties: *haiku*, *Kabuki*, *Nikkei* '[…] the Nikkei Keizai Shimbun newspaper' (*OED* s.v. *Nikkei n.2*) (henceforth *Nikkei (n.2 s3)*), *origami*, *sensei* '[…] (a respectful title or form of address for) a teacher, master, or skilled artist. Also in extended use: a guide or mentor' (*OED* s.v. *sensei n.2*) (henceforth *sensei (s1)*), *Suzuki*, and *ukiyo-e*. All of the seven except *ukiyo-e* appear in the top three most frequent LS pairs of at least one variety. As noted earlier, *origami* has a RF, 570 tokens, and is overall the most common LS pair in ARTS AND CRAFTS. In second place is *haiku* with 373 tokens and in the third *sensei (s1)* with 291. AmE, BrE, and MyE have *origami* as the most frequent LS pair while the pair is in second place for SgE and third for HKE. In the two latter varieties, the difference from the previous placement is only one token. PhE is the only variety where *origami* is not in the top three, and there is a noticeable gap in the frequencies of *origami* and *sensei (s1)*, which is in third place. The most frequent LS pair in PhE and HKE is *haiku*, followed by *Kabuki*. The borrowing *haiku* is also found in second place for AmE and BrE and *Kabuki* in third place for SgE. An interesting note about *Kabuki* is that it is also used in senses that are not recorded in the *OED*: in the area of politics, the lexeme is used in a meaning similar to *show* 'a feigned performance of an action; a pretence, an act' (*OED* s.v. *show n.1*). An example of this is "Congress

is slowly, but surely and methodically being exposed as a giant, corrupt Kabuki". There is also one token in "[…] just touch my kabuki on the powder itself […]", where the borrowing is used to mean 'kabuki brush', a type of brush used in make-up. *Nikkei (s3)* and *Suzuki* are each found in the top three of one variety each, the former in BrE and the latter in MyE.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	origami	570	origami	213	origami	239
2	haiku	373	haiku	197	haiku	90
3	sensei (s1)	282	sensei (s1)	155	Nikkei (n.2 s3)	43

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	sensei (s1)	44	haiku	44	origami	37	haiku	21
2	origami	43	Kabuki	42	sensei (s1)	23	Kabuki	18
3	Kabuki	21	sensei (s1)	32	Suzuki	19	origami	17

Table 22. Frequencies of the three most frequent Japanese borrowings in ARTS AND CRAFTS in each of the varieties studied.

5.2.3 Attributes and qualities

ATTRIBUTES AND QUALITIES has 1,159 relevant tokens and the NF of 1.2. The NFs are highest in PhE and SgE, the numbers being 3.2 and 2.5 respectively. MyE is slightly above, AmE and BrE below the texts overall in NF. HKE has the least tokens, which results in the NF of 0.4. One unusual point about the last-mentioned variety is that there are noticeably fewer relevant tokens of *ninja* '[o]f, relating to, or resembling a ninja […]' (*OED* s.v. *ninja n. and adj.*) (henceforth *ninja (adj.)*) than in the other Outer Circle varieties, as HKE has five and the others 28–103. The LS pair *kawaii* '[c]ute, esp. in a manner considered characteristic of Japanese popular culture; charming, darling; ostentatiously adorable' (*OED* s.v. *kawaii n. and adj.*) (henceforth *kawaii (adj.)*) also has somewhat fewer tokens in the variety.

Eight out of all the nine LS pairs in ATTRIBUTES AND QUALITIES have relevant tokens. The only LS pair that did not is *banzai* '(as if) shouting 'banzai', uproarious, jollificatory […]' (*OED* s.v. *banzai int.*). There are four which are used in all the varieties *kamikaze*: '[…] [r]eckless, dangerous; hazardous and potentially self-destructive' (*OED* s.v. *kamikaze n. and adj.*) (henceforth *kamikaze (adj.)*), *kawaii (adj.)*, *ninja (adj.)*, and *ninja* '[…] a person or thing considered to behave or look like a ninja, to excel at a particular skill, achieve a difficult task, etc.' (*OED* s.v. *ninja n. and adj.*) (henceforth *ninja (n. s2)*). Both of the LS pairs with *ninja* as the lexeme part are the most frequent borrowings overall, although *ninja (adj.)* is used significantly more often than *ninja (n. s2)*, and the LS pair *kamikaze (adj.)* is in third place. The Inner Circle varieties followed the same order, but the Outer Circle varieties are not equally uniform: while the top three are identical for SgE and PhE with *kawaii (adj.)* having the most tokens, followed by *ninja (adj.)* and *ninja (n. s2)*, MyE and HKE did not share the same order with any other variety. They also include LS pairs such as *Godzilla* in MyE and *gaijin* '[o]f or relating to a gaijin; foreign, alien' in HKE that are not in any other top three.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	ninja (adj.)	664	ninja (adj.)	283	ninja (adj.)	202
2	ninja (n. s2)	193	ninja (n. s2)	91	ninja (n. s2)	76
3	kamikaze (adj.)	130	kamikaze (adj.)	37	kamikaze (adj.)	59

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	kawaii (adj.)	46	kawaii (adj.)	103	ninja (adj.)	28	kamikaze (adj.)	7
2	ninja (adj.)	32	ninja (adj.)	23	kawaii (adj.)	20	ninja (adj.)	5
3	ninja (n. s2)	8	ninja (n. s2)	10	ninja (n. s2), kamikaze (adj.), Godzilla	7	gaijin (adj.), kawaii (adj.)	2

Table 23. Frequencies of the three most frequent Japanese borrowings in ATTRIBUTES AND QUALITIES in each of the varieties studied.

5.2.4 Business and finance

The number of relevant tokens is 4,262 and the NF is 4.5 for BUSINESS AND FINANCE. The tokens are distributed between the two subcategories so that 48% of the tokens are in BUSINESS AND WORKING CULTURE and 52% in FINANCE. The NFs per variety show that, despite the overall distribution between the subcategories being quite even, there is interesting variance: the portion of FINANCE is larger in the Inner Circle varieties than in the texts overall, as the percentage is approximately 58–60%. Conversely, BUSINESS AND WORKING CULTURE is the subcategory with the most tokens in the Outer Circle varieties, where 66–68% of all relevant tokens belong to the category. The exception is SgE, where the percentage is not equally high but still reaches slightly over the entire corpus with 51%. The difference between the varieties is caused by *quantitative easing* from FINANCE being the most frequent borrowing for the Inner Circle varieties but only the second for the Outer Circle ones, for which instead *tycoon (s2)* from BUSINESS AND WORKING CULTURE is the most frequent item.

The main category NFs in the varieties studied are in the range of 2.7–12.3. The highest NF is that for HKE, followed by MyE, SgE, and BrE, where the numbers are 7.7 for MyE, 6.0 for SgE and 5.0 for BrE. PhE is only slightly below the total corpus at 4.2. The NF for AmE is 2.7. It is rare for BrE to have a noticeably higher NF than AmE. There is no single LS pair that is the reason for this and instead AmE simply has overall fewer tokens than BrE. The LS pairs found in both of the varieties have a lower number of tokens for AmE than BrE in all cases, and there is only one borrowing that is not used in any other variety except AmE but it has only one instance. It thus seems that Japanese borrowings of the BUSINESS AND FINANCE category are generally used more in BrE than AmE.

There are 12 LS pairs out of the total 25 which have relevant tokens. Seven of these are found in all varieties: *kaizen*, *keiretsu*, *Nikkei* '[d]esignating an index of the relative price of representative shares on the Tokyo Stock Exchange […]' (*OED* s.v. *Nikkei n.2*) (henceforth *Nikkei (n.2 s1)*), *Nikkei (n.2 s3), quality circle*, *quantitative easing*, and *tycoon (s2)*. The order of the three most frequent LS

pairs is *quantitative easing* as the first, *tycoon (s2)* as the second, and *Nikkei (n.2 s1)* as the third. The two first-mentioned have almost 2,000 tokens and together account for 90% of the main category. *Nikkei (n.2 s1)* has significantly fewer tokens. The borrowing *kaizen* is close to the third place with 120 relevant tokens but is used comparatively less than *Nikkei (n.2 s1)* in especially HKE texts. The order of the top three items is identical between the Inner Circle varieties as well as between those of the Outer Circle: for AmE and BrE, the order is *quantitative easing*, *tycoon (s2)*, and kaizen, whereas for SgE, PhE, MyE, and HKE it is *tycoon (s2)*, *quantitative easing*, and *Nikkei (n.2 s1)*. As one may expect from the overall statistics, the LS pair in third place has remarkably fewer tokens than the pairs in the first or second place in all varieties.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	quantitative easing	1,991	quantitative easing	577	quantitative easing	767
2	tycoon (s2)	1,861	tycoon (s2)	370	tycoon (s2)	603
3	Nikkei (n.2 s1)	129	kaizen	37	kaizen	40

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	tycoon (s2)	117	tycoon (s2)	104	tycoon (s2)	207	tycoon (s2)	334
2	quantitative easing	98	quantitative easing	52	quantitative easing	78	quantitative easing	147
3	Nikkei (n.2 s1)	15	Nikkei (n.2 s1)	7	Nikkei (n.2 s1)	23	Nikkei (n.2 s1)	6

Table 24. Frequencies of the three most frequent Japanese borrowings in BUSINESS AND FINANCE in each of the varieties studied.

5.2.5 Currency and units of measurement

CURRENCY AND UNITS OF MEASUREMENT has 3,820 relevant tokens and the NF of 4.0. The distribution of the tokens is very uneven in the subcategories: tokens of the LS pairs of the CURRENCY subcategory

amount to 96% of all tokens. Virtually all of the tokens in the subcategory are instances of *yen*. The only one that is not is a single use of *koku*. Tokens of the LS pairs in UNITS OF MEASUREMENT ARE significantly rarer, as they account for 0.3–0.5% of all relevant tokens in a variety. The NFs of the varieties have a wide range of 2.0–22.7. The variety which has by far the highest NF is SgE, there being a gap of 12.7 between it and the variety with the next highest NF. However, as many as 60% of the relevant tokens in SgE texts are by the same author and from the same text component, so the NF of base occurrences, while still the highest, is not as removed from those of the other varieties as the NF of all the relevant tokens. SgE also has the highest RF of UNITS OF MEASUREMENT and is the only variety in which the subcategory has a NF reaching 1.0.

The frequencies for MyE and HKE are also relatively high as the NF for the former is 10.0 and the latter 8.7. Both varieties have the NF of 0.4 for UNITS OF MEASUREMENT. The last Outer Circle variety, PhE, has the NF of 5.9 for the main category, meaning that all of these varieties exceed the occurrence rate of five in a million words. PhE only has one instance of a LS pair, *tatami (s2)*, belonging to the UNITS OF MEASUREMENT subcategory, which is why the variety has the lowest NF for the subcategory. Both of the Inner Circle varieties have NFs that are below that of the entire corpus. BrE has the highest RF overall, but the NF is 2.7. The number is lowest for AmE, where the NF is approximately 2.0.

The number of LS pairs with relevant tokens is seven out of the 22 from the *OED*. These are *koku*, *mat*, *ri*, *tatami* in both senses, *tsubo*, and *yen*. The relatively low number of different LS pairs can be expected, as Japan has had yen as the official currency since 1871 and most of the units of measurement are used only in very specific fields, for example *tan* '[a] Japanese unit used for measuring cloth, equivalent to about ten yards in length and just over one foot in width […]'. The only LS pairs that are found in all varieties are *yen* and *tatami (s2)*. As can be seen from the figure below, the top three most frequent borrowings are quite uniform across the varieties: all have *tatami (s1)*, *tatami (s2)*, and *yen* in the top three except PhE, which lacks the first mentioned, and HKE,

which has *tsubo* instead of *tatami (s2)*. The order is *yen*, *tatami (s2)*, and *tatami (s1)* or *tsubo* for all except PhE and BrE, which has the last two in the reverse order. SgE has the most variance in LS pairs and only *koku* of the seven aforementioned is not found in texts of this variety. PhE has the least diversity, as evidenced by there being only two LS pairs with relevant tokens.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	yen	3680	yen	741	yen	415
2	tatami (s2)	90	tatami (s2)	18	tatami (s1)	21
3	tatami (s1)	38	tatami (s1)	7	tatami (s2)	11

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	yen	934	yen	253	yen	408	yen	335
2	tatami (s2)	31	tatami (s2)	1	tatami (s2)	14	tatami (s2)	10
3	tatami (s1)	8	-	-	tatami (s1)	1	tsubo	4

Table 25. Frequencies of the three most frequent Japanese borrowings in CURRENCY AND UNITS OF MEASUREMENT in each of the varieties studied.

5.2.6 Cooking and beverages

COOKING AND BEVERAGES has the most tokens in all of the main categories, amounting to 14,523 instances and subsequently the highest NF of 15.4. FOOD AND COOKING comprises 92% of the tokens in the subcategories, but BEVERAGES is still decently sized at 1,194 tokens and a NF of 1.2. There is great variance in the NFs between the varieties: SgE has by far the highest frequency, 89.4, which is also highest across all main categories. The variety has the greatest number of different LS pairs with relevant tokens, 88. Ten of these LS pairs have more than a hundred tokens in SgE while the number of LS pairs with over a hundred tokens is three to five for the other varieties. The high number of pairs exceeding a hundred tokens in SgE also shows that the tokens are spread between the LS pairs rather than being concentrated in a single pair. The other Outer Circle varieties followed SgE and all

of them are significantly above the NF for the total corpus. HKE is second highest with the NF of 47.7, MyE the third at 33.0, and PhE the fourth at 24.3. The NFs are the peak out of all the main categories for HKE and MyE. The Inner Circle varieties are both below the total corpus in NF, but the numbers are still high in comparison to most of the other main categories, as the NFs are 10.6 for AmE and 5.7 for BrE. The order of the varieties is the same as in the main category for FOOD AND COOKING but slightly different for BEVERAGES, where HKE has the highest NF and SgE is second. This is explained by HKE having the highest number of instances of *saké* among the varieties.

There are 90 out of 97 borrowings that have relevant tokens, which is the highest raw number in all of the main categories, and 42 of them are used in all varieties. The order of the three most frequent LS pairs is rather uniform: the most common LS pair in all varieties is *sushi* and the second is *tofu* except in PhE, where *ramen* is second instead. The last-mentioned pair is the third for AmE, BrE, and SgE while *tofu* is the third for PhE, *sashimi* for MyE and *saké* for HKE. In PhE, *ramen* has more tokens than *tofu,* but the borrowings would be in reverse if the ordering were by base occurrences. An interesting detail about HKE is that *ramen* is not in the top three of the variety and has significantly fewer tokens than in the other varieties. MyE also does not have *ramen* in the top three, but the LS pair is in fourth place and only has one token fewer than *sashimi*. It seems that *ramen* is simply not used as much in HKE texts as it is in the other varieties.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	sushi	3,376	sushi	563	sushi	425
2	tofu	2,115	tofu	402	tofu	188
3	ramen	1,410	ramen	371	ramen	72

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	sushi	761	sushi	233	sushi	298	sushi	348
2	tofu	477	ramen	132	tofu	222	tofu	263
3	ramen	355	tofu	112	sashimi	115	saké	196

Table 26. Frequencies of the three most frequent Japanese borrowings in COOKING AND BEVERAGES in each of the varieties studied.

5.2.7 Education

EDUCATION has 343 tokens and the NF is 0.4. There is variance in the NFs across the varieties, as SgE has 1.1 and MyE is only marginally below 1.0. The NFs are also below 1.0 for PhE and AmE, where the numbers are 0.8 and 0.5 respectively, but they are still above the total corpus. HKE and BrE have low NFs, both being below the total corpus and having frequencies of below 0.1. Both varieties have clearly fewer tokens of the LS pair *sensei* '[…] (a respectful title or form of address for) a teacher, master, or skilled artist. Also in extended use: a guide or mentor' (*OED* s.v. *sensei n.*) (henceforth *sensei (s1)*) than the other corpora of similar size. For example, BrE, which has overall the lowest NF of 0.09, has 25 relevant tokens of *sensei (s1)* while the number for AmE is over six times higher. Another point to note is that, aside from a single case in BrE, the borrowing is not used as an honorific, as in *Yamada-sensei*, in HKE and BrE while the other varieties have several such cases. If the base occurrences are examined, it may be remarked that in AmE the number is especially low in comparison to all relevant tokens, only comprising 36% of the tokens. This is because only 49 tokens of the 155 for *sensei (s1)* are base occurrences.

Three out of the six LS pairs in EDUCATION have relevant tokens, which explains why the three most frequent borrowings are so similar between the varieties. The three LS pairs are *juku* '[…] (originally) a private tutoring school run by a single tutor (hist.); (now) a private school or college attended in addition to an ordinary educational institution; spec. an institution that prepares pupils for an examination intensively over a short period of time' (*OED* s.v. *juku n.*) (henceforth *juku (s2)*), *sensei (s1)*, and *Suzuki*, of which *sensei* is the most frequent pair in all varieties and *Suzuki* is generally in second place. The exception to this is BrE, where the latter LS pair is third and *juku (s2)* is second instead. AmE, PhE, MyE, and HKE did not have any relevant tokens of the latter.

	All		AmE		BrE	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	sensei (s1)	282	sensei (s1)	155	sensei (s1)	25
2	Suzuki	54	Suzuki	27	juku (s2)	5
3	juku (s2)	7	-	-	Suzuki	3

	SgE		PhE		MyE		HKE	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	sensei (s1)	44	sensei (s1)	32	sensei (s1)	23	sensei (s1)	3
2	Suzuki	3	Suzuki	1	Suzuki	19	Suzuki	1
3	juku (s2)	2	-	-	-	-	-	-

Table 27. Frequencies of the three most frequent Japanese borrowings in EDUCATION in each of the varieties studied.

5.2.8 Electronics and technology

ELECTRONICS AND TECHNOLOGY has 462 tokens and a NF of slightly under 0.5. Interestingly, BrE, which has the NF of 0.3, is the only variety whose frequency is less than that for the total corpus. Compared to AmE, whose corpus is of a similar size, the LS pairs *otaku* and *emoji* have a lower number of tokens in BrE, which explains the low frequency. SgE has the highest NF, 1.6, followed by PhE with 0.9. The remaining varieties, HKE, MyE, and AmE, all have NFs in the range of 0.5–0.6.

There are a few main categories where all of the LS pairs from the *OED* are used in the corpus texts at least once. ELECTRONICS AND TECHNOLOGY is one of these categories, as tokens of all six pairs are found. The LS pair *otaku* has a distinctly high frequency overall while *uncanny valley* and *emoji*, which are in second and third place, only have a difference of one token. Noteworthy is that, while *otaku* and *emoji* are in the top three of each variety, the LS pair *uncanny valley* is exclusively used in the Inner Circle varieties, where its frequency is relatively high. In addition to *otaku* and *emoji*, *mechatronics* is also one of the three LS pairs that is used in all of the varieties. The pair is the most

frequent in MyE and the third for SgE and PhE. *Yagi* is the most frequent in HKE, where the LS pair has 12 tokens. This is the largest frequency the pair has across the varieties, as the others have six tokens at most.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	otaku	243	otaku	112	uncanny valley	32
2	uncanny valley	74	uncanny valley	42	otaku	31
3	emoji	73	emoji	36	emoji	19

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	otaku	52	otaku	35	mechatronics	13	Yagi	12
2	emoji	10	emoji, Yagi	2	otaku	9	emoji	5
3	mechatronics	6	keitai, mechatronics	1	emoji	1	otaku	4

Table 28. Frequencies of the three most frequent Japanese borrowings in ELECTRONICS AND TECHNOLOGY in each of the varieties studied.

5.2.9 Entertainment and leisure activities

ENTERTAINMENT AND LEISURE ACTIVITIES has the second highest number of tokens among all the main categories. The RF of all relevant tokens is 13,203 and the NF 14.0 for the entire corpus. The subcategory with the greatest frequency is COMICS AND CARTOONS, which accounts for well over half of the tokens and has the NF of 10.5. LEISURE ACTIVITIES is second, including roughly a fourth of all the tokens in the subcategories, followed by GAMES, which constitutes a tenth. The NFs for the two subcategories are 4.3 and 1.7 respectively. TRAVELLING AND TOURISM has the lowest number of tokens, which amounts to 4% of those in the subcategories and a NF of 0.7. The frequencies are not especially affected by the number of LS pairs in the subcategories, as they have roughly the same

number of pairs. The exception is TRAVELLING AND TOURISM, which has fewer LS pairs than the other categories and also the lowest frequency of tokens.

The NFs of the varieties are for the most part very high. The highest is 65.1 for SgE. The variety has a generally high frequency of tokens across LS pairs. Additionally, *ryokan* and *onsen*, the two LS pairs that cause SgE to have the highest NF in ARCHITECTURE AND INTERIOR DESIGN, are also part of this main category. PhE has the second and MyE the third highest NFs, 48.0 and 29.3 respectively. AmE and HKE are slightly below the entire corpus in NF, as the number is 12.1 for the former and 11.7 for the latter. The frequencies for PhE and AmE are the highest out of all the main categories in these varieties. BrE has the lowest NF of 5.0. The order of the subcategories from the most to least tokens is for most varieties the same as it is for the total corpus. While the order remains the same, the number of tokens in COMICS AND CARTOONS and LEISURE ACTIVITIES are quite similar in BrE, as the RF for the former category is 1,004 and the latter 961. HKE is the only variety where the order differs, and LEISURE ACTIVITIES has the highest frequency, followed by COMICS AND CARTOONS. This is because *anime* and *manga* have a noticeably smaller number of tokens in comparison to the other Outer Circle varieties.

The number of LS pairs with relevant tokens is 21 out of 24. The only pairs that did not have tokens are *gobang*, *furo*, and *ken* '[a] Japanese game of forfeits played with the hands and with gestures'. The three most frequent borrowings in ENTERTAINMENT AND LEISURE ACTIVITIES are also part of the twenty most frequent borrowings in the entire material studied, where *anime* is the second most frequent, *manga* sixth and *karaoke* eleventh. These LS pairs are also in the top three in most of the varieties. AmE and MyE share the same order as the total corpus, whereas BrE has all the LS pairs but their order from most to least frequent is *karaoke*, *anime*, and *manga*. The LS pair *karaoke* is also the most frequent in HKE, where *cosplay* and *manga* are in second place and *anime* in third. Similarly, SgE has *cosplay* in third place and is the only variety where *manga* is first, although *anime* is not too far behind in tokens. PhE has *anime* and *manga* as first and second most frequent LS pair

but, differing from the other varieties, *bonsai* is in third place in the variety. A notable number of the instances of *bonsai* were, however, from the same writer in the same text component and the number of base occurrences is merely 17. If the LS pairs were ordered by base occurrences, *karaoke* would be in third place in PhE.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	anime	4,721	anime	1,997	karaoke	714
2	manga	3,250	manga	1,472	anime	580
3	karaoke	1,955	karaoke	443	manga	268

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	manga	774	anime	812	anime	533	karaoke	172
2	anime	743	manga	439	manga	230	cosplay, manga	67
3	cosplay	385	bonsai	292	karaoke	212	anime	56

Table 29. Frequencies of the three most frequent Japanese borrowings in ENTERTAINMENT AND LEISURE ACTIVITIES in each of the varieties studied.

5.2.10 Fashion

FASHION has 870 tokens and the NF of 0.9. Almost 95% of the tokens are in the CLOTHES subcategory and the remaining in ACCESSORIES and FOOTWEAR. The subcategory HAIRSTYLES has no relevant tokens in any of the varieties studied, which is not particularly surprising, as there are only two LS pairs in the category and both of these refer to hairstyles that are used on very limited occasions today. The NFs of varieties range between 0.5 and4.3, with the highest frequency found in SgE. All tokens in the variety are in the CLOTHES and FOOTWEAR subcategories. SgE texts have an exceptionally high frequency of *yukata*, 46 tokens, compared to the 3–9 instances in the other varieties, which explains why the NF is so high. HKE has the next highest NF of 2.5, followed by MyE and PhE, where the

numbers are respectively 1.3 and 1.0. Both of the Inner Circle varieties have lower NFs than the average, 0.7 for BrE and 0.5 for AmE.

There are 15 out of the 22 LS pairs from the *OED* that have relevant tokens. Four of them are used in all varieties: *belt*, *kimono*, *obi*, and *yukata*. The most frequent of these is *kimono*, which is the LS pair with the most tokens in this main category for all varieties. The LS pair is used in a different meaning from that listed in the *OED* in the phrase *(to go) open kimono* or *to open (up) the kimono* for example in "By getting involved at an early stage and opening its kimono to credit and IBM technology, partners are more likely to stay with the company". The second most common LS pair in all varieties combined is *belt* '[i]n judo, karate, and certain other martial arts: a belt indicating by its colour the level of proficiency attained by the wearer' (*OED* s.v. *belt n.*), which is in second place in AmE and BrE and third place in SgE. In third place overall is *yukata*. AmE, BrE, and PhE have the identical top three of *kimono*, *belt*, and *yukata* while SgE and MyE have almost the same order except for the latter also having *obi* in second place together with *yukata*. HKE is the only one with *yokozuna* in the top three. SgE is the only variety where the difference between *kimono* and the second-place borrowing is not particularly large. This may be explained by the fact that the borrowing in second place is *yukata*, which refers to a type of kimono, so writers of SgE texts may have opted to use the more specific LS pair in some cases instead of *kimono*.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	kimono	414	kimono	98	kimono	157
2	belt	167	belt	60	belt	54
3	yukata	78	black belt	21	black belt	26

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	kimono	59	kimono	21	kimono	30	kimono	49
2	yukata	46	belt	7	obi, yukata	7	yokozuna	15
3	belt	38	black belt	6	belt	4	obi	11

Table 30. Frequencies of the three most frequent Japanese borrowings in FASHION in each of the varieties studied.

5.2.11 Government, administration, and politics

The number of relevant tokens is 2,480 for GOVERNMENT, ADMINISTRATION, AND POLITICS and the NF for the entire corpus is 2.6. The NFs of varieties are generally high, being over 1.5 in all of them. The highest NF is 9.4 for HKE. This is followed by MyE, where the number is 7.7. The NFs of 4.0 and 3.0, respectively for SgE and PhE, are slightly lower than for the aforementioned varieties but still above that of the entire corpus. The Inner Circle varieties both have NFs that are below that of all the texts and especially AmE has a relatively low NF of 1.5. The latter may be attributed to *tycoon (s2)* being used approximately half as many times in AmE texts as in BrE texts, which have a similar wordcount.

There are 14 out of the 31 LS pairs that have relevant tokens. Four are used in all the varieties: *daimio*, *shogun*, *thought control*, and *tycoon (s2)*. The LS pair *tycoon (s2)* has the most tokens in all the varieties, which why it makes up 75% of all relevant tokens in the main category. The LS pair in second place in the entire corpus is *thought crime*. The pair is interesting in that it is also in second place in the Inner Circle varieties but is used at most twice in the Outer Circle ones. This may be because the term is popularized by George Orwell's *Nineteen Eighty-Four* (*OED* s.v. *thought n.*), which may be more familiar to American and British audiences. Instead of *thought crime*, the borrowing on the second place for SgE and PhE is *shogun*, which is on the third place for the remaining varieties except AmE. Other LS pairs found among the three most frequent borrowings are *thought control* in AmE, *daimio* in SgE, PhE, and HKE, and *money politics* in MyE. The last-mentioned is noteworthy because the LS pair is used 0–3 times in the other varieties, but MyE has 77 relevant tokens, of which as many as 50 are base occurrences. It seems that the term is simply more common in MyE than in the other varieties.

	All		AmE		BrE	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	tycoon (s2)	1861	tycoon (s2)	370	tycoon (s2)	729
2	thought crime	219	thought crime	114	thought crime	101
3	shogun	145	thought control	35	shogun	33

	SgE		PhE		MyE		HKE	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	tycoon (s2)	117	tycoon (s2)	104	tycoon (s2)	207	tycoon (s2)	334
2	shogun	44	shogun	15	money politics	77	daimio	25
3	daimio	6	daimio, money politics	3	shogun	6	shogun	13

Table 31. Frequencies of the three most frequent Japanese borrowings in GOVERNMENT, ADMINISTRATION, AND POLITICS in each of the varieties studied.

5.2.12 Greetings, interjections, and gestures

GREETINGS, INTERJECTIONS, AND GESTURES has 91 tokens, second least in all of the main categories. The NF for all varieties combined is 0.1. The ratio of base occurrences to all relevant tokens is the highest among the main categories as base occurrences amount to 91% of the tokens. The NFs are low for all varieties and range between 0.05 and 0.28, the highest score for MyE and SgE, the latter only marginally behind the former, and the lowest for BrE.

GREETINGS, INTERJECTIONS, AND GESTURES is one of the main categories where all of the borrowings in the category have tokens. There are thus ten different LS pairs which have relevant tokens, and most of them include the lexeme *sayonara* in different senses. In fact, the three most frequent borrowings all have *sayonara* as the lexeme part: the first has the sense '[u]sed to express finality with regard to a person's departure, suggesting either leaving for good or (by extension) death […]' (*OED* s.v. *sayonara int.*) (henceforth *sayonara (int. s2b)*), the second '[u]sed to suggest that something has been finished with, abandoned, or consigned to the past' (*OED* s.v. *sayonara int.*) (henceforth *sayonara (int. s2a)*), and the third 'used to express good wishes when parting […]' (*OED* s.v. *sayonara int.*) (henceforth *sayonara (int. s1)*). There are no LS pairs that have tokens in all of the

varieties. There is some variance as to which LS pairs are in the top three for each variety, but the frequencies are so low that there are no significant differences in the RFs between the placements of the top three items in each variety other than *sayonara (int. s2b)* having clearly more tokens than the other LS pairs in AmE. As the description in the *OED* notes the use of the LS pair is "Chiefly U.S.", so it is likely that this is the reason for why the LS pair has noticeably more tokens than the other ones in AmE but is not used nearly as commonly in the other varieties.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	sayonara (int. s2b)	33	sayonara (int. s2b)	21	sayonara (int. s2a)	7
2	sayonara (int. s2a)	25	sayonara (int. s2a)	6	sayonara (int. s2b)	4
3	sayonara (int. s1)	13	ah so	4	sayonara (int. s1)	3

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	sayonara (int. s1)	5	sayonara (int. s2a)	4	sayonara (int. s2b)	7	sayonara (adj. s1)	3
2	sayonara (int. s2a)	4	banzai (int. s1), sayonara (int. s1), sayonara (int. s2b)	1	sayonara (int. s2a)	4	sayonara (int. s1)	1
3	banzai (int. s1)	2	-	-	sayonara (adj. s2)	1	-	-

Table 32. Frequencies of the three most frequent Japanese borrowings in GREETINGS, INTERJECTIONS, AND GESTURES in each of the varieties studied.

5.2.13 Historical periods

There are 524 relevant tokens for HISTORICAL PERIODS, which results in the NF of 0.6. The highest NFs are in SgE and HKE, 2.2 and 2.1 respectively. The base frequency in HKE is markedly low in comparison to all the relevant tokens, with 76% of the tokens being used by the same writers in the

same text component. As a result, HKE has the third lowest NF of base occurrences. PhE has the third highest NF of 1.7 and MyE the fourth with the frequency of 0.7 reaching slightly over that of the entire corpus. The NFs of BrE and AmE are the lowest at 0.4 for the former and 0.2 for the latter. The number of tokens in AmE is the same as in HKE despite the former corpus being almost ten times larger. BrE has twice as many tokens as AmE, but the numbers of base occurrences are very similar.

All of the LS pairs in HISTORICAL PERIODS ARE used in the corpus texts. Three of the pairs are found in all varieties. These are *Meiji* '[o]f, relating to, or characteristic of this period' (*OED* s.v. *Meiji n. and adj.*) (henceforth *Meiji (adj.)*), *Tokugawa* '[o]f or pertaining to the Japanese ruling dynasty (1603–1867) founded by Tokugawa […]' (*OED* s.v. *Tokugawa n. and adj.*) (henceforth *Tokugawa (adj.)*), and *Heian*, which are also in the top three most frequent LS pairs in the main category in most of the varieties. *Meiji (adj.)* is the most frequent in all varieties and *Tokugawa (adj.)* or *Heian* are in second or third place in each. BrE has *Jomon* in second place and MyE *Showa* in third place, but otherwise the listings are rather similar across the varieties.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	Meiji (adj)	257	Meiji (adj)	42	Meiji (adj)	85
2	Tokugawa (adj.)	120	Tokugawa (adj.)	28	Jomon	31
3	Heian	47	Heian	5	Tokugawa (adj.)	30

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	Meiji (adj)	37	Meiji (adj)	37	Meiji (adj)	12	Meiji (adj)	44
2	Tokugawa (adj.)	17	Heian	15	Tokugawa (adj.)	8	Tokugawa (adj.)	26
3	Heian	16	Tokugawa (adj.)	11	Showa	3	Heian	6

Table 33. Frequencies of the three most frequent Japanese borrowings in HISTORICAL PERIODS in each of the varieties studied.

5.2.14 Manufacturing and materials

MANUFACTURING AND MATERIALS has 389 tokens and the NF of 0.4. There are two varieties which have NFs of over one: SgE 1.6 and MyE 1.2. However, most of the tokens in MyE are from the same writer and text component, so the frequency of the base occurrences is, in fact, lower than that of PhE, whose NF is 0.9. The three remaining varieties, AmE, BrE, and HKE, have NFs of less than 0.3.

Seven out of the 22 LS pairs have relevant tokens, and there are two, *sakura* and *washi*, which are used in all varieties. Perhaps surprisingly, neither of the two is the most frequent LS pair overall, although they are in the second and third place respectively. The LS pair with the greatest RF is *kanban* '[…] [t]he coordinated manufacturing system employing kanbans, which ensures that components arrive from suppliers at the time they are required for assembly […]' (*OED* s.v. *kanban n.*), which did not have any relevant tokens in SgE or MyE texts. In fact, the LS pair is not commonly used in the remaining Outer Circle varieties either as only five tokens are found in them as opposed to the 131 tokens in the Inner Circle varieties. The LS pair *sakura* is the most frequent pair in SgE, MyE, and HKE, the second most frequent in PhE and third in AmE. The LS pair that is in the top three of all varieties is *washi*, which is the most frequent in PhE, second most frequent in AmE, and SgE, MyE. Other LS pairs in the top three are *hinoki* and *kanban* '[…] a card or sheet displaying a set of manufacturing specifications and requirements which is circulated to suppliers and sent along a production line to regulate the supply of components' (*OED* s.v. *kanban n.*).

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	kanban (s2)	136	kanban (s2)	64	kanban (s2)	67
2	sakura	122	washi	26	kanban (s1)	21
3	washi	90	sakura, kanban (s1), hinoki	5	washi	17

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	sakura	51	washi	19	sakura	40	sakura	4
2	washi	14	sakura	13	washi	12	kanban (s2)	3

3	hinoki	3	hinoki	3	-	-	washi	2

Table 34. Frequencies of the three most frequent Japanese borrowings in MANUFACTURING AND MATERIALS in each of the varieties studied.

5.2.15 Military and bushido

MILITARY AND BUSHIDO has 1,857 tokens and a NF of 2.0. There is no one variety that is significantly above or below all others in NF. Instead, the varieties are divided into those with relatively high NFs and those that are slightly below the frequency of the combined texts. SgE, PhE, and HKE belong to the former, as they are in the range of 3.1–4.4. MyE, AmE, and BrE are below the NF of all the varieties combined, but not significantly so, and the frequencies are 1.5–1.8. The difference between base occurrences and all relevant tokens is not particularly notable in other varieties than HKE, where base occurrences comprised only 39% of relevant tokens. This is mostly because of *samurai* '[i]n Japan during the continuance of the feudal system, one of the class of military retainers of the daimios; sometimes in wider sense, a member of the military caste, whether a samurai proper or a daimio. Also applied to any Japanese army officer' (*OED* s.v. *samurai n.*) having 83 relevant tokens but 29 base occurrences and to a smaller extent *Kempeitai* having 25 tokens and nine base occurrences.

The number of LS pairs with relevant tokens is 15 out of the 21 from the *OED*. Three of the LS pairs are used in all of the varieties: *bushido*, *samurai (s1a)*, and *shogun*. The latter two LS pairs are also included in the three most frequent borrowings of the combined texts together with *ninjutsu*. The most frequent LS pair by a wide margin is *samurai (s1a)*, which accounts for approximately 62% of all relevant tokens in the main category. The pair also belongs to the twenty most frequent borrowings in the entire material studied, being in the sixteenth place. As one may expect, *samurai (s1a)* is the most frequent item in all of the varieties and has almost ten times more tokens than *shogun* in second place. The third, *ninjutsu*, has slightly fewer tokens than *shogun* and is in third place for AmE and

MyE and the second for BrE. Other LS pairs in the top three listings are *hara-kiri* in second place for MyE, *kamikaze* '[d]esignating a Japanese aircraft, pilot, or military unit whose purpose is to fly a suicide mission or missions […]; (also) of or relating to a mission of this kind[…]' (*OED* s.v. *kamikaze n. and adj.*) in second place of AmE and third place of BrE, *Kempeitai* in second place of HKE, and *ronin* in third place in SgE and PhE. The borrowing *kamikaze (adj. s1a)* is especially interesting, as it has a relatively high frequency in the Inner Circle varieties but is hardly used in the Outer Circle varieties likely because of the different ways in which the countries were involved in the Second World War.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	samurai (s1a)	1,158	samurai (s1a)	416	samurai (s1a)	382
2	shogun	145	kamikaze (adj. S1a)	40	ninjutsu	59
3	ninjutsu	121	ninjutsu	39	kamikaze (adj. s1a), shogun	33

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	samurai (s1a)	112	samurai (s1a)	117	samurai (s1a)	48	samurai (s1a)	83
2	shogun	44	shogun	15	hara-kiri	8	Kempeitai	25
3	ronin	13	ronin	12	ninjutsu	7	shogun	13

Table 35. Frequencies of the three most frequent Japanese borrowings in MILITARY AND BUSHIDO in each of the varieties studied.

5.2.16 Nature

NATURE has the third largest number of tokens, 11,479 instances, and the NF is 12.2. Approximately 58% of the tokens are instances of the LS pairs in ENVIRONMENT, 28% of the pairs in FLORA, 12% of those in FAUNA, and the remaining 2% in FUNGI. The NFs are relatively high for all of the subcategories except FUNGI, where the number is 0.2. Most of the varieties follow the same order of

subcategories as the total corpus. The exceptions are PhE, where FLORA has more tokens than ENVIRONMENT, and MyE, where FAUNA is above FLORA. Especially the former case is surprising, as ENVIRONMENT has the highest frequency in the other varieties due to *tsunami*, which is the most frequent borrowing in any of the main categories, as mentioned in Section 5.1. In the case of PhE, *tsunami* actually has the most tokens, but the second most frequent borrowing, *bonsai*, is not too far behind and there are other LS pairs from FLORA, such as *nori*, *wasabi*, and *soy*, that cause the subcategory to have a higher frequency than ENVIRONMENT, which is comprised almost exclusively of instances of *tsunami*.

The NFs of the varieties are relatively high. The two highest NFs are those for SgE and MyE, which have 32.9 and 31.1 respectively. PhE and HKE are moderately lower than the other two Outer Circle varieties, but the NFs are still relatively high, being 20.8 and 19.6. The Inner Circle varieties have NFs which are below that of the entire corpus. The NF of 10.3 for AmE is quite close to that of all the texts, but the difference is more noticeable in BrE, where the number is 7.9. As approximately 41% of the tokens are base occurrences in MyE and PhE, the former is closer to the NF of base occurrences in HKE and the latter to that of AmE.

There are 59 out of 83 LS pairs that have relevant tokens in NATURE. Eighteen of them are used in all the varieties. As *tsunami* is the most frequent LS pair in all of the main categories, so it is predictably the most frequent in NATURE, where the pair amounts to 55% of all the tokens. The LS pair in second place is *soy* '[…] soybean […]' (*OED* s.v. *soy, n.*) (henceforth *soy (s2)*) and in third *koi*. There is variance in the three most frequent borrowings and, while *tsunami* is in first place in all of the varieties, none of them has the exact same LS pairs in the top three. AmE and BrE have *soy (s2)* as the second most frequent borrowing and HKE as the third. The LS pair *koi* is the second in MyE and the third in BrE and PhE, but in each variety the number of base occurrences is drastically lower than that of all relevant tokens. The base occurrences are 17% of all tokens in MyE, for example. *Wagyu* is second in HKE and third in SgE and MyE. Lexeme-sense pairs that are in the top

three of only one variety are *kudzu* in AmE and *bonsai* in PhE. The latter is a case similar to *koi* in that the number of base occurrences is extremely low, only 17 out of 292 relevant tokens. What the two LS pairs have in common is that the referents of both have to do with hobbies, as *koi* is '[a] local name in Japan for the common carp, Cyprinus carpio', which are often raised as pets, and *bonsai* '[a] Japanese potted plant or small tree, intentionally dwarfed', which are cultivated as a pastime. The tokens of these two occur in a small number of text components where they are repeated by the same writers in interaction with other enthusiasts. The LS pairs are used less outside of hobby-focused websites.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	tsunami	6,318	tsunami	2,506	tsunami	2,092
2	soy (s2)	1,094	soy (s2)	733	soy (s2)	202
3	koi	681	kudzu	89	koi	135

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	tsunami	486	tsunami	358	tsunami	534	tsunami	342
2	onsen	120	bonsai	292	koi	396	Wagyu	107
3	Wagyu	118	koi	36	Wagyu	91	soy (s2)	58

Table 36. Frequencies of the three most frequent Japanese borrowings in NATURE in each of the varieties studied.

5.2.17 Science and medicine

SCIENCE AND MEDICINE has 3,992 tokens and the NF of 4.2. The main category has one of the lowest percentages of base occurrences compared to relevant tokens, 42%. The great majority of the tokens are from ALTERNATIVE MEDICINE while the other subcategories have 1–25 tokens except MEDICINE, which has 240. ALTERNATIVE MEDICINE has the most tokens in all varieties by far, but all varieties except MyE also have tokens in at least two further subcategories in addition to the aforementioned

category. AmE has tokens in all the subcategories and the second highest frequency is that for MEDICINE. Overall, MEDICINE consists mostly of instances of *Hashimoto*, the majority of which are found in AmE texts. The variety is the only one with tokens in GEOLOGY AND MINERALOGY and PARAPSYCHOLOGY, both of which have one token, an instance of *shibuichi* for the former and *thoughtography* for the latter. The tokens in BrE texts are instances of LS pairs from all the subcategories except the two previous ones. The subcategory with the most tokens after ALTERNATIVE MEDICINE is MEDICINE similarly to AmE, but the number of tokens is low, only 63, in comparison to the 1,402 in ALTERNATIVE MEDICINE. The tokens from SgE texts related to MEDICINE in addition to CHEMISTRY AND BIOCHEMISTRY, where the variety has most of the tokens in the category. PhE has one token in MEDICINE and one in PSYCHOLOGY while the tokens in HKE are in CHEMISTRY AND BIOCHEMISTRY, MEDICINE, PHYSICS, and PSYCHOLOGY. All of the tokens in MyE texts are of the LS pairs in the ALTERNATIVE MEDICINE subcategory.

The highest NF is 11.6 in HKE, but the number of base occurrences is only 36% of all relevant tokens. The NFs for SgE, MyE, AmE, and BrE are close to that of the entire corpus, all being in the range of 3.8–4.7. PhE has a significantly lower NF than the other varieties. This is 2.5, but the NF of base occurrences, 1.5, is of similar size to those of AmE and BrE, where the numbers are respectively 1.7 and 1.6.

The number of LS pairs with relevant tokens is 25 out of 50. Four of the LS pairs, *acupuncture*, *ginkgo*, *reiki*, and *shiatsu*, are used in all the varieties. The most frequent LS pair is *acupuncture*, which yields 66% of all tokens in the main category. The two most frequent pairs are the same for all varieties: *acupuncture* is the most frequent and *reiki* the second. There is some variance in the third most frequent LS pair, which is *Hashimoto* for AmE, *ki* for BrE and SgE, *shiatsu* for PhE, MyE, and HKE. The number of base occurrences is relatively low for *Hashimoto* and *ki* in BrE and the two LS pairs have fewer base occurrences than *shiatsu* and *Ishihara*. In PhE and MyE, the second and third most frequent pairs have such a small number of tokens that there is not much difference compared

to the fourth place, which is *ki* for PhE and *ginkgo* for MyE. Most of the LS pairs in the main category could be considered terms of a special field and some of them are limited to research papers, which is likely why the number of base occurrences is so low and why most of the subcategories have so few tokens.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	acupuncture	2,615	acupuncture	393	acupuncture	991
2	reiki	773	reiki	124	reiki	306
3	Hashimoto	185	Hashimoto	66	ki	45

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	acupuncture	103	acupuncture	68	acupuncture	154	acupuncture	358
2	reiki	52	reiki	12	reiki	8	reiki	48
3	ki	25	shiatsu	8	shiatsu	5	shiatsu	31

Table 37. Frequencies of the three most frequent Japanese borrowings in SCIENCE AND MEDICINE in each of the varieties studied.

5.2.18 Seafaring, vehicles, and transportation

SEAFARING, VEHICLES, AND TRANSPORTATION has 301 relevant tokens and the NF of 0.3. SgE has a markedly high NF of 2.3 while the number is more even for the other varieties, ranging between 0.2 and 0.4. What is interesting is that SgE only has tokens from one LS pair, *Shinkansen*, but the RF is higher than in AmE or BrE, which have significantly larger corpora. Most of the tokens in SgE are, however, from the same author and text component, so the number of base occurrences is lower than in AmE and the same as in BrE.

The number of LS pairs with relevant tokens is six out of nine, with *Shinkansen* being the only borrowing that is used in all the varieties. As can be seen in Table 34, the LS pair is the most frequently used borrowing within the main category overall and in the Outer Circle varieties while it is the second for the Inner Circle varieties, where *kamikaze* '[d]esignating a Japanese aircraft, pilot, or

military unit whose purpose is to fly a suicide mission or missions […]' (*OED* s.v. *kamikaze n. and adj.*) is the most common. The third most frequent borrowing is *jinricksha*, but the tokens are almost exclusively from AmE texts, where the number of base occurrences is as low as three. Other LS pairs in the top three are *kamikaze* '[i]n the Second World War: a Japanese aircraft, typically loaded with explosives, which makes a deliberate suicidal crash into an enemy target, esp. a ship […]' (*OED* s.v. *kamikaze n. and adj.*), *Maru*, and *norimon*.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	Shinkansen	190	kamikaze (adj. s1a)	40	kamikaze (adj. s1a)	33
2	kamikaze (adj. s1a)	80	Shinkansen	35	Shinkansen	21
3	jinricksha	20	jinricksha	18	kamikaze (n. 2a)	4

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	Shinkansen	100	Shinkansen	7	Shinkansen	15	Shinkansen	12
2	-	-	kamikaze (adj. s1a)	6	-	-	jinricksha, kamikaze (adj. s1a), norimon	1
3	-	-	Maru	1	-	-	-	-

Table 38. Frequencies of the three most frequent Japanese borrowings in SEAFARING, VEHICLES, AND TRANSPORTATION in each of the varieties studied.

5.2.19 Society

SOCIETY is fifth among the main categories in terms of the number of tokens, although the category has the most LS pairs from the *OED*. The number of tokens is nevertheless quite high, 9,168 instances, and the NF is 9.7. The tokens are divided into each of the six subcategories: PROFESSIONALS AND HOBBYISTS is the subcategory with the largest percentage of tokens, amounting to 36%. The number is only a little lower for STATUS AND RANK, where it is 30%, while OTHER TYPES OF PEOPLE constitutes

17% of all tokens and TITLES, FORMS OF ADDRESS, AND NAMES 11%. All of the aforementioned subcategories have more than a thousand tokens. The two remaining subcategories, GROUPS AND ORGANIZATIONS and ETHNICITY AND RESIDENCY, have 461 and 248 instances respectively, which corresponds to 4% and 2% of the tokens. The order of the subcategories by number of tokens is the same as by the number of LS pairs, so it is likely that the frequency is affected by how many LS pairs a subcategory has. There was, however, variance in the order of the categories between the varieties, as will be seen below, so the number of LS pairs does not as such explain the order.

The highest NF is recorded for SgE, 26.0. The number of tokens for each LS pair tends to be the highest or second highest among the Outer Circle varieties, and SgE also has the most LS pairs with relevant tokens among the four varieties. PhE is second with the NF of 20.1, followed by MyE and HKE, having frequencies slightly over 17.0. The Inner Circle varieties are both below the NF of the entire corpus, as the number is 8.1 for AmE and 6.7 for BrE. MyE is the only variety where the order of the subcategories from most to least frequent is the same as in the entire corpus. The other varieties generally have one or two of the subcategories changing places. AmE has ETHNICITY AND RESIDENCY below GROUPS AND ORGANIZATIONS, BrE and HKE have STATUS AND RANK as the most frequent subcategory and PROFESSIONALS AND HOBBYISTS as the second, and the latter variety has additionally OTHER TYPES OF PEOPLE in fourth place instead of TITLES, FORMS OF ADDRESS, AND NAMES, which is in third place, as in SgE, and STATUS AND RANK and OTHER TYPES OF PEOPLE are reversed in PhE.

The number of LS pairs with relevant tokens is 74 out of 111. There are 21 that are used in all the varieties. As with GOVERNMENT, ADMINISTRATION, AND POLITICS, the LS pair with the most tokens is *tycoon (s2)*, which is also one of the three most frequent borrowings of this main category in each variety. The LS pairs in the second and third place are *samurai (s1a)* and *ninja* '[a] person trained in the feudal Japanese art of ninjutsu or a modern version of it' (*OED* s.v. *ninja n.*) (henceforth *ninja (n. s1)*), both of which yield over a thousand tokens. The three LS pairs are in the top three for

AmE, BrE, and PhE, although in a different order, as *samurai (s1a)* and *ninja (n. s1)* are reversed in BrE, and AmE and PhE have *ninja (n. s1)* as the most frequent item, followed by *samurai (s1a)* and *tycoon (s2)*. There is more variance in the top three of the remaining Outer Circle varieties, but *tycoon (s2) is* nevertheless in all of them. The LS pair is in first place in MyE and HKE and in third place in SgE. SgE is the only variety where *cosplayer* has the most tokens, closely followed by the honorific *san*. The reason why *cosplayer* 'a person who participates in cosplay' (*OED* s.v. *cosplay n.*) is so high in SgE specifically may be because the variety also has an exceptionally high NF of ENTERTAINMENT AND LEISURE ACTIVITIES. As many other LS pairs related to the borrowing, such as *anime*, *manga*, and *cosplay*, also have high frequencies in the main category, it is not surprising that *cosplayer* would also have quite a few tokens. MyE also has a LS pair that is not included in the top three of any other variety. This is *geisha*, which is in second place, although the borrowing would be the sixth by the number of base occurrences. In general, the base occurrences of *geisha* comprise a smaller percentage of all tokens in some of the Outer Circle varieties than in the Inner Circle ones. The percentage is 25–47% in MyE, PhE and SgE, whereas the percentage is 64–70% for AmE, BrE, and HKE. The last-mentioned has *samurai (s1a)* in the second and *san* in third place.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	tycoon (s2)	1,861	ninja (n. s1)	520	tycoon (s2)	729
2	ninja (n. s1)	1,244	samurai (s1a)	416	samurai (s1a)	382
3	samurai (s1a)	1,158	tycoon (s2)	370	ninja (n. s1)	350

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	cosplayer	197	ninja (n. s1)	198	tycoon (s2)	207	tycoon (s2)	334
2	san	190	samurai (s1a)	117	geisha	99	samurai (s1a)	83
3	tycoon (s2)	117	tycoon (s2)	104	ninja (n. s1)	93	san	27

Table 39. Frequencies of the three most frequent Japanese borrowings in SOCIETY in each of the varieties studied.

5.2.20 Sports and martial arts

SPORTS AND MARTIAL ARTS is the fourth largest main category in terms of the frequency of tokens, which is 10,542 for the category. The overwhelming majority of tokens came from the MARTIAL ARTS subcategory. The number of instances of the LS pairs in Sports is 145, amounting to 1.2% of all relevant tokens. The low frequency of the subcategory is to be expected because of the low number of different LS pairs: there are only three LS pairs, *keirin*, *sayonara* '[d]esignating a hit, esp. a home run, that wins the game' (*OED* s.v. *sayonara int., n., and adj.*), and *Tsukahara*, in the category, compared to the 86 in MARTIAL ARTS. The latter subcategory has the NF of 11.0. Relevant tokens are concentrated in four LS pairs, which all exceed a thousand tokens and comprise 75% of all relevant tokens: *martial art* with 4,342 tokens, *ninja (n. s1)* with 1,244, *karate* with 1,181, and *judo* with 1,124. There is a noticeable gap between these four and the other LS pairs, as the number of tokens for the remaining items is 1–375.

The NFs for all the varieties is generally very high. SgE is far ahead of the other varieties with a NF of 24.7 while the rest are in the range of 9.5–17.2. The NF of the base occurrences in the variety was, however, not equally removed from the others, as the number is 11.5 for SgE and 5.5–9.2 for the other varieties. The relatively low number of base occurrences is mostly because of *martial art*, for which approximately 40% of all tokens are base occurrences. The remaining Outer Circle varieties are close to one another in NFs, as the numbers are 17.2 for MyE, 16.0 for PhE, and 14.8 for HKE. Both Inner Circle varieties are slightly below ten, the figure for BrE being 9.8 and for AmE 9.5. In BrE, the NF is highest among all main categories, which means that one can expect to find more instances of the LS pairs of SPORTS AND MARTIAL ARTS than those of the other main categories in a BrE text.

There are 52 out of the 89 LS pairs from the *OED* that have relevant tokens. Ten of them are found in all the varieties. As noted before, the most frequent LS pairs are *martial art*, *ninja (n. s1)*,

and *karate*. It is not surprising that the first-mentioned has the highest frequency in all varieties, as it is also the fourth most common Japanese borrowing in the entire corpus. The same order is shared by AmE, PhE, and MyE, but there are no major differences between the three and the remaining varieties: BrE and HKE have *judo* in second place and SgE in the third, which is not surprising, as *judo* has almost as high a number of tokens as *karate*. HKE also has *kendo* in third place, which is slightly more unusual, as *kendo* is ninth in the entire corpus. However, the LS pair is used in the same text component by the same author in most cases in HKE, as evidenced by the number of base occurrences being six out of all 30 tokens. The second place in HKE would be shared by *judo* and *karate* and *ninja (n. s1)* would be in third place if the borrowings are ordered by base occurrences.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	martial art	4,342	martial art	1,366	martial art	1,255
2	ninja (n. s1)	1,244	ninja (n. s1)	520	judo	698
3	karate	1,181	karate	379	karate	600

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	martial art	690	martial art	296	martial art	338	martial art	397
2	ninja (n. s1)	62	ninja (n. s1)	198	ninja (n. s1)	93	judo	31
3	judo	58	karate	53	karate	81	kendo	30

Table 40. Frequencies of the three most frequent Japanese borrowings in SPORTS AND MARTIAL ARTS in each of the varieties studied.

5.2.21 Tradition, mythology, and religion

TRADITION, MYTHOLOGY, AND RELIGION has 3,791 tokens and the NF of 4.0. The majority of the tokens are in the RELIGION subcategory, which yields 93% of all relevant tokens in the subcategories. Tokens of the LS pairs in TRADITION made up 6% while MYTHOLOGY has the least occurrences, amounting to approximately one percent. The order is the same as the order by the number of LS

pairs, but the distribution of the tokens is not even, as the LS pairs in RELIGION constitute 70% of the pairs while the portion of TRADITION is 28%, which is noticeably more than in the case of the tokens.

The highest NF is 9.5 for HKE, followed by SgE 8.9. The number of base occurrences is less than half of all the relevant tokens in HKE, which leads to SgE having the highest NF of base occurrences. What is notable about SgE is that the NF of the TRADITION subcategory is quite high at 2.3 while the other varieties are in the range 0.1–0.5. The NF of this subcategory in SgE can mostly be attributed to *hanami* and *kaiseki*, which have 48 and 31 tokens respectively. PhE is the third with a NF of 6.2, but similarly to HKE the number of base occurrences is significantly lower than all of the relevant tokens and the NF of the base occurrences is closer to that of AmE and MyE. The latter two are slightly above the NF of the combined texts. BrE has a markedly low NF of 1.9. The number of instances in BrE texts is comparatively low across the board and almost all of the LS pairs that have relevant tokens have 1–37 instances. The only exception is *Zen*, which has the RF of 466, but this is still only a fourth of the tokens in AmE, the corpus of which is of similar size to that of BrE, for the same LS pair.

The number of LS pairs which have relevant tokens is 52 out of 67. Thirteen of these are found in all varieties. Over half of the tokens are instances of the LS pair *Zen*, which is the most frequent in this main category in all of the varieties. Otherwise there is considerable variance as to which LS pairs belong to the three most frequent LS pairs in each variety. AmE followed the same order as the combined texts with *sensei (s1)* being in second place and *koan* in the third. The LS pair *sensei (s1)* is also the second in PhE and MyE and the third in SgE while *koan* shares the second place with the attributive *Shinto* in BrE. The attributive *Shinto* is also in second place in HKE. Borrowings that are only found in the top three listings of one variety are *Shinto* in the meaning '[t]he native religious system of Japan […]' in BrE, *hanami* in SgE, *bonze* in PhE, *Nichiren* in MyE, and *roshi* in HKE. Noteworthy is that aside from *hanami*, *sensei (s1)* and the two LS pairs with *Shinto* as the lexeme

part, all of the borrowings in the top three are associated Buddhism. This is likely because Shinto is not practiced as much outside of Japan as Buddhism.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	Zen	2059	Zen	1098	Zen	466
2	sensei (s1)	282	sensei (s1)	155	koan, Shinto (n. s1b)	37
3	koan	217	koan	129	Shinto (n. s1a)	27

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	Zen	113	Zen	105	Zen	73	Zen	204
2	hanami	48	sensei (s1)	32	sensei (s1)	23	Shinto (n. s1b)	39
3	sensei (s1)	44	bonze	28	Nichiren	13	roshi	22

Table 41. Frequencies of the three most frequent Japanese borrowings in TRADITION, MYTHOLOGY, AND RELIGION in each of the varieties studied.

5.2.22 Values and concepts

VALUES AND CONCEPTS has 1,280 relevant tokens and the NF 1.4. Aside from SgE, where the number is significantly higher at 3.3, and MyE, which conversely is moderately lower at 0.9, the NFs of the varieties are all quite close to the combined frequency and range between 1.2 and 1.5. An interesting detail is that fewer than half of the relevant tokens are base occurrences. A more careful inspection shows that the number of base occurrences out of all relevant tokens is markedly low specifically in AmE, BrE, and SgE texts and pertains to *reiki* in all of the varieties in addition to *budo* in AmE, *kaizen* and *ki* in BrE, and *ki* in SgE. Consequently, the NFs of the base occurrences for AmE and BrE are actually lower than those of MyE, which means that while AmE and BrE texts are likely to use the LS pairs in the main category more than MyE, the latter variety has more separate text components where the pairs are used.

Most of the LS pairs in VALUES AND CONCEPTS have at least one relevant token, as 15 of all the 19 pairs are found in the corpus texts. Four of the 15 items, *bushido*, *kaizen*, *kawaii (adj.)*, and *reiki*, are used in all the varieties. The three last-mentioned borrowings are overall the most frequent borrowings in this main category. The LS pair *reiki* is noticeably more common than the second and third places, *kawaii (adj.)* and *kaizen*, which are closer to each other in numbers. The top three is identical for the two Inner Circle varieties, where the order from the most to least frequent is *reiki*, *kaizen*, and *ki*. Predictably, *reiki* has markedly more tokens than the other LS pairs, as with the entire corpus. This is not the case with most of the Outer Circle varieties. Although *reiki* is included in the top three most frequent LS pairs in all of them, it is only second in PhE and MyE. While the LS pair is in first place in SgE and HKE, there is not much difference in its frequency compared to that of the second place *kawaii (adj.)*, which in fact has more base occurrences than *reiki*. Only HKE has a similar gap between the first and second place to the Inner Circle varieties. The LS pair *kawaii (adj.)* is the most frequent pair in PhE and MyE, and even if it is in second place in frequency in SgE, the variety has the most tokens of the pair. Although *kawaii (adj.)* is not part of ENTERTAINMENT AND LEISURE ACTIVITIES, it is often used in the same context as LS pairs of that main category, which may explain why it is more common in SgE, PhE, and MyE, which have the highest NFs in ENTERTAINMENT AND LEISURE ACTIVITIES. The LS pair *kaizen* is the second most frequent pair in AmE, BrE, and MyE and the third in MyE and HKE, and *ki* is mostly in third place with the exception of HKE, where it is on the second place.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	reiki	773	reiki	347	reiki	306
2	kawaii (adj.)	130	kaizen	37	kaizen	59
3	kaizen	120	ki	30	ki	45

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	reiki	52	kawaii (adj.)	23	kawaii (adj.)	20	reiki	48
2	kawaii (adj.)	46	reiki	12	kaizen	10	ki	7

3	ki	25	kaizen, ki	6	reiki	8	kawaii (adj.), kaizen	2

Table 42. Frequencies of the three most frequent Japanese borrowings in VALUES AND CONCEPTS in each of the varieties studied.

5.2.23 Weaponry

WEAPONRY has 395 relevant tokens. The NF is 0.4 overall and 0.2–1.4 across the varieties studied. PhE has the highest NF, 1.4, and is the only variety to reach over 1.0. MyE is close, as the frequency is only marginally below 1.0, followed by SgE, which has the frequency 0.8, and AmE with slightly over 0.4. The NFs of HKE and BrE are below that of the entire corpus at 0.3 and 0.2 respectively. The latter variety has the greatest number of different LS pairs with relevant tokens, but half of these yield only one token. Something to note is that the Outer Circle varieties excluding PhE have base occurrences which are fewer than half of those of raw tokens. Especially MyE has a low base occurrence in comparison to the frequency of raw tokens. This stems from the fact that all the tokens of *shuriken* originate from the same author and text component.

There are ten out of all 12 borrowings in the main category that have relevant tokens but only one of them, *katana*, is used in all of the varieties. The LS pair accounts for 67% of all tokens in WEAPONRY and is also the most frequent LS pair in the main category in all the varieties. There are approximately 40 cases where *katana* is used in the context of video games. The other LS pairs have significantly fewer tokens in most of the varieties. The LS pair *shuriken* is in second place overall and similarly placed in AmE, BrE, PhE, and MyE. In third place overall and in AmE is *nunchaku*, which is also in second place for BrE and HKE. Almost half of the tokens of the LS pair come from BrE texts. The other LS pairs in the top three, *menuki*, *tanto*, *tsuba*, and *wakizashi*, are more specific terms for swords or their parts, so it is understandable that they are not as frequent as the three aforementioned pairs.

	All		AmE		BrE	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	katana	263	katana	116	katana	46
2	shuriken	58	shuriken	30	nunchaku, shuriken	12
3	nunchaku	25	nunchaku, wakizashi	6	wakizashi	2

	SgE		PhE		MyE		HKE	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	katana	23	katana	53	katana	16	katana	9
2	tsuba	6	shuriken	4	shuriken	11	nunchaku	4
3	menuki	3	tanto, wakizashi	2	tanto	5	-	-

Table 43. Frequencies of the three most frequent Japanese borrowings in WEAPONRY in each of the varieties studied.

5.2.24 Writing, language, and book-binding

WRITING, LANGUAGE, AND BOOKBINDING has 1,199 tokens and 489 base occurrences, which is notable, because it is only 41% of all relevant tokens. This is the lowest percentage of base occurrences in any main category. Another interesting point is that only LS pairs of one subcategory, WRITING AND LANGUAGE, has tokens while BOOKBINDING has none in any of the varieties studied. BOOKBINDING is a small subcategory of two LS pairs, *orihon* and *Nippon* 'Japanese vellum' (*OED* s.v. *Nippon n. (and adj.)*), so the result is not entirely unexpected even if there are other subcategories of a similar size with tokens.

The NFs vary from 0.3 to 9.8 with PhE having the highest frequency. This frequency is considerably higher compared to the 0.3–3.9 for the other varieties, but the NF of the base occurrences is relatively close to that of SgE, which comes next. AmE is slightly below the combined texts at 1.1. Both SgE and AmE also have relatively few base occurrences in comparison to all the relevant tokens, similarly to PhE. The varieties with a NF of below one are MyE, HKE, and BrE, with the frequencies of 0.9, 0.5, and 0.3 respectively.

There are eight out of 13 LS pairs that have relevant tokens: *Ainu* '[t]he language traditionally spoken by this people [Ainu]' (*OED* s.v. *Ainu n. and adj.*), *Ainu* '[o]f or relating to the Ainu or their language' (*OED* s.v. *Ainu n. and adj.*), *hiragana*, *kana*, *kanji*, *katakana*, *romaji*, and *sumi*. Of these, *hiragana*, *kanji*, *katakana*, and *romaji* are found in all varieties. The three most frequent LS pairs are *kanji* in first place and *hiragana* and *katakana* in the second or third place for all varieties. It may be surprising that *katakana* has 42 tokens fewer than *hiragana* despite the LS pairs referring to analogous writing systems. The reason mainly lies in PhE having many text components where *hiragana* is repeated by the same author, but this is not the case for *katakana*. The LS pairs are close to one another in terms of base occurrences as it is 22 for *hiragana* and 17 for *katakana* in PhE.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	kanji	610	kanji	265	kanji	61
2	hiragana	220	katakana	55	katakana	21
3	katakana	178	hiragana	48	hiragana	11

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	kanji	98	kanji	163	kanji	15	kanji	8
2	hiragana	31	hiragana	119	katakana	9	katakana	5
3	katakana	24	katakana	64	hiragana	8	hiragana	3

Table 44. Frequencies of the three most frequent Japanese borrowings in WRITING, LANGUAGE, AND BOOKBINDING in each of the varieties studied.

5.2.25 Other objects

As noted before, OTHER OBJECTS has the lowest number of tokens in all the main categories, only eight, which gives the NF of 0.008. SgE, PhE, and MyE lack tokens entirely and the other varieties have 1–4 tokens. The small number of tokens is caused by there only being three LS pairs, of which two have relevant tokens in the category. These are *hibachi* and *soroban*, of which the former is found

in BrE and HKE and the latter in AmE and BrE. In BrE, which has both the LS pairs, *hibachi* has one token while *soroban* has three, but the base occurrence of the latter is also one.

	All		**AmE**		**BrE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	hibachi (s1), soroban	4	soroban	1	soroban	3
2	-	-	-	-	hibachi (s1)	1
3	-	-	-	-	-	-

	SgE		**PhE**		**MyE**		**HKE**	
	borrowing	**RF**	**borrowing**	**RF**	**borrowing**	**RF**	**borrowing**	**RF**
1	-	-	-	-	-	-	hibachi (s1)	3
2	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-

Table 45. Frequencies of the three most frequent Japanese borrowings in OTHER OBJECTS in each of the varieties studied.

6 Conclusion

The purpose of this thesis was to investigate the semantic fields of Japanese borrowings in English. The study set out to answer questions regarding the distribution of the borrowings into semantic fields during different time periods and in different Present-day English varieties. The concept of an LS pair was adopted to enable a precise account of the distributions.

The 705 LS pairs from the *OED* belong to 25 semantic main categories, of which 13 have a total of 51 subcategories and 12 have no subcategories. The main categories with the most LS pairs were SOCIETY, ARTS AND CRAFTS, COOKING AND BEVERAGES, SPORTS AND MARTIAL ARTS, NATURE, TRADITION, MYTHOLOGY, AND RELIGION, and SCIENCE AND MEDICINE, which has more than 50 LS pairs each and together accounted for 66.6% of the pairs in all categories. A chronological examination of the years of first attested use reveals that the semantic fields of borrowings were affected by the historical context at the time. Borrowings were adopted sporadically in the time periods before 1850, when the contact between English speakers and Japanese speakers was limited. It is rare for new LS pairs to have entered in consecutive years, and there are many cases where there are multiple years, even decades, with no new pairs attested. The main category of TRADITION, MYTHOLOGY, AND RELIGION dominates with 25 pairs. Notable is also the relatively large number of LS pairs in CURRENCY AND UNITS OF MEASUREMENT, which is the only main category that has more pairs in the years before 1850 than after it. These are likely because the first known Europeans in Japan were merchants and clergy, who would take notice of these aspects. A sizable number of the LS pairs originate from Kæmpfer's *The history of Japan*, which was published in English in 1727 and contained semantically varied LS pairs, but not all of the categories are represented in the 50-year periods between 1550–1850, as only 16 main categories and 35 subcategories have borrowings.

An event that had the greatest effect on the number of new LS pairs was the Opening of Japan starting in 1854, after which the rate of borrowing increases substantially and becomes more regular

than it was before 1850. There are only two years at the most without new LS pairs being attested. The number of LS pairs from the years 1850–1999 is 600, which is 85% of all the LS pairs studied. The range of semantic fields was markedly wide and all the main and subcategories had LS pairs. For example, ARTS AND CRAFTS, MARTIAL ARTS, and SCIENCE AND MEDICINE have particularly many LS pairs in the periods after 1850 compared to before the year.

The use of Japanese borrowings was examined in the Internet texts of six varieties in the GloWbE corpus: two Inner Circle varieties, i.e. American English and British English, and four Outer Circle varieties, i.e. Singaporean English, Philippine English, Malaysian English, and Hong Kong English, which are spoken in regions which are geographically close to Japan and have a history of being under Japanese rule. The results show that the LS pairs are used more in the Outer Circle varieties than in the Inner Circle ones. The highest normalized frequency appears in Singaporean English texts, where the number is 271 per one million words. The normalized frequencies of the other Outer Circle varieties are more even and range between 142–150. The number was significantly lower for the Inner Circle varieties, of which American English has 65 and British English 48. The semantic categories with the most tokens of the LS pairs in all the texts combined are COOKING AND BEVERAGES, ENTERTAINMENT AND LEISURE ACTIVITIES, NATURE, SPORTS AND MARTIAL ARTS, and SOCIETY. The most frequent LS pairs include for example *tsunami*, *anime*, *martial art*, *yen*, and *sushi*. There is variance in the frequency of the LS pairs, which also leads to differences in the frequency of the semantic categories between the varieties.

There are aspects of Japanese borrowings and their use that could not be explored in the frame of this thesis but may be worthwhile to research. The focus of this study was on the use of Japanese borrowings in English varieties which have had a certain degree of language contact with Japanese. It may be interesting to examine how Japanese borrowings are used in varieties spoken in regions which are geographically more removed from Japan and have had less close relations with the country. GloWbE contains texts from 14 varieties besides the six investigated here, so comparisons between

findings from them and the results of this study would be straightforward. This study also made no distinction between the text types in which the borrowings were used and was restricted to written language. Further research into the lexical semantics of Japanese borrowings by text type or in spoken language would give more insight into the subject.

Japanese borrowings have had a relatively short history in English, but they still make up a surprisingly large share of borrowings in the language and span a wide range of semantic fields. Globalisation and the World Wide Web have made the diffusion of lexical innovations possible even without physical interaction between speakers, so one may expect that the years 2000–2049 will bring new borrowings which will characterize this time period.

References

Primary sources

The Oxford English Dictionary Online (*OED*). 2000–. 3rd edition. Oxford: Oxford University Press. Available from http://oed.com/.

Davies, Mark. 2013. *GloWbE: The Corpus of Global Web-Based English*: 1.9 billion words from speakers in 20 countries, 2012–2013. Available from https://corpus.byu.edu/glowbe/.

Secondary sources

”The Internet Archive.” 1996–. Available from https://archive.org/ [Accessed 6 February 2021].

Carman, Christopher P. 1991. “Japanese Loanwords in English.” *Journal of UOEH* 13/3: 217–226.

Census and Statistics Department Hong Kong Special Administrative Region. 2019. *Thematic Household Survey Report No. 66*. Available from https://www.censtatd.gov.hk/en/data/stat_report/product/C0000086/att/B11302662019XXXXB0100.pdf

Cheng, Winnie. 2011. *Exploring Corpus Linguistics: Language in Action.* New York: Taylor & Francis.

Church, Peter. 2017. *A Short History of South-East Asia*. New York: John Wiley & Sons.

Cruse, Alan. 2000. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Cary: Oxford University Press.

Cullen, L. M. 2003. *A History of Japan, 1582–1941: Internal and External Worlds*. Cambridge: Cambridge University Press.

Davies, Mark & Robert Fuchs. 2015. “Expanding Horizons in the Study of World Englishes with the 1.9 Billion Word Global Web-Based English Corpus (GloWbE) [Special Section].” *English World-Wide: A Journal of Varieties of English* 36, 1: 1–47.

Davies, Mark. 2016– *Corpus of News on the Web (NOW)*. Available from https://www.english-corpora.org/now/.

—. 2018-. *The iWeb Corpus*. Available from https://www.english-corpora.org/iWeb/ [Accessed February 6, 2021].

Department of Statistics, Ministry of Trade & Industry, Republic of Singapore. 2011. *Statistical Release 1: Demographic Characteristics, Education, Language and Religion*. Available from https://www.singstat.gov.sg/-/media/files/publications/cop2010/census_2010_release1/cop2010sr1.pdf

Doi, Schun. 2013. "Japanese borrowings found in the Oxford English dictionary and Kæmpfer's the History of Japan" (Doctoral dissertation). Nagoya: Nagoya University Press.

Durkin, Philip. 2014. *Borrowed Words: A History of Loanwords in English*. Oxford: Oxford University Press.

Fennell, Barbara A. 2001. *A History of English: A Sociolinguistic Approach.* Malden: Blackwell.

Friday, Karl, 2001. "Koryū bugei, Japanese." In *Judo in Martial Arts of the World: A–Q*, ed. Thomas A. Green, 301–305. California: ABC-CLIO.

Hadamitzky, Wolfgang and Mark Spahn. 2012. *Japanese Kanji and Kana: A Complete Guide to the Japanese Writing System*. Tokyo: Tuttle Publishing.

Haspelmath, Martin & Uri Tadmor. 2009. *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: De Gruyter Mouton.

Henshall, Kenneth. 2014. *Historical Dictionary of Japan to 1945*. Plymouth: Scarecrow Press.

Garland Cannon. 1984. "Zero Plurals Among the Japanese Loanwords in English." *American Speech* 59/2: 149–158.

Gelderen, Elly van. 2006. *A History of the English Language*. Amsterdam: John Benjamins.

Goto-Jones, Christopher S. 2009. *Modern Japan: A Very Short Introduction*. Oxford: Oxford University Press.

Haspelmath, Martin, and Uri Tadmor. 2009. *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: De Gruyter Mouton.

Henshall, Kenneth. 2014. *Historical Dictionary of Japan to 1945*. Plymouth: Scarecrow Press.

Hickey, Raymond. 2004. *Legacies of Colonial English Studies in Transported Dialects*. Cambridge: Cambridge University Press.

Igarashi, Yuko. 2007. "The Changing Role of Katakana in the Japanese Writing System: Processing and Pedagogical Dimensions for Native Speakers and Foreign Learners" (Doctoral dissertation). Available from http://dspace.library.uvic.ca/bitstream/handle/1828/189/PhD%20dissertation.pdf.

Iwasaki, Shoichi. 2013. *Japanese: Revised edition*. Amsterdam: John Benjamins.

Kæmpfer, Engelbert & Beatrice M. Bodart-Bailey. 1999. *Kaempfer's Japan: Tokugawa Culture Observed*. Honolulu: University of Hawaii Press.

Lipka, Leonhard. 1992. *An Outline of English Lexicology: Lexical Structure, Word Semantics, and Word-Formation*. Second edition. Berlin: De Gruyter.

Long, William J. 2001. "Judo." in *Judo in Martial Arts of the World: A-Q*, ed. Thomas A. Green, 210–216. California: ABC-CLIO.

Lyons, John. 2009. *Semantics*. New York: Cambridge University Press.

McEnery, Tony, and Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

McNally, Mark T. 2015. *Like No Other: Exceptionalism and Nativism in Early Modern Japan*. Honolulu: University of Hawaii Press.

Melchers, Gunnel et al. 2019. *World Englishes*. New York: Taylor and Francis.

Meyer, Milton W. 2012. *Japan: A Concise History.* Maryland: Rowman & Littlefield Publishers.

Minkova, Donka. 2004. "Old English," in *Encyclopedia of Linguistics* ed. Philipp Strazny. 777–780. Chicago: Taylor & Francis.

Millward, C.M. and Mary Hayes. 2012. *A Biography of the English Language.* 3rd edition. Boston: Wadsworth.

Ogilvie, Sarah. 2012. *Words of the World: A Global History of the Oxford English Dictionary*. Cambridge: Cambridge University Press.

Perez, Loius G. 1998. *History of Japan.* West Port: Greenwood Publishing Group.

Russell, David W. 2014. *Blackwell History of the World: History of Japan*. Hoboken: John Wiley & Sons.

Schultz, Julia. 2012. *Twentieth Century Borrowings from French to English: Their Reception and Development*. Newcastle upon Tyne: Cambridge Scholars Publishing.

—. 2017. "The impact of Japanese on the English lexicon since 1900." *Otar Lezunery Bardzraguyn Dprots'um* (transl. *Foreign Languages in Higher Education*) 22: 41–58.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Stefanowitsch, Anatol. 2020. *Corpus Linguistics: A Guide to the Methodology*. Berlin: Language Science Press.

Stubbs, Michael. 1998. "German Loanwords and Cultural Stereotypes." *English Today* 14, 1: 19-26.

Szudarski, Paweł. 2020. *Corpus Linguistics for Vocabulary: A Guide for Research*. London: Routledge.

The Asiatic Society of Japan. 2021. Available from http://asjapan.org/index.html [Accessed March 25, 2021]

The ICE Project. 2016. “The International Corpus of English (ICE).” Available from http://ice-corpora.net/ice/index.html [Accessed February 6, 2021].

Tsuchihashi, Shoko. 1997. “History of Japanese loanwords in English.” *The 23rd LACUS Forum 1996*: 683–692.

University Centre for Computer Corpus Research on Language (UCREL). 2003– “UCREL Home Page.” Available from http://ucrel.lancs.ac.uk/ [Accessed February 6, 2021].

Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Malden: Blackwell.

Yamaguchi, Toshiko and David Deterding. 2016. *English in Malaysia: Current Use and Status*. Leiden: Brill.

Yamamoto, Miyuki. 2020. “English Words of Japanese Origin: A Study of Dictionaries and Corpora” (Master’s thesis). Available from https://andor.tuni.fi/permalink/358FIN_TAMPO/1j3mh4m/alma9911185997805973.

Yule, George. 2012. *The Study of Language*. 4th edition. Cambridge: Cambridge University Press.

Appendices

Appendix 1: Semantic fields of Japanese borrowings in three studies

Tsuchihashi (1997)	Doi (2013)	Schultz (2017)
ARTS	Administrative units	Technology Transport, Nautics, Manufacturing and Engineering, Electronics and Telecommunication
LANGUAGE	Arts and crafts	
RELIGION	Botany	
DAILY LIFE	Bushido	Leisure and Pleasure Tourism, Sports, Theatre, Cartoons, Comics, and Film, Games and Leisure Activities
HISTORY	Characters and letters	
CULINARY	Chemistry and physics	
SPORTS MARTIAL ARTS	Clothing and footwear	
SCIENCE TECHNOLOGY	Commodities	The Fine Arts and Crafts Music, Literature, Art, Flower Arrangement, Jewellery, Cultural Artefacts and Artistic Remains
BUSINESS	Culture	
OTHERS	Customs	
	Economy and business	
	Entertainment	
	Events	
	Food and drink	Civilization and Politics Politics War and the Military, Business, Economy, and Finances
	Games	
	Geography	
	Greetings and chants	Gastronomy Drink, Restaurants and Bars, Cookery
	Historical periods	
	Housing	
	Lineage and family	The Natural Sciences Physics, Geography and Geology, Chemistry and Biochemistry, Mineralogy, Medicine, Psychiatry and Psychology, Parapsychology, Pharmacology, Biology, Microbiology, Zoology, Botany
	Literature	
	Measurement	
	Medicine	
	Mineralogy	
	Monetary	
	Music	
	Nature	
	Professions and status	
	Religion	People and Everyday Life Animal rearing Love and Sexuality Clothing Health and Vitality Educational Systems Buddhism Society, Human Behaviour and Feelings
	Residents and emigrants	
	Social systems	
	Sports and martial arts	
	Transportation	
	Weaponry	The Martial Arts Weaponry Karate Sumo Judo
	Zoology	
	Others	

Appendix 2: Discarded entries and sense and the criteria for exclusion

#	Headword	Criterion for exclusion
1	acupressure n.	1 No etymological connection to Japanese
2	Ashkenazim n.	1 No etymological connection to Japanese
3	bantam n.	1 No etymological connection to Japanese
4	† beˈjape v.	1 No etymological connection to Japanese
5	Capitan China n.	1 No etymological connection to Japanese
6	Chinglish n. and adj.	1 No etymological connection to Japanese
7	diosgenin n.	1 No etymological connection to Japanese
8	Discman n.	1 No etymological connection to Japanese
9	East Indies n.	1 No etymological connection to Japanese
10	emperor n.	1 No etymological connection to Japanese
11	empire n. and adj.	1 No etymological connection to Japanese
12	funny n.1	1 No etymological connection to Japanese
13	ginseng n.	1 No etymological connection to Japanese
14	Ginsu n.	1 No etymological connection to Japanese
15	Goanese n. and adj.	1 No etymological connection to Japanese
16	hapkido n.	1 No etymological connection to Japanese
17	Hinglish n.2 and adj.2	1 No etymological connection to Japanese
18	Huk n.	1 No etymological connection to Japanese
19	hunk n.2 and adj.	1 No etymological connection to Japanese
20	jandal n.	1 No etymological connection to Japanese
21	Jap n.1 and adj.	1 No etymological connection to Japanese
22	jap v.	1 No etymological connection to Japanese
23	Japan n.	1 No etymological connection to Japanese
24	japan v.	1 No etymological connection to Japanese
25	Japanese adj. and n.	1 No etymological connection to Japanese
26	Japaneseness n.	1 No etymological connection to Japanese
27	Japanesery n.	1 No etymological connection to Japanese
28	Japanesey | Japanesy adj.	1 No etymological connection to Japanese
29	Japanesque adj. and n.	1 No etymological connection to Japanese
30	Japanimation n.	1 No etymological connection to Japanese
31	Japanism n.	1 No etymological connection to Japanese
32	Japanize v.	1 No etymological connection to Japanese
33	japanned adj.	1 No etymological connection to Japanese
34	Japanner n.	1 No etymological connection to Japanese
35	japanning n.	1 No etymological connection to Japanese

#	Headword	Criterion for exclusion
58	Neil Robertson n.	1 No etymological connection to Japanese
59	(-)nese, suffix	1 No etymological connection to Japanese
60	oink, int. and n.	1 No etymological connection to Japanese
61	paulownia n.	1 No etymological connection to Japanese
62	polynosic adj. and n.	1 No etymological connection to Japanese
63	ponasterone n.	1 No etymological connection to Japanese
64	prefect n.	1 No etymological connection to Japanese
65	prefecture n.	1 No etymological connection to Japanese
66	primo adj. and n.	1 No etymological connection to Japanese
67	retinospora n.	1 No etymological connection to Japanese
68	rice n.2	1 No etymological connection to Japanese
69	rising sun n.	1 No etymological connection to Japanese
70	Russo-Japanese adj.	1 No etymological connection to Japanese
71	sachaline n.	1 No etymological connection to Japanese
72	Scoville n.	1 No etymological connection to Japanese
73	sexagenal adj.	1 No etymological connection to Japanese
74	Singlish n.1	1 No etymological connection to Japanese
75	Singlish n.2 and adj.	1 No etymological connection to Japanese
76	ˈskimmia n.	1 No etymological connection to Japanese
77	Terra Japonica n.	1 No etymological connection to Japanese
78	vaaljapie n.	1 No etymological connection to Japanese
79	white adj. (and adv.) and n.	1 No etymological connection to Japanese
80	-worthy, comb. Form	1 No etymological connection to Japanese
81	yugawaralite n.	2 Formed within English
82	nunchuck n.	2 Formed within English
83	rickshaw n.	2 Formed within English
84	moose n.3	2 Formed within English
85	reap n.3	2 Formed within English
86	reaping n.	2 Formed within English
87	parametron n.	2 Formed within English
88	Mikadoism n.	2 Formed within English
89	Nipponese adj. and n.	2 Formed within English
90	Nipponian adj.	2 Formed within English
91	Nipponism n.	2 Formed within English
92	Nipponize v.	2 Formed within English

36	Japannish adj.	1 No etymological connection to Japanese
37	jape n.	1 No etymological connection to Japanese
38	jape v.	1 No etymological connection to Japanese
39	japer n.	1 No etymological connection to Japanese
40	† japery n.	1 No etymological connection to Japanese
41	Japhetian adj. and n.	1 No etymological connection to Japanese
42	Japhetic adj.	1 No etymological connection to Japanese
43	ˈjapish adj.	1 No etymological connection to Japanese
44	Japlish n.	1 No etymological connection to Japanese
45	japonate n.	1 No etymological connection to Japanese
46	† Jaˈponian adj. and n.	1 No etymological connection to Japanese
47	Japonic adj.	1 No etymological connection to Japanese
48	japonica n.	1 No etymological connection to Japanese
49	Japonize v.	1 No etymological connection to Japanese
50	ketchup n.	1 No etymological connection to Japanese
51	Kurile n. and adj.	1 No etymological connection to Japanese
52	lotus n.	1 No etymological connection to Japanese
53	macao n.	1 No etymological connection to Japanese
54	mako n.1	1 No etymological connection to Japanese
55	Marrism n.	1 No etymological connection to Japanese
56	mesotron n.	1 No etymological connection to Japanese
57	mow v.5	1 No etymological connection to Japanese

93	Nipponized adj.	2 Formed within English
94	Okinawan n. and adj	2 Formed within English
95	Ryukyuan n. and adj.	2 Formed within English
96	Tokyoite n.	2 Formed within English
97	urushiol n.	2 Formed within English
98	fuji n.	3 Uncertain or unknown etymology
99	pom-pom n.3	3 Uncertain or unknown etymology
100	rumaki n.	3 Uncertain or unknown etymology
101	shaka, int. and n.	3 Uncertain or unknown etymology
102	Roshambo n. (and int.)	3 Uncertain or unknown etymology
103	Betamax n.	4 Proprietary name
104	Kumon n.	4 Proprietary name
105	Midori n.	4 Proprietary name
106	Mikimoto pearl n.	4 Proprietary name
107	Nintendo n.	4 Proprietary name
108	Othello n.	4 Proprietary name
109	Pac-Man n.	4 Proprietary name
110	Pentel n.	4 Proprietary name
111	Tamagotchi n.	4 Proprietary name
112	Walkman n.	4 Proprietary name
113	bukkake n. sense 2	5 After 1999
114	Sudoku n.	5 After 1999

Appendix 3: All lexeme-sense pairs from the *OED*. The markings *lw*, *cq*, *sl*, and *hb* in the *Type of borrowing* column signify *loanword*, *calque*, *semantic loan*, and *hybrid* respectively.

#	Sense number/ section	Headword	Specific lexeme[15]	Subcategories	Time period of first attested use	Type of borrow-ing	*OED* edition
1	1	Kuge, n.		Status and rank	1550–1599	lw	2nd
2	1	bonze, n.		Professionals and hobbyists, Religion	1550–1599	lw	2nd
3	1	kami, n.1		Religion	1600–1649	lw	3rd
4	1	wacadash, n.		Weaponry	1600–1649	lw	2nd
5	1	katana, n.		Weaponry	1600–1649	lw	2nd
6	8	mat, n.1		Units of measurement	1600–1649	sl	3rd
7	1	Nippon, n. (and adj.)		Ethnicity and residency	1600–1649	lw	3rd
8	1	oban, n.		Currency	1600–1649	lw	3rd
9	1	tatami, n.		Architecture, Units of measurement	1600–1649	lw	2nd
10	1	furo, n.		Buildings and property, Interior design and furniture, Leisure activities	1600–1649	lw	3rd
11	1	miso, n.		Food and cooking	1600–1649	lw	3rd
12	1	shogun, n.		Government, administration, and politics, Military and bushido, Status and rank	1600–1649	lw	2nd
13	1	bento, n.		Food and cooking	1600–1649	lw	3rd
14	1	itzebu | itzeboo, n.		Currency	1600–1649	lw	2nd
15	1	kobang, n.		Currency	1600–1649	lw	2nd
16	1	maki-e, n.		Pottery and lacquerware	1600–1649	lw	3rd
17	1	mochi, n.1		Food and cooking	1600–1649	lw	3rd
18	1	norimon, n.		Seafaring, vehicles and transportation	1600–1649	lw	3rd
19	1	samisen, n.		Music	1600–1649	lw	2nd
20	1	tabi, n.		Footwear	1600–1649	lw	2nd
21	1	inro, n.		Pottery and lacquerware, Accessories	1600–1649	lw	2nd
22	1	menuki, n.		Weaponry	1600–1649	lw	3rd
23	1	mikan, n.		Food and cooking, Flora	1600–1649	lw	3rd
24	1	tai, n.1		Food and cooking, Fauna	1600–1649	lw	2nd
25	1	dairi, n.		Buildings and property, Titles, forms of address and names	1650–1699	lw	2nd
26	1	kami-sama, n.1		Government, administration, and politics, Titles, forms of address, and names	1650–1699	lw	3rd
27	1	moxa, n.		Alternative medicine	1650–1699	lw	3rd
28	1	acupuncture, n.		Alternative medicine	1650–1699	hb	3rd
29	1	soy, n.1		Food and cooking	1650–1699	lw	2nd
30	1	akoya, n.		Gemstones, Fauna	1700–1749	lw	3rd
31	1	chanoyu, n.		Beverages	1700–1749	lw	3rd
32	1	hatamoto, n.		Military and bushido, Status and rank	1700–1749	lw	3rd
33	1	hinoki, n.		Manufacturing and materials, Flora	1700–1749	lw	2nd
34	1	Hizen, n.		Pottery and lacquerware	1700–1749	lw	2nd
35	1a	Jōdo, n.		Religion	1700–1749	lw	2nd
36	1	kaki, n.		Food and cooking, Flora	1700–1749	lw	2nd
37	1	kami, n.2		Government, administration, and politics, Status and rank, Titles, forms of address, and names	1700–1749	lw	3rd
38	1	kana, n.		Writing and language	1700–1749	lw	2nd
39	1	katakana, n.		Writing and language	1700–1749	lw	2nd
40	1	katsuo, n.		Food and cooking, Fauna	1700–1749	lw	2nd
41	1	ken, n.3		Units of measurement	1700–1749	lw	2nd
42	1	kiri, n.		Flora	1700–1749	lw	2nd
43	1	kirin, n.		Painting, pictures, and printing, Pottery and lacquerware, Mythology	1700–1749	lw	2nd
44	1	koi, n.		Fauna	1700–1749	lw	2nd
45	1	koi-cha, n.		Beverages	1700–1749	lw	2nd

[15] *Specific lexeme* here refers to a lexeme that is found in the *OED* under the headword but is not the same as the headword. For example, the specific lexeme *rain door* is found under the headword *rain* in the *OED* (s.v. rain n.). Lexemes that are the same as the headword have not been included in the Specific lexeme column.

46	1a	koku, n.		Units of measurement	1700–1749	lw	2nd
47	1	kuruma, n.		Seafaring, vehicles and transportation	1700–1749	lw	2nd
48	1	matsuri, n.		Religion	1700–1749	lw	3rd
49	1	Mikado, n.		Government, administration, and politics, Status and rank, Titles, forms of address, and names	1700–1749	lw	3rd
50	1	mikoshi, n.		Religion	1700–1749	lw	3rd
51	1	miya, n.		Buildings and property, Religion	1700–1749	lw	3rd
52	1	momme, n.		Units of measurement	1700–1749	lw	3rd
53	1	nembutsu, n.		Religion	1700–1749	lw	3rd
54	1	Ritsu, n.		Religion	1700–1749	lw	3rd
55	1a	samurai, n.		Military and bushido, Professionals and hobbyists, Status and rank	1700–1749	lw	2nd
56	1	satori, n.		Religion	1700–1749	lw	2nd
57	1	sen, n.1		Currency	1700–1749	lw	2nd
58	1	shaku, n.		Units of measurement	1700–1749	lw	2nd
59	1	shikimi, n.		Flora, Religion	1700–1749	lw	2nd
60	1	Shingon, n.		Religion	1700–1749	lw	2nd
61	1	Shinshū, n.		Religion	1700–1749	lw	2nd
62	1a	Shinto, n.		Religion	1700–1749	lw	2nd
63	1b	Shinto, n.		Religion	1700–1749	lw	2nd
64	1	shoyu, n.		Food and cooking	1700–1749	lw	2nd
65	1	sugi, n.		Manufacturing and materials, Flora	1700–1749	lw	2nd
66	1	sun, n.2		Units of measurement	1700–1749	lw	2nd
67	1	Tendai, n.		Religion	1700–1749	lw	2nd
68	1	tokonoma, n.		Architecture	1700–1749	lw	2nd
69	1	torii, n.		Religion	1700–1749	lw	2nd
70	1	tsubo, n.		Units of measurement	1700–1749	lw	2nd
71	1	urushi, n.		Pottery and lacquerware, Flora	1700–1749	lw	2nd
72	1	wakizashi, n.		Weaponry	1700–1749	lw	3rd
73	1	zazen, n.		Religion	1700–1749	lw	2nd
74	1	Zen, n.		Religion	1700–1749	lw	2nd
75	1	adzuki, n.		Food and cooking, Flora	1700–1749	lw	3rd
76	1	ginkgo, n.		Flora	1750–1799	lw	3rd
77	1	Nandina, n.		Flora	1750–1799	hb	3rd
78	comp	miso, n.	miso soup n.	Food and cooking	1750–1799	hb	3rd
79	1	koto, n.		Music	1750–1799	lw	2nd
80	1	saké, n.2		Beverages	1750–1799	lw	2nd
81	1	Ryukyu, adj.		Ethnicity and residency	1800–1849	lw	3rd
82	1	Ainu, n.		Ethnicity and residency	1800–1849	lw	3rd
83	1	ri, n.		Units of measurement	1800–1849	lw	3rd
84	1	aucuba, n.		Flora	1800–1849	lw	2nd
85	1	gobo, n.1		Food and cooking, Flora	1800–1849	lw	3rd
86	1	hiragana, n.		Writing and language	1800–1849	lw	2nd
87	1	koro, n.		Pottery and lacquerware	1800–1849	lw	2nd
88	1	noshi, n.		Paper folding and packaging	1800–1849	lw	3rd
89	1	Roju, n.		Government, administration, and politics, Groups and organizations, Status and rank	1800–1849	lw	3rd
90	1	ume, n.		Flora	1800–1849	lw	3rd
91	1	umeboshi, n.		Food and cooking	1800–1849	lw	3rd
92	1	yukata, n.		Clothes	1800–1849	lw	2nd
93	1	zori, n.		Footwear	1800–1849	lw	2nd
94	2	moxa, n.		Alternative medicine	1800–1849	lw	3rd
95	1	iroha, n.		Writing and language	1800–1849	lw	2nd
96	1	jito, n.		Government, administration, and politics, Military and bushido, Status and rank	1800–1849	lw	2nd
97	2	Shinto, n.		Other people, Religion	1800–1849	lw	2nd
98	1	Obaku, n.		Religion	1800–1849	lw	3rd
99	1	Rinzai, n.		Religion	1800–1849	lw	3rd
100	3	Ainu, adj.		Ethnicity and residency, Writing and language	1800–1849	lw	3rd
101	1	daimio, n.		Government, administration, and politics, Titles, forms of address, and names	1800–1849	lw	2nd
102	1	go, n.2		Games	1800–1849	lw	3rd
103	1	manyogana, n.		Writing and language	1800–1849	lw	3rd

104	1	akebia, n.		Flora	1800–1849	hb	3rd
105	1	renga, n.		Literature	1850–1899	lw	3rd
106	1	uta, n.		Literature	1850–1899	lw	2nd
107	1	hara-kiri, n.		Military and bushido	1850–1899	lw	2nd
108	1	kago, n.		Seafaring, vehicles and transportation	1850–1899	lw	2nd
109	1	tycoon, n.		Government, administration, and politics, Status and rank, Titles, forms of address, and names	1850–1899	lw	2nd
110	1	ronin, n.		Military and bushido	1850–1899	lw	3rd
111	1	shogi, n.		Games	1850–1899	lw	2nd
112	1	shakudo, n.		Other crafts, Manufacturing and materials, Geology and mineralogy	1850–1899	lw	2nd
113	1	tempo, n.2		Currency	1850–1899	lw	2nd
114	1	mon, n.2		Tradition	1850–1899	lw	3rd
115	1	mousmé, n.		Professionals and hobbyists, Other people	1850–1899	lw	3rd
116	2	tycoon, n.		Business and working culture, Government, administration, and politics, Status and rank	1850–1899	lw	2nd
117	1	aka-matsu, n.		Manufacturing and materials, Flora	1850–1899	lw	3rd
118	1	mebos, n.		Food and cooking	1850–1899	lw	3rd
119	1(c)	red pine, n.		Manufacturing and materials, Flora	1850–1899	sl	3rd
120	1	ryo, n.		Currency	1850–1899	lw	3rd
121	1	hibachi, n.		Other objects	1850–1899	lw	2nd
122	1	sayonara, int.		Greetings, interjections, and gestures	1850–1899	lw	3rd
123	1	yashiki, n.		Buildings and property	1850–1899	lw	2nd
124	1	Yokohama, n.		Fauna	1850–1899	lw	3rd
125	2	Yokohama, n.		Manufacturing and materials	1850–1899	lw	3rd
126	1	sasanqua, n.		Manufacturing and materials, Flora	1850–1899	lw	2nd
127	1	nandin, n.		Flora	1850–1899	lw	3rd
128	comp	rain, n.1	rain door n.	Architecture	1850–1899	cq	3rd
129	1	rin, n.2		Currency	1850–1899	lw	3rd
130	1	o-muraji, n.		Government, administration, and politics, Status and rank	1850–1899	lw	3rd
131	1	keyaki, n.		Manufacturing and materials, Flora	1850–1899	lw	3rd
132	1	inkyo, n.		Business and working culture	1850–1899	lw	2nd
133	1	junshi, n.		Military and bushido	1850–1899	lw	2nd
134	1	ki-mon, n.		Tradition	1850–1899	lw	2nd
135	1	kyogen, n.		Theater, puppetry, and dance	1850–1899	lw	2nd
136	1	mitsumata, n.		Manufacturing and materials, Flora	1850–1899	lw	3rd
137	1	Noh, n.		Theater, puppetry, and dance	1850–1899	lw	3rd
138	1	oiran, n.		Professionals and hobbyists	1850–1899	lw	3rd
139	comp	reign, n.	reign name n.	Government, administration, and politics, Historical periods	1850–1899	cq	3rd
140	1	seppuku, n.		Military and bushido	1850–1899	lw	2nd
141	1	tan, n.3		Units of measurement	1850–1899	lw	2nd
142	1	to, n.		Units of measurement	1850–1899	lw	2nd
143	1	uguisu, n.		Fauna	1850–1899	lw	2nd
144	1	obi, n.1		Clothes	1850–1899	lw	3rd
145	1	Satsuma, n.		Pottery and lacquerware	1850–1899	lw	2nd
146	1	sayonara, n.		Greetings, interjections, and gestures	1850–1899	lw	3rd
147	1	kamidana, n.		Religion	1850–1899	lw	3rd
148	1	Meiji, n.		Historical periods	1850–1899	lw	3rd
149	1	fundoshi, n.		Clothes	1850–1899	lw	3rd
150	1	jinricksha | jinrikisha, n.		Seafaring, vehicles and transportation	1850–1899	lw	2nd
151	1	kakke, n.		Medicine	1850–1899	lw	2nd
152	1	miko, n.		Professionals and hobbyists, Religion	1850–1899	lw	3rd
153	1	mirin, n.		Food and cooking	1850–1899	lw	3rd
154	1	sencha, n.		Beverages	1850–1899	lw	3rd
155	1	sensei, n.		Painting, pictures, and printing, Education, Professionals and hobbyists, Other people, Titles, forms of address, and names, Religion	1850–1899	lw	3rd
156	1	yen, n.1		Currency	1850–1899	lw	2nd
157	1	aikuchi, n.		Weaponry	1850–1899	lw	3rd

158	1	gobang, n.		Games	1850–1899	lw	3rd
159	1	heimin, n.		Status and rank	1850–1899	lw	2nd
160	1	Imari, n.		Pottery and lacquerware	1850–1899	lw	2nd
161	1	ju-jitsu, n.		Martial arts	1850–1899	lw	2nd
162	1	raku, n.		Pottery and lacquerware	1850–1899	lw	3rd
163	1	ryugi, n.		Arrangement, Music, Painting, pictures, and printing, Theater, puppetry, and dance, Martial arts	1850–1899	lw	3rd
164	1	sennin, n.		Other people, Mythology, Religion	1850–1899	lw	2nd
165	2	shaku, n.		Tradition	1850–1899	lw	2nd
166	1	shippo, n.		Pottery and lacquerware	1850–1899	lw	2nd
167	1	yokan, n.		Food and cooking	1850–1899	lw	3rd
168	1	bugaku, n.		Music, Theater, puppetry, and dance	1850–1899	lw	3rd
169	1	futon, n.		Interior design and furniture	1850–1899	lw	2nd
170	1	Genro, n.		Government, administration, and politics, Groups and organizations, Status and rank	1850–1899	lw	2nd
171	1	geta, n.		Footwear	1850–1899	lw	3rd
172	comp	god, n. and int.	god shelf n.	Religion	1850–1899	cq	3rd
173	1	kotatsu, n.		Interior design and furniture	1850–1899	lw	2nd
174	1	magatama, n.		Gemstones	1850–1899	lw	3rd
175	1	monogatari, n.		Literature	1850–1899	lw	3rd
176	1	mura, n.		Government, administration, and politics	1850–1899	lw	3rd
177	1	netsuke, n.		Accessories	1850–1899	lw	3rd
178	1	Nichiren, n.		Groups and organizations, Religion	1850–1899	lw	3rd
179	1	Ramanas rose, n.		Flora	1850–1899	hb	3rd
180	1	sho, n.1		Units of measurement	1850–1899	lw	2nd
181	1	tan, n.4		Units of measurement, Manufacturing and materials	1850–1899	lw	2nd
182	1	Tokugawa, adj.		Historical periods	1850–1899	lw	2nd
183	1	uji, n.		Groups and organizations, Titles, forms of address and names	1850–1899	lw	2nd
184	1	haori, n.		Clothes	1850–1899	lw	2nd
185	1	matsutake, n.		Food and cooking, Fungi	1850–1899	lw	3rd
186	1	pillow word, n.		Literature	1850–1899	cq	3rd
187	1	shiitake, n.		Food and cooking, Fungi	1850–1899	lw	2nd
188	1	Shin, n.3		Religion	1850–1899	lw	2nd
189	1	tanka, n.2		Literature	1850–1899	lw	2nd
190	1	uchiwa, n.		Other objects	1850–1899	lw	2nd
191	1	koji, n.		Beverages, Food and cooking, Chemistry and biochemistry	1850–1899	lw	2nd
192	1	san, n.3		Titles, forms of address and names	1850–1899	lw	2nd
193	1	Arita, n.		Pottery and lacquerware	1850–1899	lw	2nd
194	1	Tosa, n.1		Painting, pictures, and printing	1850–1899	lw	2nd
195	1	ukiyo-e, n.		Painting, pictures, and printing	1850–1899	lw	2nd
196	1	Yamato, n.		Painting, pictures, and printing	1850–1899	lw	2nd
197	2	Ainu, n.		Writing and language	1850–1899	lw	3rd
198	1	amado, n.		Architecture	1850–1899	lw	2nd
199	1	fusuma, n.		Architecture	1850–1899	lw	2nd
200	1	happi, n.		Clothes	1850–1899	lw	3rd
201	1	Hirado, n.		Pottery and lacquerware	1850–1899	lw	2nd
202	1	kura, n.		Buildings and property	1850–1899	lw	2nd
203	1	Kutani, n.		Pottery and lacquerware	1850–1899	lw	2nd
204	1	maguro, n.		Food and cooking, Fauna	1850–1899	lw	3rd
205	1	makimono, n.		Painting, pictures, and printing	1850–1899	lw	3rd
206	1	marumage, n.		Hairstyles	1850–1899	lw	3rd
207	1	Nabeshima, adj. and n.		Pottery and lacquerware	1850–1899	lw	3rd
208	1	nakodo, n.		Other people	1850–1899	lw	3rd
209	1	nashiji, n.		Pottery and lacquerware	1850–1899	lw	3rd
210	1	okimono, n.		Interior design and furniture	1850–1899	lw	3rd
211	1	sashimi, n.		Food and cooking	1850–1899	lw	2nd

212	1	shibuichi, n.		Other crafts, Manufacturing and materials, Geology and mineralogy	1850–1899	lw	2nd
213	1	shoji, n.		Architecture	1850–1899	lw	2nd
214	2	soy, n.1		Flora	1850–1899	lw	2nd
215	1	sumo, n.		Martial arts	1850–1899	lw	2nd
216	1	Tanabata, n.		Tradition	1850–1899	lw	2nd
217	1	temmoku, n.		Pottery and lacquerware	1850–1899	lw	2nd
218	1	tofu, n.		Food and cooking	1850–1899	lw	2nd
219	1	matcha, n.		Beverages, Food and cooking	1850–1899	lw	3rd
220	1	Seto, n.		Pottery and lacquerware	1850–1899	lw	2nd
221	1	togidashi, n.		Pottery and lacquerware	1850–1899	lw	2nd
222	1	ken, n.4		Government, administration, and politics	1850–1899	lw	2nd
223	1	Nara, adj.		Historical periods	1850–1899	lw	3rd
224	1	orihon, n.		Bookbinding	1850–1899	lw	3rd
225	2	Satsuma, n.		Food and cooking, Flora	1850–1899	lw	2nd
226	1	hechima, n.		Flora	1850–1899	lw	2nd
227	1	hinin, n.		Status and rank	1850–1899	lw	2nd
228	1	joro, n.		Professionals and hobbyists	1850–1899	lw	2nd
229	1	kagura, n.		Theater, puppetry, and dance, Religion	1850–1899	lw	2nd
230	1	kikyo, n.		Flora	1850–1899	lw	2nd
231	1	kombu, n.		Food and cooking, Flora	1850–1899	lw	2nd
232	1	koniak | koniaku, n.		Food and cooking, Flora	1850–1899	lw	2nd
233	1	matsu, n.		Manufacturing and materials, Flora	1850–1899	lw	3rd
234	1	miai, n.		Tradition	1850–1899	lw	3rd
235	1	mokume, n.		Other crafts	1850–1899	lw	3rd
236	1	nanten, n.		Flora	1850–1899	lw	3rd
237	1	sakura, n.		Manufacturing and materials, Flora	1850–1899	lw	2nd
238	1	Shijō, n.		Painting, pictures, and printing	1850–1899	lw	2nd
239	1	Kuroshiwo, n.		Environment	1850–1899	lw	2nd
240	1	romaji, n.		Writing and language	1850–1899	lw	3rd
241	1	tansu, n.		Interior design and furniture	1850–1899	lw	2nd
242	1	tanto, n.2		Weaponry	1850–1899	lw	2nd
243	1	tsukemono, n.		Food and cooking	1850–1899	lw	2nd
244	1b	Jōdo, n.		Religion	1850–1899	lw	2nd
245	1	kimono, n.		Clothes	1850–1899	lw	2nd
246	1	Nabeshimayaki, n.		Pottery and lacquerware	1850–1899	lw	3rd
247	1	shikimic, adj.		Chemistry and biochemistry	1850–1899	hb	2nd
248	2	shoji, n.		Architecture	1850–1899	lw	2nd
249	2b	tea man, n.		Beverages, Professionals and hobbyists	1850–1899	sl	3rd
250	comp	head, n.1	head drop n.	Medicine	1850–1899	cq	3rd
251	2	nembutsu, n.		Religion	1850–1899	lw	3rd
252	1	shō, n.2		Music	1850–1899	lw	2nd
253	1	† mokum, n.		Other crafts	1850–1899	lw	3rd
254	1	awabi, n.		Food and cooking, Fauna	1850–1899	lw	2nd
255	1	bekko, n.		Manufacturing and materials	1850–1899	lw	2nd
256	1	judo, n.		Martial arts	1850–1899	lw	3rd
257	1	kaya, n.		Manufacturing and materials, Flora	1850–1899	lw	2nd
258	1	Maru, n.		Seafaring, vehicles and transportation	1850–1899	lw	3rd
259	1	nageire, n.		Arrangement	1850–1899	lw	3rd
260	1	ojime, n.		Accessories	1850–1899	lw	3rd
261	1	reishi, n.		Fungi, Alternative medicine	1850–1899	lw	3rd
262	1	rikka, n.		Arrangement	1850–1899	lw	3rd
263	2	shogun, n.		Painting, pictures, and printing	1850–1899	lw	2nd
264	1	tsuba, n.		Weaponry	1850–1899	lw	2nd
265	1	zabuton, n.		Interior design and furniture	1850–1899	lw	2nd
266	1	jōruri, n.		Music, Theater, puppetry, and dance	1850–1899	lw	2nd
267	1	kabane, n.		Government, administration, and politics, Titles, forms of address, and names	1850–1899	lw	2nd
268	1	kakemono, n.		Interior design and furniture, Painting, pictures, and printing	1850–1899	lw	2nd
269	1	Kakiemon, n.		Pottery and lacquerware	1850–1899	lw	2nd
270	1	ken, n.5		Games	1850–1899	lw	2nd
271	1	mitsuba, n.		Food and cooking, Flora	1850–1899	lw	3rd

272	1	suimono, n.		Food and cooking	1850–1899	lw	2nd
273	2	thought control, n.		Government, administration, and politics	1850–1899	sl	3rd
274	1	geisha, n.		Professionals and hobbyists	1850–1899	lw	2nd
275	1	hanami, n.		Tradition	1850–1899	lw	2nd
276	1	hanashika, n.		Theater, puppetry, and dance, Professionals and hobbyists	1850–1899	lw	2nd
277	der	katsuo, n.		Food and cooking	1850–1899	lw	2nd
278	1	maiko, n.		Professionals and hobbyists	1850–1899	lw	3rd
279	1	misoshiru, n.		Food and cooking	1850–1899	lw	3rd
280	1	muraji, n.		Government, administration, and politics, Status and rank, Titles, forms of address, and names	1850–1899	lw	3rd
281	1	omi, n.1		Government, administration, and politics, Status and rank, Titles, forms of address, and names	1850–1899	lw	3rd
282	1	sika, n.1		Fauna	1850–1899	lw	2nd
283	1	soroban, n.		Other objects	1850–1899	lw	2nd
284	1	kami-sama, n.2		Religion	1850–1899	lw	3rd
285	1	maegashira, n.		Status and rank, Martial arts	1850–1899	lw	3rd
286	1	nashi, n.		Food and cooking, Flora	1850–1899	lw	3rd
287	1	nori, n.		Food and cooking, Flora	1850–1899	lw	3rd
288	1	banzai, int.		Greetings, interjections, and gestures	1850–1899	lw	2nd
289	1	Heian, adj.		Historical periods	1850–1899	lw	2nd
290	1	ki, n.2		Alternative medicine, Martial arts, Values and concepts	1850–1899	lw	3rd
291	1	kudzu, n.		Flora	1850–1899	lw	2nd
292	1	shakuhachi, n.		Music	1850–1899	lw	2nd
293	1	shugo, n.		Government, administration, and politics, Military and bushido, Status and rank	1850–1899	lw	2nd
294	1	Soto, n.		Religion	1850–1899	lw	2nd
295	1	sushi, n.		Food and cooking	1850–1899	lw	2nd
296	1	janken, n. (and int.)		Games	1850–1899	lw	3rd
297	comp	plum, n. and adj.2	plum rains n.	Environment	1850–1899	cq	3rd
298	1	soshi, n.		Government, administration, and politics, Professionals and hobbyists, Other people	1850–1899	lw	2nd
299	1	yokozuna, n.		Clothes, Professionals and hobbyists, Status and rank, Martial arts	1850–1899	lw	3rd
300	1	baren, n.		Painting, pictures, and printing	1850–1899	lw	2nd
301	1	habu, n.		Fauna	1850–1899	lw	2nd
302	1	habutai, n.		Manufacturing and materials	1850–1899	lw	2nd
303	1	hakama, n.		Clothes	1850–1899	lw	2nd
304	1	kamikaze, n.		Military and bushido, Environment, Religion	1850–1899	lw	3rd
305	1	metake, n.		Flora	1850–1899	lw	3rd
306	1	onsen, n.		Buildings and property, , Environment	1850–1899	lw	3rd
307	1	soba, n.		Food and cooking	1850–1899	lw	2nd
308	1	Taka-diastase, n.		Chemistry and biochemistry	1850–1899	lw	2nd
309	1	Eta | eta, n.3		Status and rank	1850–1899	lw	2nd
310	1	tsunami, n.		Environment	1850–1899	lw	2nd
311	1	ujigami, n.		Religion	1850–1899	lw	2nd
312	1	bushido, n.		Military and bushido, Values and concepts	1850–1899	lw	2nd
313	1	Kōrin, n.2		Painting, pictures, and printing	1850–1899	lw	2nd
314	1	makunouchi, n.		Martial arts	1850–1899	lw	3rd
315	2	Meiji, adj.		Historical periods	1850–1899	lw	3rd
316	B7	pink, n.5 and adj.2		Leisure activities	1850–1899	sl	3rd
317	1	Bon, n.		Religion	1850–1899	lw	2nd
318	1	haiku, n.		Literature	1850–1899	lw	2nd
319	1	Kabuki, n.		Theater, puppetry, and dance	1850–1899	lw	2nd
320	1	natto, n.		Food and cooking	1850–1899	lw	3rd
321	1	omochi, n.		Food and cooking	1850–1899	lw	3rd
322	2	pivot word, n.		Literature	1850–1899	sl	3rd

323	1	surimono, n.		Painting, pictures, and printing	1850–1899	lw	2nd
324	1	Kobe, n.		Food and cooking, Fauna	1900–1949	lw	3rd
325	1	Shiga, n.		Medicine	1900–1949	lw	2nd
326	1	ikebana, n.		Arrangement	1900–1949	lw	2nd
327	1	onnagata, n.		Theater, puppetry, and dance, Professionals and hobbyists	1900–1949	lw	3rd
328	1	o-omi, n.		Government, administration, and politics, Status and rank	1900–1949	lw	3rd
329	1	sayonara, adj.		Greetings, interjections, and gestures	1900–1949	lw	3rd
330	1	Kamakura, n.		Painting, pictures, and printing	1900–1949	lw	2nd
331	1	kombucha, n.		Beverages	1900–1949	lw	3rd
332	1	nembutsu, adj.		Religion	1900–1949	lw	3rd
333	1	obento, n.		Food and cooking	1900–1949	lw	3rd
334	1	sentoku, n.		Manufacturing and materials	1900–1949	lw	2nd
335	1	yūzen, n.		Manufacturing and materials	1900–1949	lw	2nd
336	comp	goblin, n.1	goblin shark n.	Fauna	1900–1949	cq	3rd
337	2	Mikado, n.		Seafaring, vehicles and transportation	1900–1949	lw	3rd
338	1	wasabi, n.		Food and cooking, Flora	1900–1949	lw	2nd
339	1	† shimose, n.		Military and bushido, Chemistry and biochemistry	1900–1949	lw	2nd
340	2b	culture, n.	culture pearl n.	Gemstones	1900–1949	cq	3rd
341	1	makikomi, n.		Martial arts	1900–1949	lw	3rd
342	1	mama-san, n.		Professionals and hobbyists, Status and rank, Other people	1900–1949	lw	3rd
343	1a	papasan, n.		Other people, Titles, forms of address and names	1900–1949	lw	3rd
344	1	rotenone, n.		Chemistry and biochemistry	1900–1949	lw	3rd
345	1c	Shinto, n.		Religion	1900–1949	lw	2nd
346	1	budo, n.		Martial arts, Values and concepts	1900–1949	lw	3rd
347	1	maitake, n.		Food and cooking, Fungi	1900–1949	lw	3rd
348	1	mawashi, n.		Clothes, Martial arts	1900–1949	lw	3rd
349	1b	samurai, n.		Other people	1900–1949	lw	2nd
350	2	year man, n.		Other people	1900–1949	cq	3rd
351	1	medaka, n.		Fauna	1900–1949	lw	3rd
352	3	Mikado, n.		Fauna	1900–1949	lw	3rd
353	1	nage-no-kata, n.		Martial arts	1900–1949	lw	3rd
354	1	sutemi-waza, n.		Martial arts	1900–1949	lw	2nd
355	1	tomoe-nage, n.		Martial arts	1900–1949	lw	2nd
356	1	tsurikomi, n.		Martial arts	1900–1949	lw	2nd
357	1	tsutsugamushi, n.		Medicine	1900–1949	lw	2nd
358	1	uchimata, n.		Martial arts	1900–1949	lw	2nd
359	1	uki, n.		Martial arts	1900–1949	lw	2nd
360	1	ura-nage, n.		Martial arts	1900–1949	lw	3rd
361	1	rikishi, n.		Professionals and hobbyists, Martial arts	1900–1949	lw	3rd
362	1	katsura, n.		Accessories	1900–1949	lw	2nd
363	1	mompe, n.		Clothes	1900–1949	lw	3rd
364	1	roshi, n.		Professionals and hobbyists, Titles, forms of address and names, Religion	1900–1949	lw	3rd
365	1	terakoya, n.		Education	1900–1949	lw	2nd
366	1	bai-u, n.		Environment	1900–1949	lw	2nd
367	1	Nanga, n.2		Painting, pictures, and printing	1900–1949	lw	3rd
368	1	randori, n.		Martial arts	1900–1949	lw	3rd
369	2	sayonara, n.		Greetings, interjections, and gestures	1900–1949	lw	3rd
370	1	shimada, n.		Hairstyles	1900–1949	lw	2nd
371	1	yuzu, n.		Food and cooking, Flora	1900–1949	lw	3rd
372	2c	belt, n.1		Clothes, Martial arts	1900–1949	sl	3rd
373	3	black belt, n.		Clothes, Professionals and hobbyists, Status and rank, Martial arts	1900–1949	sl	3rd
374	1	dotaku, n.		Other crafts	1900–1949	lw	2nd
375	1b	koku, n.		Units of measurement	1900–1949	lw	2nd
376	1	mizuna, n.		Food and cooking, Flora	1900–1949	lw	3rd
377	1	sewamono, n.		Theater, puppetry, and dance	1900–1949	lw	2nd

378	1	shosagoto, n.		Theater, puppetry, and dance	1900–1949	lw	2nd
379	1	sumi, n.		Writing and language	1900–1949	lw	2nd
380	1	suiboku, n.		Painting, pictures, and printing	1900–1949	lw	2nd
381	1	mu, n.3		Religion, Values and concepts	1900–1949	lw	3rd
382	1	ozeki, n.		Status and rank, Titles, forms of address and names, Martial arts	1900–1949	lw	3rd
383	1	shodan, n.		Professionals and hobbyists, Status and rank, Martial arts	1900–1949	lw	2nd
384	1	yondan, n.		Martial arts	1900–1949	lw	2nd
385	1	maki zushi, n.		Food and cooking	1900–1949	lw	3rd
386	1	nightingale floor, n.		Architecture	1900–1949	cq	3rd
387	1	ryokan, n.		Buildings and property, Travelling and tourism	1900–1949	lw	3rd
388	1	katsuramono, n.		Theater, puppetry, and dance	1900–1949	lw	2nd
389	1	nogaku, n.		Music, Theater, puppetry, and dance	1900–1949	lw	3rd
390	1	o-matsu, n.		Manufacturing and materials, Flora	1900–1949	lw	3rd
391	1	shubunkin, n.		Fauna	1900–1949	lw	2nd
392	1	narikin, n.		Finance, Other people	1900–1949	lw	3rd
393	2	ume, n.		Flora	1900–1949	lw	3rd
394	1	yoko-shiho-gatame, n.		Martial arts	1900–1949	lw	3rd
395	1	bunraku, n.		Theater, puppetry, and dance	1900–1949	lw	2nd
396	1	kaiseki, n.		Food and cooking	1900–1949	lw	3rd
397	1	kanji, n.		Writing and language	1900–1949	lw	2nd
398	1	Kurume, n.		Flora	1900–1949	lw	2nd
399	1	martial art, n.		Martial arts	1900–1949	cq	3rd
400	1	protoanemonin, n.		Chemistry and biochemistry	1900–1949	cq	3rd
401	2a	sayonara, int.		Greetings, interjections, and gestures	1900–1949	lw	3rd
402	1	sukiyaki, n.		Food and cooking	1900–1949	lw	2nd
403	1	tempura, n.		Food and cooking	1900–1949	lw	2nd
404	1	udon, n.		Food and cooking	1900–1949	lw	2nd
405	1	kendo, n.		Martial arts	1900–1949	lw	2nd
406	1	moribana, n.		Arrangement	1900–1949	lw	3rd
407	1	yugen, n.		Literature, Theater, puppetry, and dance, Values and concepts	1900–1949	lw	2nd
408	2	aucuba, n.		Biology	1900–1949	lw	2nd
409	1	ishikawaite, n.		Chemistry and biochemistry	1900–1949	hb	2nd
410	1	daisho, n.		Weaponry	1900–1949	lw	2nd
411	1	Ishihara, n.		Medicine	1900–1949	lw	2nd
412	1	oyama, n.		Theater, puppetry, and dance, Professionals and hobbyists	1900–1949	lw	3rd
413	2	Nippon, n. (and adj.)		Bookbinding, Writing and language	1900–1949	hb	3rd
414	1	Showa, n.		Historical periods	1900–1949	lw	3rd
415	1	sodoku, n.		Medicine	1900–1949	lw	2nd
416	1	ah so, int.		Greetings, interjections, and gestures	1900–1949	lw	3rd
417	1	Minseito, n.		Government, administration, and politics, Groups and organizations	1900–1949	lw	3rd
418	1	mondo, n.1		Religion	1900–1949	lw	3rd
419	2	Nabeshima, adj. and n.		Pottery and lacquerware	1900–1949	lw	3rd
420	1	Showa, adj.		Historical periods	1900–1949	lw	3rd
421	1	Akita, n.		Fauna	1900–1949	lw	3rd
422	2	kombucha, n.		Beverages	1900–1949	lw	3rd
423	1	Oranda, n.		Fauna	1900–1949	lw	3rd
424	2	pillow book, n.		Literature	1900–1949	cq	3rd
425	1	Shibayama, n.		Pottery and lacquerware	1900–1949	lw	2nd
426	2a	banzai, int.		Attributes and qualities, Greetings, interjections, and gestures	1900–1949	lw	2nd
427	1	Momoyama, adj.		Historical periods	1900–1949	lw	3rd
428	1	suiseki, n.		Arrangement	1900–1949	lw	2nd
429	1	oyakata, n.		Status and rank, Titles, forms of address and names, Martial arts	1900–1949	lw	3rd
430	1	haniwa, n.		Pottery and lacquerware	1900–1949	lw	2nd

431	1	juku, n.		Education	1900–1949	lw	3rd
432	1	thoughtography, n.		Parapsychology	1900–1949	cq	3rd
433	1	kesa-gatame, n.		Martial arts	1900–1949	lw	2nd
434	1	nagewaza, n.		Martial arts	1900–1949	lw	3rd
435	1	o-goshi, n.		Martial arts	1900–1949	lw	3rd
436	1	osaekomi-waza, n.		Martial arts	1900–1949	lw	3rd
437	1	sabi, n.		Religion, Values and concepts	1900–1949	lw	2nd
438	1	seoi nage, n.		Martial arts	1900–1949	lw	2nd
439	der	tryptophan, n.	tryptophanase n.	Chemistry and biochemistry	1900–1949	lw	2nd
440	1	waka, n.2		Literature	1900–1949	lw	2nd
441	1	kyudo, n.		Martial arts	1900–1949	lw	3rd
442	1	Mitsuda, n.		Medicine	1900–1949	lw	3rd
443	1	Kikuchi, n.		Physics	1900–1949	lw	2nd
444	2	mu, int.		Greetings, interjections, and gestures	1900–1949	lw	3rd
445	1	mushin, n.		Literature, Martial arts, Values and concepts	1900–1949	lw	3rd
446	1	nisei, adj.		Ethnicity and residency	1900–1949	lw	3rd
447	2	nisei, n.		Ethnicity and residency	1900–1949	lw	3rd
448	1	ofuro, n.		Interior design and furniture, Leisure activities	1900–1949	lw	3rd
449	2	Ryukyu,n.		Writing and language	1900–1949	lw	3rd
450	1	thought, n.	thought crime n.	Government, administration, and politics	1900–1949	cq	2nd
451	1	todorokite, n.		Geology and mineralogy	1900–1949	hb	3rd
452	1	wabi, n.		Religion, Values and concepts	1900–1949	lw	2nd
453	1	Hashimoto, n.		Medicine	1900–1949	lw	2nd
454	1	Nagami, n.		Food and cooking, Flora	1900–1949	hb	3rd
455	1	ryu, n.		Arrangement, Music, Painting, pictures, and printing, Theater, puppetry, and dance, Martial arts	1900–1949	lw	3rd
456	1	gagaku, n.		Music, Religion	1900–1949	lw	2nd
457	1	napa, n.2		Food and cooking, Flora	1900–1949	lw	3rd
458	1	kyu, n.		Martial arts	1900–1949	lw	2nd
459	1	mizutaki, n.		Food and cooking	1900–1949	lw	3rd
460	1	zaibatsu, n.		Business and working culture, Professionals and hobbyists	1900–1949	lw	2nd
461	1	senryu, n.		Literature	1900–1949	lw	2nd
462	1	shochu, n.		Beverages	1900–1949	lw	2nd
463	1	sumi-e, n.		Painting, pictures, and printing	1900–1949	lw	2nd
464	2	waka, n.2		Literature	1900–1949	lw	2nd
465	1	Yukawa, n.		Physics	1900–1949	lw	2nd
466	3	Nippon, n. (and adj.)		Flora	1900–1949	lw	3rd
467	1	basho, n.		Martial arts	1900–1949	lw	2nd
468	1	mabe, n.		Gemstones	1900–1949	lw	3rd
469	1	mono no aware, n.		Literature, Painting, pictures, and printing, Values and concepts	1900–1949	lw	3rd
470	1	oshi, n.		Martial arts	1900–1949	lw	3rd
471	1	oshi-dashi, n.		Martial arts	1900–1949	lw	3rd
472	1	oshi-taoshi, n.		Martial arts	1900–1949	lw	3rd
473	1	dan, n.5		Status and rank, Martial arts	1900–1949	lw	2nd
474	1	harai goshi, n.		Martial arts	1900–1949	lw	2nd
475	1	osae-waza, n.		Martial arts	1900–1949	lw	3rd
476	1	osotogari, n.		Martial arts	1900–1949	lw	3rd
477	1	sumi-gaeshi, n.		Martial arts	1900–1949	lw	2nd
478	1	tsukuri, n.		Martial arts	1900–1949	lw	2nd
479	1a	dojo, n.		Buildings and property, Martial arts	1900–1949	lw	2nd
480	1	johachidolite, n.		Chemistry and biochemistry	1900–1949	hb	2nd
481	1	Tojo, n.		Military and bushido, Groups and organizations, Professionals and hobbyists	1900–1949	lw	2nd
482	2	Yamato, n.		Values and concepts	1900–1949	lw	2nd
483	1	Yagi, n.		Electronics and technology	1900–1949	lw	2nd
484	1a	kamikaze, adj.		Military and bushido, Groups and organizations, Professionals and hobbyists, Seafaring, vehicles, and transportation	1900–1949	lw	3rd

485	2a	kamikaze, n.		Military and bushido, Groups and organizations, Seafaring, vehicles, and transportation	1900–1949	lw	3rd
486	2b	banzai, int.		Military and bushido	1900–1949	lw	2nd
487	1	comfort girl, n.		Other people	1900–1949	cq	3rd
488	1	honcho, n.		Status and rank	1900–1949	lw	3rd
489	2b	kamikaze, n.		Military and bushido, Professionals and hobbyists, Other people	1900–1949	lw	3rd
490	1	sansei, n.		Ethnicity and residency	1900–1949	lw	2nd
491	1	Tosa, n.2		Fauna	1900–1949	lw	2nd
492	1	Jomon, n.		Pottery and lacquerware, Historical periods	1900–1949	lw	2nd
493	1	koan, n.		Religion	1900–1949	lw	2nd
494	1	on, n.2		Values and concepts	1900–1949	lw	3rd
495	1	Kempeitai, n.		Military and bushido, Groups and organizations	1900–1949	lw	2nd
496	1	shibui, adj.		Values and concepts	1900–1949	lw	2nd
497	2	tatami, n.		Architecture	1900–1949	lw	2nd
498	1	tenko, n.		Military and bushido	1900–1949	lw	2nd
499	1	tonari gumi, n.		Groups and organizations	1900–1949	lw	2nd
500	comp	body, n.	body drop n.	Martial arts	1900–1949	cq	3rd
501	1	judogi, n.		Clothes, Martial arts	1900–1949	lw	3rd
502	1	one-worldism, n.		Government, administration, and politics, Values and concepts	1900–1949	cq	3rd
503	1	origami, n.		Paper folding and packaging	1900–1949	lw	3rd
504	1	oyabun, n.		Status and rank	1900–1949	lw	3rd
505	1	tachi, n.		Weaponry	1900–1949	lw	2nd
506	2	Nanga, adj.		Painting, pictures, and printing	1900–1949	lw	3rd
507	1	pachinko, n.		Games	1900–1949	lw	3rd
508	1	pan-pan, n.		Professionals and hobbyists, Other people	1900–1949	lw	3rd
509	1	bonsai, n.		Leisure activities, Flora	1950–1999	lw	2nd
510	1	jigotai, n.		Martial arts	1950–1999	lw	2nd
511	2	jōruri, n.		Theater, puppetry, and dance	1950–1999	lw	2nd
512	1	kobeite, n.		Chemistry and biochemistry	1950–1999	hb	2nd
513	1	kuzushi, n.		Martial arts	1950–1999	lw	2nd
514	1	ninjutsu, n.		Military and bushido, Martial arts	1950–1999	lw	3rd
515	1	reap, v.1		Martial arts	1950–1999	lw	3rd
516	1	tai-otoshi, n.		Martial arts	1950–1999	lw	2nd
517	1	tsugi ashi, n.		Martial arts	1950–1999	lw	2nd
518	1	wakame, n.		Food and cooking, Flora	1950–1999	lw	2nd
519	1	edamame, n.		Food and cooking	1950–1999	lw	3rd
520	1	manga, n.2		Comics and cartoons	1950–1999	lw	3rd
521	1	Mutsu, n.		Food and cooking, Flora	1950–1999	lw	3rd
522	2	honcho, n.		Status and rank	1950–1999	lw	3rd
523	1	hoochie, n.1		Buildings and property	1950–1999	lw	2nd
524	1	judoka, n.		Professionals and hobbyists, Martial arts	1950–1999	lw	3rd
525	2	kamikaze, adj.		Attributes and qualities	1950–1999	lw	3rd
526	1	Takayasu, n.		Medicine	1950–1999	lw	2nd
527	1	Zengakuren, n.		Government, administration, and politics	1950–1999	lw	2nd
528	1	Morita, n.2		Psychology	1950–1999	lw	3rd
529	1	ryotei, n.		Buildings and property, Food and cooking	1950–1999	lw	3rd
530	1	Sohyo, n.		Business and working culture, Government, administration, and politics	1950–1999	lw	2nd
531	1	aikido, n.		Martial arts	1950–1999	lw	3rd
532	1	ama, n.		Professionals and hobbyists	1950–1999	lw	2nd
533	1	kainic, adj.		Chemistry and biochemistry	1950–1999	hb	2nd
534	1	kata, n.		Martial arts	1950–1999	lw	2nd
535	1	Shihan, n.		Titles, forms of address and names, Martial arts	1950–1999	lw	2nd
536	1	shime-waza, n.		Martial arts	1950–1999	lw	2nd
537	1a	ude, n.	ude-garami n.	Martial arts	1950–1999	lw	2nd
538	1b	ude, n.	ude-gatame n.	Martial arts	1950–1999	lw	2nd
539	1	waza-ari, n.		Martial arts	1950–1999	lw	2nd
540	1	aikijutsu, n.		Martial arts	1950–1999	lw	3rd

541	1	karate, n.		Martial arts	1950–1999	lw	2nd
542	1	mingei, n.		Painting, pictures, and printing, Pottery and lacquerware, Other crafts	1950–1999	lw	3rd
543	1	tori, n.		Professionals and hobbyists, Martial arts	1950–1999	lw	2nd
544	1	mondo, n.2		Flora	1950–1999	lw	3rd
545	1	oshibori, n.		Food and cooking	1950–1999	lw	3rd
546	1	seiza, n.		Martial arts	1950–1999	lw	2nd
547	1	sosaku hanga, n.		Painting, pictures, and printing	1950–1999	lw	2nd
548	1	uke, n.2		Professionals and hobbyists, Martial arts	1950–1999	lw	2nd
549	1	ukemi, n.		Martial arts	1950–1999	lw	2nd
550	1	ippon, n.		Martial arts	1950–1999	lw	2nd
551	1	keirin, n.		Sports	1950–1999	lw	3rd
552	1	makuuchi, n.		Martial arts	1950–1999	lw	3rd
553	1	Minamata disease, n.		Medicine	1950–1999	hb	3rd
554	1	yakisoba, n.		Food and cooking	1950–1999	lw	3rd
555	1	andosol, n.		Geology and mineralogy	1950–1999	hb	3rd
556	1	emakimono, n.		Painting, pictures, and printing	1950–1999	lw	2nd
557	1	Sendai, n.		Biology, Medicine	1950–1999	lw	2nd
558	der	soba, n.	sobaya n.	Buildings and property, Food and cooking	1950–1999	lw	2nd
559	1	Soka Gakkai, n.		Religion	1950–1999	lw	2nd
560	1	chankonabe, n.		Food and cooking, Martial arts	1950–1999	lw	3rd
561	1	ikunolite, n.		Chemistry and biochemistry	1950–1999	hb	2nd
562	1	kokeshi, n.		Other crafts	1950–1999	lw	2nd
563	1	makiwara, n.		Martial arts	1950–1999	lw	3rd
564	1	ningyoite, n.		Geology and mineralogy	1950–1999	hb	3rd
565	1	shuto, n.		Martial arts	1950–1999	lw	2nd
566	1	skosh, n.		Attributes and qualities	1950–1999	lw	2nd
567	1	zendo, n.		Buildings and property, Religion	1950–1999	lw	2nd
568	1	mae-geri, n.		Martial arts	1950–1999	lw	3rd
569	2	ronin, n.		Education, Status and rank, Other people	1950–1999	lw	3rd
570	1	shibui, n.		Values and concepts	1950–1999	lw	2nd
571	2	busgirl, n.		Professionals and hobbyists	1950–1999	sl	3rd
572	1	comfort woman, n.		Other people	1950–1999	cq	3rd
573	1	Rashomon, n.		Attributes and qualities	1950–1999	lw	3rd
574	2b	sayonara, int.		Greetings, interjections, and gestures	1950–1999	lw	3rd
575	draft	teriyaki, n.		Food and cooking	1950–1999	lw	2nd
576	1	ibotenic, adj.		Chemistry and biochemistry	1950–1999	hb	2nd
577	2	juku, n.		Education	1950–1999	lw	3rd
578	1	ramen, n.		Food and cooking	1950–1999	lw	3rd
579	2	sayonara, adj.		Greetings, interjections, and gestures, Sports	1950–1999	lw	3rd
580	1	teriyaki, n.		Food and cooking	1950–1999	lw	2nd
581	1	yakitori, n.		Food and cooking	1950–1999	lw	2nd
582	1	dashi, n.		Food and cooking	1950–1999	lw	2nd
583	3	kamikaze, n.		Sports	1950–1999	lw	3rd
584	special uses	living, adj. and n.1	living treasure n.	Other people, Values and concepts	1950–1999	sl	3rd
585	1	sanpaku, n.		Alternative medicine	1950–1999	lw	2nd
586	1	Shotokan, n.		Martial arts	1950–1999	lw	2nd
587	1	sudoite, n.		Geology and mineralogy	1950–1999	hb	2nd
588	1	umami, n.		Food and cooking	1950–1999	lw	3rd
589	1	Wagyu, n.		Food and cooking, Fauna	1950–1999	lw	3rd
590	B	gaijin, adj.		Attributes and qualities, Ethnicity and residency	1950–1999	lw	2nd
591	1	ninja, n.		Professionals and hobbyists, Other people, Martial arts	1950–1999	lw	3rd
592	1	shunga, n.		Painting, pictures, and printing	1950–1999	lw	2nd
593	1	Suzuki, n.		Music, Education	1950–1999	lw	2nd
594	1	window, n.	window guidance n.	Finance	1950–1999	cq	3rd
595	1	yakuza, n.		Groups and organizations, Other people	1950–1999	lw	2nd
596	1	Yamaguchi-gumi, n.		Groups and organizations	1950–1999	lw	2nd

597	1	Godzilla, n.		Attributes and qualities, Other people	1950–1999	lw	3rd
598	1	gyoza, n.		Food and cooking	1950–1999	lw	3rd
599	1	keiretsu, n.		Business and working culture	1950–1999	lw	3rd
600	2	sensei, n.		Professionals and hobbyists, Titles, forms of address, and names, Martial arts	1950–1999	lw	3rd
601	1b	dojo, n.		Martial arts	1950–1999	lw	2nd
602	1b	kamikaze, adj.		Military and bushido	1950–1999	lw	3rd
603	der	karate, n.	karateka n.	Professionals and hobbyists, Martial arts	1950–1999	lw	2nd
604	1	ninja, adj.		Attributes and qualities	1950–1999	lw	3rd
605	1	ponzu, n.		Food and cooking	1950–1999	lw	3rd
606	1	Tokugawa, n.		Government, administration, and politics, Historical periods, Groups and organizations	1950–1999	lw	2nd
607	1	shiatsu, n.		Alternative medicine	1950–1999	lw	2nd
608	1	shunto, n.		Business and working culture	1950–1999	lw	2nd
609	1	sogo shosha, n.		Business and working culture	1950–1999	lw	2nd
610	1	suzuribako, n.		Pottery and lacquerware, Writing and language	1950–1999	lw	2nd
611	A	gaijin, n.		Ethnicity and residency	1950–1999	lw	2nd
612	1	Okazaki, n.		Biology	1950–1999	lw	3rd
613	1	rishitin, n.		Chemistry and biochemistry	1950–1999	hb	3rd
614	1	seitan, n.		Food and cooking	1950–1999	lw	3rd
615	1	Shinkansen, n.		Seafaring, vehicles and transportation	1950–1999	lw	2nd
616	1	zaikai, n.		Business and working culture, Finance	1950–1999	lw	2nd
617	1	itai-itai, n.		Medicine	1950–1999	lw	2nd
618	1	nikkei, n.1		Ethnicity and residency	1950–1999	lw	3rd
619	1	nunchaku, n.		Weaponry	1950–1999	lw	3rd
620	1	Red Army Faction, n.		Government, administration, and politics, Groups and organizations	1950–1999	cq	3rd
621	1	yusho, n.		Medicine	1950–1999	lw	2nd
622	1	hijiki, n.		Food and cooking, Flora	1950–1999	lw	3rd
623	1	kogai, n.		Environment	1950–1999	lw	2nd
624	1	minshuku, n.		Buildings and property, Travelling and tourism	1950–1999	lw	3rd
625	1	omiai, n.		Tradition	1950–1999	lw	3rd
626	1	panko, n.		Food and cooking	1950–1999	lw	3rd
627	1b	papasan, n.		Professionals and hobbyists, Status and rank	1950–1999	lw	3rd
628	1	Red Army, n.		Government, administration, and politics, Groups and organizations	1950–1999	cq	3rd
629	1	shabu-shabu, n.		Food and cooking	1950–1999	lw	2nd
630	1	shishi, n.		Pottery and lacquerware	1950–1999	lw	2nd
631	der	sushi, n.	sushiya n.	Buildings and property, Food and cooking	1950–1999	lw	2nd
632	1	teppan-yaki, n.		Food and cooking	1950–1999	lw	2nd
633	2c	kamikaze, n.		Other people	1950–1999	lw	3rd
634	1	meishi, n.		Business and working culture	1950–1999	lw	3rd
635	1	shokku, n.		Finance, Government, administration, and politics	1950–1999	lw	2nd
636	1	sokaiya, n.		Finance, Other people	1950–1999	lw	2nd
637	1	tataki, n.		Food and cooking	1950–1999	lw	3rd
638	1	toro, n.2		Food and cooking	1950–1999	lw	3rd
639	draft	futon, n.		Interior design and furniture	1950–1999	lw	2nd
640	2	ninja, n.		Attributes and qualities, Other people	1950–1999	lw	3rd
641	1	Tsukahara, n.		Sports	1950–1999	lw	2nd
642	initialism	O, n.1 OE	OL n.	Professionals and hobbyists	1950–1999	lw	3rd
643	comp	office, n.	office lady n.	Professionals and hobbyists	1950–1999	lw	3rd
644	1	sai, n.3		Weaponry	1950–1999	lw	2nd
645	1	sumotori, n.		Professionals and hobbyists, Martial arts	1950–1999	lw	2nd
646	1	surimi, n.		Food and cooking	1950–1999	lw	3rd
647	comp	mikoshi, n.		Religion	1950–1999	lw	3rd
648	1	Nikkei, n.2		Finance	1950–1999	lw	3rd

649	comp	quality, n. and adj.	quality circle n.	Business and working culture	1950–1999	cq	3rd
650	1	robata-yaki, n.		Food and cooking	1950–1999	lw	3rd
651	1	Shorin ryu, n.		Martial arts	1950–1999	lw	2nd
652	C2	money, n.	money politics n.	Government, administration, and politics	1950–1999	cq	3rd
653	2a	Nikkei, n.2		Finance	1950–1999	lw	3rd
654	1	reiki, n.		Alternative medicine, Values and concepts	1950–1999	lw	3rd
655	1	shishito, n.		Food and cooking, Flora	1950–1999	lw	3rd
656	1	tsutsumu, n.		Paper folding and packaging	1950–1999	lw	2nd
657	1	mechatronics, n.		Electronics and technology	1950–1999	cq	3rd
658	1	shosha, n.		Business and working culture	1950–1999	lw	2nd
659	1	kanban, n.		Manufacturing and materials	1950–1999	lw	2nd
660	2	kanban, n.		Manufacturing and materials	1950–1999	lw	2nd
661	1	karaoke, n.		Leisure activities	1950–1999	lw	2nd
662	1	tamari, n.		Food and cooking	1950–1999	lw	2nd
663	1	hamachi, n.		Food and cooking, Fauna	1950–1999	lw	3rd
664	1	shuriken, n.		Weaponry	1950–1999	lw	2nd
665	1	uncanny valley, n.		Electronics and technology	1950–1999	cq	3rd
666	1	washi, n.		Manufacturing and materials	1950–1999	lw	2nd
667	4	kamikaze, n.		Beverages	1950–1999	lw	3rd
668	1	mokume gane, n.		Other crafts	1950–1999	lw	3rd
669	1	nikkeijin, n.		Ethnicity and residency	1950–1999	lw	3rd
670	1	omakase, n.		Food and cooking	1950–1999	lw	3rd
671	2	papasan, n.		Interior design and furniture	1950–1999	lw	3rd
672	1	shojo, n.		Comics and cartoons	1950–1999	lw	3rd
673	1	okadaic, adj.		Chemistry and biochemistry	1950–1999	hb	3rd
674	2	hand roll, n.		Food and cooking	1950–1999	cq	3rd
675	3	Nikkei, n.2		Literature, Finance	1950–1999	lw	3rd
676	1	shonen, n.		Comics and cartoons	1950–1999	lw	3rd
677	comp	skin, n.	skin lady n.	Professionals and hobbyists	1950–1999	cq	3rd
678	1	softnomics, n.		Finance	1950–1999	lw	2nd
679	1	bukkake, n.		Food and cooking	1950–1999	lw	3rd
680	1	kawaii, adj.		Attributes and qualities, Values and concepts	1950–1999	lw	3rd
681	1	anime, n.3		Comics and cartoons	1950–1999	lw	3rd
682	1	kaizen, n.		Business and working culture, Values and concepts	1950–1999	lw	2nd
683	comp	makuuchi, n.		Martial arts	1950–1999	lw	3rd
684	1	shiso, n.		Food and cooking, Flora	1950–1999	lw	3rd
685	1	tokkin, n.		Finance	1950–1999	lw	2nd
686	2	kawaii, n.		Values and concepts	1950–1999	lw	3rd
687	1	mecha, n.		Comics and cartoons	1950–1999	lw	3rd
688	1	zaitech, n.		Finance	1950–1999	lw	2nd
689	1	izakaya, n.		Buildings and property, Beverages	1950–1999	lw	3rd
690	2b	Nikkei, n.2		Finance	1950–1999	lw	3rd
691	1	karoshi, n.		Business and working culture	1950–1999	lw	3rd
692	comp	functional, adj. and n.	functional food n.	Food and cooking	1950–1999	cq	3rd
693	1	hentai, n.		Comics and cartoons	1950–1999	lw	3rd
694	2	ginkgo, n.		Alternative medicine	1950–1999	lw	3rd
695	1	quantitative easing, n.		Finance	1950–1999	cq	3rd
696	1	replotment, n.		Buildings and property	1950–1999	sl	3rd
697	1	otaku, n.		Electronics and technology, Professionals and hobbyists	1950–1999	lw	3rd
698	1	cosplay, n.		Comics and cartoons, Leisure activities, Games	1950–1999	lw	3rd
699	comp	compensate, v.	compensated dating n.	Leisure activities	1950–1999	cq	Draft add.
700	1	enjo kosai, n.		Leisure activities	1950–1999	lw	3rd
701	1	emoji, n.		Electronics and technology	1950–1999	lw	3rd
702	2	cosplay, n.	cosplayer n.	Comics and cartoons, Leisure activities, Games, Professionals and hobbyists	1950–1999	lw	3rd
703	1	hikikomori, n.		Psychology, Other people	1950–1999	lw	3rd
704	1	keitai, n.		Electronics and technology	1950–1999	lw	3rd
705	der	dairi, n.	dairi-sama n.	Titles, forms of address and names	N/A	lw	2nd

Appendix 4: Frequencies of the lexeme-sense pairs from the *OED* searched in the AmE, BrE, SgE, PhE, MyE, and HKE sections of GloWbE. The table includes only those LS pairs that have relevant tokens.

		All varieties		AmE		BrE		SgE		PhE		MyE		HKE	
Headword	**S#**	**Raw**	**Base**	**Raw**	**Base**	**Raw**	**Base**	**Raw**	**Base**	**Raw**	**Base**	**Raw**	**Base**	**Raw**	**Base**
bonze n.	2.1	31	4	0	0	3	2	0	0	28	2	0	0	0	0
kami n.1	3.1	24	15	7	6	3	3	10	4	0	0	0	0	4	2
katana n.	5.1	263	166	116	76	46	33	23	13	53	33	16	8	9	3
mat n.1	6.8	2	2	1	1	0	0	1	1	0	0	0	0	0	0
Nippon n. (and adj.)	7.1	2	2	1	1	0	0	0	0	0	0	0	0	1	1
tatami n.	9.1	38	28	7	7	21	13	8	6	0	0	1	1	1	1
miso n.	11.1	386	208	88	58	66	40	121	50	29	11	23	14	59	35
shogun n.	12.1	145	75	34	21	33	17	44	15	15	10	6	5	13	7
bento n.	13.1	378	188	54	38	41	23	234	99	24	12	14	10	11	6
mochi n.1	17.1	167	90	14	12	4	2	32	22	22	10	31	15	64	29
norimon n.	18.1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
samisen n.	19.1	31	15	5	5	5	5	0	0	15	2	4	2	2	1
tabi n.	20.1	17	13	9	5	5	5	1	1	0	0	0	0	2	2
inro n.	21.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
menuki n.	22.1	4	2	0	0	1	1	3	1	0	0	0	0	0	0
mikan n.	23.1	10	3	0	0	0	0	1	1	0	0	5	1	4	1
tai n.1	24.1	12	10	3	2	1	1	8	7	0	0	0	0	0	0
moxa n.	27.1	10	4	0	0	5	2	0	0	5	2	0	0	0	0
acupuncture n.	28.1	2615	1085	941	393	991	418	103	54	68	40	154	58	358	122
soy n.1	29.1	132	108	27	22	22	19	39	29	16	13	6	6	22	19
akoya n.	30.1	12	10	3	2	2	2	1	1	0	0	3	3	3	2
chanoyu n.	31.1	26	9	4	3	20	4	0	0	1	1	0	0	1	1
hinoki n.	33.1	12	9	5	3	1	1	3	3	3	2	0	0	0	0
Hizen n.	34.1	9	1	0	0	0	0	0	0	0	0	9	1	0	0
Jōdo n.	35.1a	4	3	1	1	0	0	0	0	0	0	2	1	1	1
kaki n.	36.1	2	1	0	0	0	0	2	1	0	0	0	0	0	0
kana n.	38.1	82	29	30	13	5	4	0	0	43	9	3	2	1	1
katakana n.	39.1	178	92	55	32	21	17	24	15	64	17	9	8	5	3
katsuo n.	40.1	8	4	0	0	2	1	6	3	0	0	0	0	0	0
kiri n.	42.1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
kirin n.	43.1	16	11	4	3	1	1	7	4	4	3	0	0	0	0
koi n.	44.1	681	227	54	42	135	58	37	31	36	13	396	69	23	14
koi-cha n.	45.1	4	4	1	1	0	0	2	2	1	1	0	0	0	0
koku n.	46.1a	1	1	0	0	1	1	0	0	0	0	0	0	0	0
matsuri n.	48.1	9	7	1	1	3	2	4	3	1	1	0	0	0	0
Mikado n.	49.1	19	10	13	5	4	3	1	1	0	0	0	0	1	1
mikoshi n.	50.1	9	5	0	0	0	0	6	3	2	1	0	0	1	1
nembutsu n.	53.1	4	2	3	1	0	0	0	0	1	1	0	0	0	0
Ritsu n.	54.1	3	3	0	0	0	0	0	0	0	0	1	1	2	2
samurai n.	55.1a	1158	672	416	243	382	227	112	74	117	68	48	31	83	29
satori n.	56.1	78	42	47	24	24	12	3	3	2	1	1	1	1	1
Shingon n.	60.1	14	11	0	0	5	5	7	4	0	0	1	1	1	1
Shinshū n.	61.1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
Shinto n.	62.1a	119	60	56	29	27	16	9	4	7	3	5	3	15	5
Shinto n.	63.1b	215	137	76	51	37	31	41	28	16	12	6	5	39	10
shoyu n.	64.1	56	42	8	6	16	14	24	18	6	3	0	0	2	1
Tendai n.	67.1	8	6	5	3	2	2	1	1	0	0	0	0	0	0
tokonoma n.	68.1	6	5	2	2	2	2	2	1	0	0	0	0	0	0
torii n.	69.1	33	18	4	3	9	5	9	1	3	2	2	2	6	5
tsubo n.	70.1	7	3	1	1	0	0	2	1	0	0	0	0	4	1
urushi n.	71.1	38	5	0	0	25	1	2	2	1	1	0	0	10	1
wakizashi n.	72.1	14	10	6	5	2	2	0	0	2	2	4	1	0	0
zazen n.	73.1	56	37	21	20	10	7	3	2	5	3	2	1	15	4
Zen n.	74.1	2059	1001	1098	502	466	253	113	76	105	48	73	39	204	83
adzuki n.	75.1	48	23	7	4	14	8	23	8	2	1	1	1	1	1
ginkgo n.	76.1	157	91	30	17	23	20	73	28	10	7	5	4	16	15
Nandina n.	77.1	2	1	2	1	0	0	0	0	0	0	0	0	0	0
miso n.	78. comp	165	127	36	26	28	22	45	36	23	20	16	10	17	13
koto n.	79.1	17	9	2	2	4	4	0	0	11	3	0	0	0	0
saké n.2	80.1	599	330	105	71	104	82	155	79	10	9	29	23	196	66
Ryukyu adj.	81.1	15	11	7	5	1	1	2	1	1	1	2	2	2	1
Ainu n.	82.1	35	20	26	12	3	3	1	1	1	1	1	1	3	2

ri n.	83.1	2	2	0	0	0	0	1	1	0	0	0	0	1	1
aucuba n.	84.1	2	2	2	2	0	0	0	0	0	0	0	0	0	0
gobo n.1	85.1	4	2	0	0	2	1	2	1	0	0	0	0	0	0
hiragana n.	86.1	220	77	48	26	11	8	31	15	119	22	8	5	3	1
noshi n.	88.1	1	1	0	0	0	0	0	0	0	0	1	1	0	0
ume n.	90.1	5	5	1	1	0	0	3	3	0	0	1	1	0	0
umeboshi n.	91.1	29	22	21	16	3	3	4	2	0	0	0	0	1	1
yukata n.	92.1	78	39	5	3	8	1	46	22	3	2	7	3	9	8
zori n.	93.1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
moxa n.	94.2	4	3	0	0	3	2	0	0	0	0	1	1	0	0
Shinto n.	97.2	13	13	8	8	3	3	1	1	0	0	0	0	1	1
Rinzai n.	99.1	22	18	16	13	1	1	2	2	2	1	0	0	1	1
Ainu adj.	100.3	31	24	13	9	8	8	6	5	3	1	0	0	1	1
daimio n.	101.1	53	20	11	6	7	4	6	3	3	3	1	1	25	3
go n.2	102.1	51	30	17	13	27	11	6	5	0	0	0	0	1	1
renga n.	105.1	14	6	1	1	1	1	0	0	12	4	0	0	0	0
hara-kiri n.	107.1	46	42	11	11	17	16	2	2	8	6	8	7	0	0
ronin n.	110.1	51	33	18	15	7	7	13	6	12	4	0	0	1	1
shogi n.	111.1	11	10	5	5	4	3	2	2	0	0	0	0	0	0
tycoon n.	116.2	1861	1423	370	298	729	603	117	88	104	82	207	150	334	202
hibachi n.	121.1	4	4	0	0	1	1	0	0	0	0	0	0	3	3
sayonara int.	122.1	13	13	3	3	3	3	5	5	1	1	0	0	1	1
yashiki n.	123.1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
sasanqua n.	126.1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
kyogen n.	135.1	11	5	0	0	3	2	2	1	4	1	0	0	2	1
Noh n.	137.1	52	31	11	9	15	13	6	3	11	4	0	0	9	2
oiran n.	138.1	13	7	1	1	0	0	12	6	0	0	0	0	0	0
reign n.	139. comp	1	1	0	0	0	0	0	0	0	0	1	1	0	0
seppuku n.	140.1	50	42	33	29	9	7	0	0	5	3	3	3	0	0
obi n.1	144.1	39	25	1	1	10	6	5	4	5	2	7	6	11	6
sayonara n.	146.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
kamidana n.	147.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
Meiji n.	148.1	14	9	0	0	8	5	0	0	2	2	1	1	3	1
fundoshi n.	149.1	4	2	0	0	0	0	0	0	0	0	3	1	1	1
jinricksha | jinrikisha n.	150.1	20	5	18	3	1	1	0	0	0	0	0	0	1	1
miko n.	152.1	7	5	4	3	0	0	1	1	2	1	0	0	0	0
mirin n.	153.1	62	47	18	12	10	9	20	16	2	2	9	6	3	2
sencha n.	154.1	22	10	6	3	6	3	2	2	0	0	0	0	8	2
sensei n.	155.1	291	130	159	49	30	20	44	27	32	16	23	15	3	3
yen n.1	156.1	3680	1561	741	363	1009	415	934	362	253	97	408	165	335	159
aikuchi n.	157.1	2	1	0	0	0	0	0	0	0	0	2	1	0	0
heimin n.	159.1	2	2	0	0	1	1	0	0	0	0	0	0	1	1
Imari n.	160.1	2	2	0	0	1	1	0	0	1	1	0	0	0	0
ju-jitsu n.	161.1	375	222	201	114	86	53	52	30	22	14	6	5	8	6
raku n.	162.1	5	5	2	2	2	2	0	0	1	1	0	0	0	0
sennin n.	164.1	2	2	0	0	0	0	0	0	2	2	0	0	0	0
yokan n.	167.1	5	1	0	0	0	0	5	1	0	0	0	0	0	0
futon n.	169.1	52	43	17	15	6	5	19	16	0	0	3	3	7	4
Genro n.	170.1	1	1	0	0	0	0	0	0	1	1	0	0	0	0
geta n.	171.1	13	10	0	0	6	3	5	5	1	1	0	0	1	1
kotatsu n.	173.1	6	4	1	1	1	1	0	0	0	0	0	0	4	2
magatama n.	174.1	2	2	2	2	0	0	0	0	0	0	0	0	0	0
monogatari n.	175.1	1	1	0	0	0	0	0	0	1	1	0	0	0	0
netsuke n.	177.1	12	12	2	2	9	9	0	0	0	0	0	0	1	1
Nichiren n.	178.1	33	19	5	5	6	5	1	1	2	2	13	4	6	2
Tokugawa adj.	182.1	120	67	28	22	30	17	17	14	11	6	8	6	26	2
haori n.	184.1	7	7	1	1	1	1	2	2	2	2	0	0	1	1
matsutake n.	185.1	8	8	1	1	2	2	2	2	0	0	0	0	3	3
pillow word n.	186.1	3	3	0	0	0	0	0	0	3	3	0	0	0	0
shiitake n.	187.1	150	103	53	33	21	18	24	17	15	11	16	10	21	14
Shin n.3	188.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
tanka n.2	189.1	24	13	4	4	0	0	1	1	16	6	0	0	3	2
koji n.	191.1	13	6	0	0	2	1	10	4	0	0	0	0	1	1
san n.3	192.1	621	304	192	96	88	67	190	69	38	21	86	37	27	14
ukiyo-e n.	195.1	43	20	6	5	21	8	8	3	2	1	1	1	5	2
Yamato n.	196.1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
Ainu n.	197.2	7	6	4	3	1	1	0	0	0	0	1	1	1	1
fusuma n.	199.1	4	3	0	0	1	1	3	2	0	0	0	0	0	0
happi n.	200.1	4	4	1	1	1	1	0	0	0	0	0	0	2	2
maguro n.	204.1	20	15	2	2	1	1	7	5	6	3	1	1	3	3

makimono n.	205.1	3	2	2	1	0	0	0	0	1	1	0	0	0	0
okimono n.	210.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
sashimi n.	211.1	687	448	82	65	67	61	194	122	60	38	115	56	169	106
shibuichi n.	212.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
soy n.1	214.2	1094	663	733	436	202	139	50	31	25	16	26	20	58	21
sumo n.	215.1	314	1	120	65	110	79	23	13	10	10	35	17	16	9
Tanabata n.	216.1	15	9	4	4	0	0	8	3	0	0	2	1	1	1
tofu n.	218.1	2115	1153	753	402	288	188	477	232	112	66	222	122	263	143
matcha n.	219.1	202	101	54	16	34	18	47	32	24	10	16	9	27	16
Seto n.	220.1	2	1	2	1	0	0	0	0	0	0	0	0	0	0
ken n.4	222.1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
Nara, adj.	223.1	13	12	1	1	0	0	5	5	3	2	0	0	4	4
Satsuma n.	225.2	83	69	17	11	65	57	0	0	0	0	1	1	0	0
hinin n.	227.1	3	1	0	0	3	1	0	0	0	0	0	0	0	0
kagura n.	229.1	6	2	1	1	0	0	0	0	5	1	0	0	0	0
kombu n.	231.1	78	36	22	9	6	3	40	16	3	2	4	4	3	2
koniak |koniaku n.	232.1	95	42	1	1	4	3	34	20	1	1	37	13	18	4
sakura n.	237.1	122	57	5	4	9	7	51	27	13	8	40	7	4	4
Shijō n.	238.1	2	2	0	0	1	1	0	0	1	1	0	0	0	0
Kuroshiwo n.	239.1	13	9	9	5	1	1	3	3	0	0	0	0	0	0
romaji n.	240.1	52	24	13	8	1	1	6	6	30	7	1	1	1	1
tansu n.	241.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
tanto n.2	242.1	12	6	4	2	1	1	0	0	2	1	5	2	0	0
tsukemono n.	243.1	4	3	1	1	0	0	2	1	0	0	0	0	1	1
kimono n.	245.1	414	289	98	69	157	108	59	45	21	15	30	18	49	34
shikimic adj.	247.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
shoji n.	248.2	13	11	2	2	7	5	1	1	0	0	1	1	2	2
tea man n.	249.2b	1	1	0	0	0	0	1	1	0	0	0	0	0	0
head n.1	250. comp	1	1	0	0	1	1	0	0	0	0	0	0	0	0
nembutsu n.	251.2	6	3	3	1	0	0	0	0	2	1	1	1	0	0
awabi n.	254.1	2	2	0	0	1	1	0	0	0	0	0	0	1	1
judo n.	256.1	1124	627	223	140	698	379	58	36	40	26	74	28	31	18
Maru n.	258.1	4	2	3	1	0	0	0	0	1	1	0	0	0	0
nageire n.	259.1	3	1	0	0	0	0	0	0	0	0	3	1	0	0
reishi n.	261.1	13	11	5	5	5	4	0	0	1	1	0	0	2	1
shogun n.	263.2	3	3	1	1	1	1	0	0	0	0	1	1	0	0
tsuba n.	264.1	10	5	2	2	1	1	6	1	0	0	1	1	0	0
zabuton n.	265.1	6	5	3	3	0	0	3	2	0	0	0	0	0	0
Kakiemon n.	269.1	3	2	0	0	3	2	0	0	0	0	0	0	0	0
mitsuba n.	271.1	3	3	0	0	2	2	1	1	0	0	0	0	0	0
suimono n.	272.1	2	2	0	0	0	0	2	2	0	0	0	0	0	0
thought control n.	273.2	76	64	35	31	27	24	1	1	2	2	5	5	6	1
geisha n.	274.1	375	186	84	55	72	51	76	32	30	14	99	25	14	9
hanami n.	275.1	59	18	4	3	1	1	48	11	3	2	3	1	0	0
katsuo n.	277.der	8	7	3	2	1	1	4	4	0	0	0	0	0	0
maiko n.	278.1	28	10	1	1	0	0	24	6	2	2	1	1	0	0
sika n.1	282.1	10	9	0	0	6	6	2	2	0	0	0	0	2	1
soroban n.	283.1	4	2	1	1	3	1	0	0	0	0	0	0	0	0
kami-sama n.2	284.1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
nashi n.	286.1	4	4	0	0	1	1	2	2	0	0	0	0	1	1
nori n.	287.1	221	118	58	39	31	14	81	39	35	16	3	3	13	7
banzai int.	288.1	5	5	0	0	2	2	2	2	1	1	0	0	0	0
Heian adj.	289.1	47	33	5	5	4	4	16	10	15	8	1	1	6	5
ki n.2	290.1	113	37	30	16	45	14	25	3	6	3	0	0	7	1
kudzu n.	291.1	118	72	89	57	12	8	4	2	0	0	6	1	7	4
shakuhachi n.	292.1	28	13	3	2	11	8	0	0	13	2	1	1	0	0
Soto n.	294.1	34	23	19	10	5	5	1	1	2	1	3	3	4	3
sushi n.	295.1	3376	1650	1135	563	601	425	761	273	233	89	298	134	348	166
janken n. (and int.)	296.1	6	5	0	0	1	1	5	4	0	0	0	0	0	0
plum n. and adj.2	297. comp	1	1	1	1	0	0	0	0	0	0	0	0	0	0
yokozuna n.	299.1	19	4	0	0	0	0	2	1	0	0	2	1	15	2
baren n.	300.1	5	3	3	2	2	1	0	0	0	0	0	0	0	0
habu n.	301.1	2	2	2	2	0	0	0	0	0	0	0	0	0	0
hakama n. 1859	303.1	25	9	5	3	0	0	16	4	0	0	0	0	4	2
kamikaze n.	304.1	3	3	1	1	0	0	0	0	2	2	0	0	0	0
onsen n.	306.1	314	99	28	14	23	14	120	34	3	3	91	17	49	17
soba n.	307.1	193	106	28	25	15	10	99	32	6	6	23	16	22	17
Eta |eta n.3	309.1	17	5	13	4	4	1	0	0	0	0	0	0	0	0
tsunami n.	310.1	6318	3659	2506	1406	2092	1305	486	249	358	207	534	297	342	195

bushido n.	312.1	61	48	29	22	23	17	4	4	3	3	1	1	1	1
meiji adj.	315.2	257	124	42	31	85	33	37	28	37	16	12	8	44	8
pink n.5 and adj.2	316.B7	9	7	3	2	3	3	1	1	0	0	0	0	2	1
Bon n.	317.1	53	26	18	7	2	2	11	9	10	4	11	3	1	1
haiku n.	318.1	373	243	197	130	90	66	16	13	44	13	5	5	21	16
Kabuki n.	319.1	145	93	25	22	32	28	21	16	42	18	7	4	18	5
natto n.	320.1	72	43	47	29	9	5	15	8	0	0	0	0	1	1
omochi n.	321.1	1	1	0	0	0	0	0	0	0	0	1	1	0	0
pivot word n.	322.2	6	2	1	1	0	0	5	1	0	0	0	0	0	0
Kobe n.	324.1	111	79	56	34	21	17	14	11	8	8	4	4	8	5
Shiga n.	325.1	10	7	10	7	0	0	0	0	0	0	0	0	0	0
ikebana n.	326.1	37	24	10	7	7	7	5	3	9	5	6	2	0	0
onnagata n.	327.1	2	2	1	1	0	0	0	0	1	1	0	0	0	0
sayonara adj.	329.1	4	4	0	0	0	0	1	1	0	0	0	0	3	3
Kamakura n.	330.1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
kombucha n.	331.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
nembutsu, adj.	332.1	5	2	0	0	0	0	0	0	0	0	5	2	0	0
obento n.	333.1	8	6	1	1	1	1	2	2	2	1	2	1	0	0
goblin n.1	336. comp	5	3	3	1	2	2	0	0	0	0	0	0	0	0
wasabi n.	338.1	385	270	78	60	78	64	116	73	35	20	44	26	34	27
mama-san n.	342.1	22	17	4	2	3	3	6	4	3	3	3	3	3	2
rotenone n.	344.1	7	5	2	2	4	2	0	0	0	0	0	0	1	1
Shinto n.	345.1c	12	9	2	1	6	4	0	0	3	3	0	0	1	1
budo n.	346.1	37	11	28	5	3	2	0	0	5	3	1	1	0	0
maitake n.	347.1	14	10	8	5	0	0	1	1	0	0	1	1	4	3
samurai n.	349.1b	60	50	13	9	29	25	3	3	4	3	9	8	2	2
medaka n.	351.1	14	7	7	4	7	3	0	0	0	0	0	0	0	0
Mikado n.	352.3	3	1	0	0	0	0	3	1	0	0	0	0	0	0
ura-nage n.	360.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
rikishi n.	361.1	14	2	0	0	0	0	1	1	0	0	0	0	13	1
roshi n.	364.1	107	53	76	41	5	5	0	0	0	0	4	2	22	5
randori n.	368.1	12	8	3	3	7	3	1	1	0	0	1	1	0	0
sayonara n.	369.2	1	1	0	0	1	1	0	0	0	0	0	0	0	0
belt n.1	372.2c	167	98	60	32	54	38	38	13	7	7	4	4	4	4
black belt n.	373.3	65	53	21	14	26	21	9	9	6	6	3	3	0	0
mizuna n.	376.1	20	14	2	1	8	8	8	3	1	1	1	1	0	0
shosagoto n.	378.1	1	1	0	0	0	0	0	0	1	1	0	0	0	0
sumi n.	379.1	19	7	14	5	0	0	4	1	1	1	0	0	0	0
mu n.3	381.1	6	3	5	2	0	0	0	0	1	1	0	0	0	0
ozeki n.	382.1	2	2	0	0	1	1	0	0	0	0	0	0	1	1
shodan n.	383.1	16	10	3	3	8	3	0	0	0	0	3	2	2	2
yondan n.	384.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
maki zushi n.	385.1	14	8	3	3	0	0	8	4	0	0	3	1	0	0
nightingale floor n.	386.1	5	4	2	1	0	0	0	0	3	3	0	0	0	0
ryokan n.	387.1	315	71	21	6	24	10	250	44	5	2	3	2	12	7
shubunkin n.	391.1	2	2	0	0	2	2	0	0	0	0	0	0	0	0
ume n.	393.2	4	4	1	1	1	1	0	0	0	0	0	0	2	2
bunraku n.	395.1	17	11	1	1	4	3	2	2	4	3	0	0	6	2
kaiseki n.	396.1	79	46	19	10	8	6	31	18	1	1	3	3	17	8
kanji n.	397.1	610	230	265	103	61	43	98	35	163	30	15	11	8	8
martial art n.	399.1	4342	2453	1366	870	1255	772	690	279	296	148	338	187	397	197
sayonara int.	401.2a	25	23	6	6	7	7	4	4	4	2	4	4	0	0
sukiyaki n.	402.1	47	31	3	2	2	2	20	11	9	6	6	4	7	6
tempura n.	403.1	400	258	43	35	77	56	116	53	52	38	56	37	48	39
udon n.	404.1	192	107	12	11	27	19	66	38	4	4	18	14	65	21
kendo n.	405.1	215	53	137	16	24	13	12	10	8	4	4	4	30	6
moribana n.	406.1	1	1	0	0	0	0	0	0	0	0	1	1	0	0
yugen n.	407.1	3	2	1	1	2	1	0	0	0	0	0	0	0	0
Ishihara n.	411.1	17	16	1	1	16	15	0	0	0	0	0	0	0	0
Showa n.	414.1	4	3	0	0	0	0	1	1	2	1	1	1	0	0
ah so, int.	416.1	6	6	4	4	2	2	0	0	0	0	0	0	0	0
Showa adj.	420.1	15	12	3	3	1	1	7	4	1	1	3	3	0	0
Akita n.	421.1	101	39	15	15	74	16	2	2	7	3	2	2	1	1
kombucha n.	422.2	117	65	101	54	9	5	3	3	0	0	3	2	1	1
Oranda n.	423.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
pillow book n.	424.2	16	2	1	1	15	1	0	0	0	0	0	0	0	0
Momoyama adj.	427.1	6	4	0	0	3	1	1	1	2	2	0	0	0	0
oyakata n.	429.1	3	1	0	0	0	0	0	0	0	0	0	0	3	1
haniwa n.	430.1	1	1	0	0	0	0	0	0	1	1	0	0	0	0
thoughtography n.	432.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0

kesa-gatame n.	433.1	2	1	2	1	0	0	0	0	0	0	0	0	0	0
nagewaza n.	434.1	7	2	0	0	7	2	0	0	0	0	0	0	0	0
sabi n.	437.1	4	3	1	1	0	0	3	2	0	0	0	0	0	0
waka n.2	440.1	7	3	0	0	0	0	1	1	6	2	0	0	0	0
kyudo n.	441.1	2	2	1	1	0	0	1	1	0	0	0	0	0	0
mu int.	444.2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
mushin n.	445.1	5	3	4	2	1	1	0	0	0	0	0	0	0	0
nisei adj.	446.1	38	11	37	10	0	0	1	1	0	0	0	0	0	0
nisei n.	447.2	41	18	36	14	3	3	0	0	2	1	0	0	0	0
ofuro n.	448.1	5	5	1	1	0	0	2	2	0	0	2	2	0	0
thought n.	450. comp	219	187	114	93	101	90	0	0	2	2	1	1	1	1
wabi n.	452.1	9	3	6	1	0	0	3	2	0	0	0	0	0	0
Hashimoto n.	453.1	185	78	144	66	39	10	2	2	0	0	0	0	0	0
ryu n.	455.1	22	4	17	2	3	1	0	0	0	0	2	1	0	0
gagaku n.	456.1	7	7	1	1	3	3	1	1	1	1	0	0	1	1
napa n.2	457.1	18	13	8	6	1	1	2	2	6	3	0	0	1	1
kyu n.	458.1	28	14	6	5	17	5	1	1	0	0	4	3	0	0
zaibatsu n.	460.1	20	14	5	4	6	3	5	4	3	2	1	1	0	0
senryu n.	461.1	4	3	3	2	1	1	0	0	0	0	0	0	0	0
shochu n.	462.1	38	24	4	4	11	6	2	2	4	1	7	3	10	8
sumi-e n.	463.1	24	8	18	3	3	3	1	1	2	1	0	0	0	0
waka n.2	464.2	7	3	0	0	0	0	1	1	6	2	0	0	0	0
Yukawa n.	465.1	13	6	3	2	4	3	0	0	0	0	0	0	6	1
basho n.	467.1	7	3	0	0	6	2	0	0	0	0	0	0	1	1
mono no aware n.	469.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
dan n.5	473.1	13	9	1	1	10	6	2	2	0	0	0	0	0	0
dojo n.	479.1a	24	11	7	5	17	6	0	0	0	0	0	0	0	0
Tojo n.	481.1	3	2	3	2	0	0	0	0	0	0	0	0	0	0
Yagi n.	483.1	23	15	2	2	6	3	1	1	2	2	0	0	12	7
kamikaze adj.	484.1a	80	58	40	29	33	25	0	0	6	3	0	0	1	1
kamikaze n.	485.2a	6	6	2	2	4	4	0	0	0	0	0	0	0	0
banzai int.	486.2b	14	11	8	5	6	6	0	0	0	0	0	0	0	0
comfort girl n.	487.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
honcho n.	488.1	223	198	120	100	65	62	7	6	19	18	9	9	3	3
kamikaze n.	489.2b	37	32	18	17	17	13	1	1	1	1	0	0	0	0
sansei n.	490.1	9	4	9	4	0	0	0	0	0	0	0	0	0	0
Tosa n.2	491.1	9	8	0	0	7	6	1	1	0	0	0	0	1	1
Jomon n.	492.1	44	13	3	3	31	6	8	2	1	1	1	1	0	0
koan n.	493.1	217	135	129	88	37	24	5	5	18	4	7	6	21	8
Kempeitai n.	495.1	39	21	2	2	3	2	7	6	0	0	2	2	25	9
tatami n.	497.2	90	67	23	18	11	9	31	24	1	1	14	8	10	7
body n.	500. comp	10	9	6	5	4	4	0	0	0	0	0	0	0	0
judogi n.	501.1	4	3	1	1	3	2	0	0	0	0	0	0	0	0
one-worldism n.	502.1	4	4	0	0	4	4	0	0	0	0	0	0	0	0
origami n.	503.1	570	375	213	145	239	148	43	28	21	21	37	18	17	15
oyabun n.	504.1	4	3	2	1	1	1	1	1	0	0	0	0	0	0
tachi n.	505.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
pachinko n.	507.1	70	40	31	13	18	9	6	6	8	7	2	2	5	3
bonsai n.	509.1	566	180	82	48	115	69	24	17	292	17	35	13	18	16
jōruri n.	511.2	3	3	0	0	0	0	1	1	2	2	0	0	0	0
kuzushi n.	513.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
ninjutsu n.	514.1	121	56	39	21	59	20	7	4	9	9	7	2	0	0
reap v.1	515.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
wakame n.	518.1	45	21	11	8	10	4	12	5	2	2	10	2	0	0
edamame n.	519.1	134	105	56	46	24	22	27	20	3	2	21	12	3	3
manga n.2	520.1	3250	1853	1472	859	268	189	774	423	439	201	230	150	67	31
Mutsu n.	521.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
honcho n.	522.2	356	328	112	105	160	145	28	27	16	15	33	30	7	6
hoochie n.1	523.1	9	7	6	6	3	1	0	0	0	0	0	0	0	0
judoka n.	524.1	80	61	9	8	62	45	3	3	2	2	4	3	0	0
kamikaze adj.	525.2	112	104	37	35	59	56	1	1	1	1	7	4	7	7
Takayasu n.	526.1	2	2	1	1	1	1	0	0	0	0	0	0	0	0
Zengakuren n.	527.1	2	1	0	0	2	1	0	0	0	0	0	0	0	0
ryotei n.	529.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
aikido n.	531.1	306	132	203	63	70	48	6	6	7	5	16	9	4	1
ama n.	532.1	6	3	0	0	5	2	0	0	0	0	0	0	1	1
Shihan n.	535.1	25	6	23	4	0	0	0	0	0	0	2	2	0	0
waza-ari n.	539.1	4	4	1	1	3	3	0	0	0	0	0	0	0	0
aikijutsu n.	540.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
karate n.	541.1	1181	787	379	298	600	349	46	32	53	42	81	48	22	18

mingei n.	542.1	3	2	2	1	1	1	0	0	0	0	0	0	0	0
tori n.	543.1	3	2	1	1	2	1	0	0	0	0	0	0	0	0
mondo n.2	544.1	3	3	1	1	1	1	1	1	0	0	0	0	0	0
oshibori n.	545.1	3	1	2	0	0	0	1	1	0	0	0	0	0	0
seiza n.	546.1	11	9	6	4	2	2	2	2	0	0	1	1	0	0
uke n.2	548.1	29	14	23	10	5	3	0	0	0	0	1	1	0	0
ukemi n.	549.1	36	11	32	9	1	1	0	0	0	0	3	1	0	0
ippon n.	550.1	20	13	1	1	17	11	0	0	0	0	0	0	2	1
keirin n.	551.1	143	88	0	0	111	71	3	1	0	0	24	14	5	2
Minamata disease n.	553.1	11	3	0	0	1	1	0	0	1	1	0	0	9	1
yakisoba n.	554.1	19	10	3	3	5	3	9	2	0	0	1	1	1	1
Sendai n.	557.1	10	4	5	2	5	2	0	0	0	0	0	0	0	0
Soka Gakkai n.	559.1	38	21	6	6	9	5	1	1	2	2	6	1	14	6
kokeshi n.	562.1	12	7	4	3	4	2	1	1	0	0	3	1	0	0
makiwara n.	563.1	5	2	0	0	1	1	4	1	0	0	0	0	0	0
skosh n.	566.1	9	9	8	8	1	1	0	0	0	0	0	0	0	0
zendo n.	567.1	15	13	12	7	2	2	1	1	0	0	0	0	0	0
mae-geri n.	568.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
comfort woman n.	572.1	64	47	26	17	4	4	3	3	24	16	3	3	4	4
Rashomon n.	573.1	12	8	10	6	1	1	1	1	0	0	0	0	0	0
sayonara int.	574.2b	33	26	21	16	4	4	0	0	1	1	7	5	0	0
teriyaki n.	575. draft	87	62	30	24	17	12	20	11	5	4	11	7	4	4
juku n.	577.2	7	2	0	0	5	1	2	1	0	0	0	0	0	0
ramen n.	578.1	1410	728	593	371	154	72	355	140	132	49	114	59	62	37
sayonara adj.	579.2	1	1	0	0	0	0	0	0	0	0	1	1	0	0
teriyaki n.	580.1	134	111	52	46	19	16	20	19	13	11	21	13	9	6
yakitori n.	581.1	96	68	7	5	9	8	30	20	3	2	11	11	36	22
dashi n.	582.1	95	63	32	22	7	5	38	21	8	5	3	3	7	7
living, adj. and n.1	584. special uses	1	1	0	0	1	1	0	0	0	0	0	0	0	0
sanpaku n.	585.1	3	1	0	0	3	1	0	0	0	0	0	0	0	0
Shotokan n.	586.1	41	20	3	3	34	14	3	2	1	1	0	0	0	0
umami n.	588.1	226	138	42	26	62	33	64	46	22	11	14	8	22	14
Wagyu n.	589.1	380	226	51	27	34	27	118	68	24	15	46	30	107	59
gaijin, adj.	590.B	23	19	12	8	1	1	6	6	0	0	2	2	2	2
ninja n.	591.1	1244	724	520	326	350	203	62	35	198	113	93	34	21	13
shunga n.	592.1	3	3	1	1	1	1	1	1	0	0	0	0	0	0
Suzuki n.	593.1	54	29	27	18	3	3	3	3	1	1	19	3	1	1
yakuza n.	595.1	285	148	73	53	126	43	33	17	19	15	25	14	9	6
Godzilla n.	597.1	16	12	4	4	4	3	1	1	0	0	7	4	0	0
gyoza n.	598.1	131	69	16	10	35	16	34	21	13	8	26	8	7	6
keiretsu n.	599.1	13	12	2	2	6	5	1	1	1	1	1	1	2	2
sensei n.	600.2	308	104	169	52	73	32	9	6	23	6	18	4	16	4
kamikaze adj.	602.1b	43	36	24	19	16	14	1	1	1	1	1	1	0	0
karate n.	603.der	23	15	2	2	19	11	1	1	1	1	0	0	0	0
ninja, adj.	604.1	664	475	283	216	202	138	43	32	103	64	28	20	5	5
ponzu n.	605.1	46	25	6	5	3	3	25	10	6	4	0	0	6	3
Tokugawa n.	606.1	3	3	1	1	1	1	1	1	0	0	0	0	0	0
shiatsu n.	607.1	102	68	17	12	36	27	5	5	8	7	5	5	31	12
sogo shosha n.	609.1	1	1	0	0	0	0	0	0	0	0	1	1	0	0
gaijin n.	611.A	49	40	24	18	10	9	10	9	0	0	4	3	1	1
seitan n.	614.1	67	38	34	27	10	9	1	0	21	2	0	0	1	0
Shinkansen n.	615.1	190	102	35	26	21	13	100	43	7	5	15	8	12	7
itai-itai n.	617.1	4	4	2	2	0	0	1	1	0	0	0	0	1	1
nikkei n.1	618.1	4	2	4	2	0	0	0	0	0	0	0	0	0	0
nunchaku n.	619.1	25	19	6	6	12	7	0	0	1	1	2	2	4	3
hijiki n.	622.1	19	9	1	1	1	1	12	3	3	2	0	0	2	2
minshuku n.	624.1	14	6	1	1	5	2	8	3	0	0	0	0	0	0
omiai n.	625.1	3	2	2	1	0	0	1	1	0	0	0	0	0	0
panko n.	626.1	82	49	34	19	19	14	13	8	0	0	7	5	9	3
Red Army n.	628.1	11	11	5	5	2	2	2	2	0	0	1	1	1	1
shabu-shabu n.	629.1	64	39	3	3	2	2	27	16	13	7	7	4	12	7
sushi n.	631.der	2	2	1	1	0	0	1	1	0	0	0	0	0	0
teppan-yaki n.	632.1	116	75	11	6	12	8	18	12	9	3	29	18	37	28
kamikaze n.	633.2c	6	6	4	4	2	2	0	0	0	0	0	0	0	0
meishi n.	634.1	2	2	0	0	0	0	1	1	1	1	0	0	0	0
tataki n.	637.1	30	19	2	2	6	5	5	5	1	1	14	4	2	2
toro n.2	638.1	38	27	9	6	3	2	7	4	1	1	2	2	16	12

futon n.	639. draft	164	118	96	70	41	31	9	7	12	5	4	4	2	1
ninja n.	640.2	193	168	91	81	76	64	8	7	10	8	7	7	1	1
Tsukahara n.	641.1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
O n.1	642. ini-tialism	5	5	1	1	0	0	0	0	0	0	2	2	2	2
office n.	643. comp	20	20	4	4	2	2	6	6	0	0	6	6	2	2
sai n.3	644.1	6	3	5	2	1	1	0	0	0	0	0	0	0	0
surimi n.	646.1	23	14	9	5	3	3	3	2	0	0	8	4	0	0
Nikkei n.2	648.1	129	98	29	22	49	40	15	11	7	5	23	15	6	5
quality n. and adj.	649. comp	32	25	6	6	15	8	2	2	4	4	1	1	4	4
robata-yaki n.	650.1	24	14	1	1	0	0	0	0	0	0	0	0	23	13
Shorin ryu n.	651.1	3	2	2	1	0	0	0	0	0	0	1	1	0	0
money n.	652.C2	84	57	2	2	1	1	1	1	3	3	77	50	0	0
Nikkei n.2	653.2a	5	5	0	0	2	2	2	2	0	0	1	1	0	0
reiki n.	654.1	773	285	347	124	306	107	52	16	12	8	8	6	48	24
shishito n.	655.1	17	7	6	1	4	2	4	2	0	0	2	1	1	1
mechatronics n.	657.1	45	38	10	8	12	12	6	5	1	1	13	9	3	3
kanban n.	659.1	27	13	5	4	21	8	1	1	0	0	0	0	0	0
kanban n.	660.2	136	44	64	16	67	24	0	0	2	1	0	0	3	3
karaoke n.	661.1	1955	1508	443	360	714	581	223	168	191	115	212	148	172	136
tamari n.	662.1	30	23	16	11	9	9	5	3	0	0	0	0	0	0
hamachi n.	663.1	40	26	11	9	1	1	5	5	5	2	0	0	18	9
shuriken n.	664.1	58	38	30	21	12	12	1	1	4	3	11	1	0	0
uncanny valley n.	665.1	74	65	42	42	32	23	0	0	0	0	0	0	0	0
washi n.	666.1	90	52	26	16	17	14	14	10	19	5	12	5	2	2
kamikaze n.	667.4	6	6	4	4	1	1	1	1	0	0	0	0	0	0
mokume gane n.	668.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
nikkeijin n.	669.1	1	1	0	0	1	1	0	0	0	0	0	0	0	0
omakase n.	670.1	41	22	5	5	1	1	9	4	4	3	4	3	18	6
shojo n.	672.1	91	52	66	31	1	1	20	16	3	3	1	1	0	0
papasan n.	671.2	8	5	5	2	2	2	0	0	1	1	0	0	0	0
okadaic adj.	673.1	4	3	2	2	2	1	0	0	0	0	0	0	0	0
hand roll n.	674.2	19	18	4	4	1	1	4	4	0	0	2	2	8	7
Nikkei n.2	675.3	87	62	24	17	43	29	9	7	2	2	4	3	5	4
shonen n.	676.1	80	61	61	44	2	2	3	3	10	8	4	4	0	0
bukkake n.	679.1	1	1	0	0	0	0	0	0	0	0	1	1	0	0
kawaii adj.	680.1	130	94	19	13	20	15	46	29	23	17	20	18	2	2
anime n.3	681.1	4721	2414	1997	1146	580	332	743	363	812	305	533	230	56	38
kaizen n.	682.1	120	66	37	22	59	26	6	4	6	5	10	7	2	2
shiso n.	684.1	54	42	14	13	5	4	19	12	1	1	4	3	11	9
kawaii n.	686.2	13	10	4	3	4	3	2	2	2	1	0	0	1	1
mecha n.	687.1	213	144	86	63	15	13	19	12	29	17	63	38	1	1
izakaya n.	689.1	99	59	18	12	7	5	36	18	8	3	7	4	23	17
karoshi n.	691.1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
functional adj. and n.	692. comp	55	39	12	8	20	14	3	3	1	1	14	11	5	2
hentai n.	693.1	54	46	26	22	9	9	8	6	5	3	4	4	2	2
ginkgo n.	694.2	64	33	38	13	8	4	6	6	5	4	4	3	3	3
quantitative easing n.	695.1	1991	1393	577	389	1039	767	98	66	52	32	78	47	147	92
otaku n.	697.1	243	149	112	74	31	9	52	35	35	20	9	8	4	3
cosplay n.	698.1	949	387	184	138	86	55	385	97	178	61	49	21	67	15
compensate v.	699. comp	2	2	0	0	0	0	1	1	0	0	0	0	1	1
enjo kosai n.	700.1	4	3	3	2	0	0	0	0	0	0	0	0	1	1
emoji n.	701.1	73	33	36	16	19	9	10	5	2	1	1	1	5	1
cosplay n.	702.2	523	267	147	109	43	25	197	68	103	50	13	8	20	7
hikikomori n.	703.1	15	9	7	6	6	1	0	0	1	1	0	0	1	1
keitai n.	704.1	4	4	0	0	3	3	0	0	1	1	0	0	0	0

Appendix 5: Summary of the order of varieties from highest to lowest normalized frequency in each main category.

	Architecture and interior design	Arts and crafts	Attributes and qualities	Business and finance	Cooking and beverages
1	SgE	PhE	PhE	HKE	SgE
2	MyE	SgE	SgE	BrE	HKE
3	HKE	MyE	MyE	SgE	MyE
4	PhE	HKE	AmE	MyE	PhE
5	AmE	AmE	BrE	PhE	AmE
6	BrE	BrE	HKE	AmE	BrE

	Currency and units of measurement	Education	Electronics and technology	Entertainment and leisure activities	Fashion
1	SgE	SgE	SgE	SgE	SgE
2	MyE	MyE	PhE	PhE	HKE
3	HKE	PhE	HKE	MyE	MyE
4	PhE	AmE	MyE	AmE	PhE
5	BrE	HKE	AmE	HKE	BrE
6	AmE	BrE	BrE	BrE	AmE

	Government, administration, and politics	Greetings, interjections, and gestures	Historical periods	Manufacturing and materials	Military and bushido
1	HKE	MyE	SgE	SgE	SgE
2	MyE	SgE	HKE	MyE	PhE
3	SgE	PhE	PhE	PhE	HKE
4	PhE	HKE	MyE	BrE	MyE
5	BrE	AmE	BrE	AmE	AmE
6	AmE	BrE	AmE	HKE	BrE

	Nature	Science and medicine	Seafaring, vehicles, and transportation	Society	Sports and martial arts
1	SgE	HKE	SgE	SgE	SgE
2	MyE	SgE	HKE	PhE	MyE
3	PhE	MyE	MyE	MyE	PhE
4	HKE	AmE	PhE	HKE	HKE
5	AmE	PhE	AmE	AmE	BrE
6	BrE	BrE	BrE	BrE	AmE

	Tradition, mythology, and religion	Values and concepts	Weaponry	Writing, language, and bookbinding	Other objects
1	HKE	SgE	PhE	PhE	HKE
2	SgE	HKE	MyE	SgE	BrE
3	PhE	PhE	SgE	AmE	AmE
4	AmE	AmE	AmE	MyE	SgE/PhE/MyE
5	MyE	BrE	HKE	HKE	-
6	BrE	MyE	BrE	BrE	-

Appendix 6: Combined raw and normalized frequencies of LS pairs in the main and subcategories per variety.

		Architecture and interior design		Architecture		Buildings and property		Interior design and furniture	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	259	0.670	41	0.106	95	0.246	124	0.321
	base	179	0.463	34	0.088	53	0.137	93	0.240
BrE	tokens	183	0.472	51	0.132	81	0.209	51	0.132
	base	115	0.297	35	0.090	40	0.103	40	0.103
SgE	tokens	503	11.705	56	1.303	416	9.680	33	0.768
	base	163	3.793	37	0.861	101	2.350	27	0.628
PhE	tokens	36	0.832	7	0.162	16	0.370	13	0.301
	base	20	0.462	6	0.139	8	0.185	6	0.139
MyE	tokens	128	3.017	20	0.471	101	2.381	9	0.212
	base	44	1.037	14	0.330	23	0.542	9	0.212
HKE	tokens	117	2.892	19	0.470	85	2.101	13	0.321
	base	64	1.582	15	0.371	42	1.038	7	0.173
All	tokens	1,226	1.299	194	0.206	794	0.842	243	0.258
	base	585	0.620	141	0.149	267	0.283	182	0.193

		Arts and crafts		Arrangement		Literature		Music		Painting, pictures, and printing	
		RF	NF	RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	765	1.978	27	0.070	237	0.613	55	0.142	210	0.543
	base	454	1.174	9	0.023	160	0.414	30	0.078	65	0.168
BrE	tokens	606	1.563	10	0.026	153	0.395	29	0.075	59	0.152
	base	376	0.970	8	0.021	100	0.258	24	0.062	36	0.093
SgE	tokens	190	4.421	5	0.116	33	0.768	4	0.093	61	1.419
	base	124	2.885	3	0.070	24	0.558	4	0.093	36	0.838
PhE	tokens	278	6.428	9	0.208	90	2.081	41	0.948	42	0.971
	base	127	2.936	5	0.116	33	0.763	9	0.208	23	0.532
MyE	tokens	131	3.088	12	0.283	9	0.212	26	0.613	27	0.636
	base	65	1.532	5	0.118	8	0.189	7	0.165	18	0.424
HKE	tokens	108	2.670	0	0.000	29	0.717	4	0.099	10	0.247
	base	60	1.483	0	0.000	22	0.544	3	0.074	7	0.173
All	tokens	2,078	2.202	63	0.067	551	0.584	159	0.169	409	0.433
	base	1,206	1.278	30	0.032	347	0.368	77	0.082	185	0.196

		Paper folding and packaging		Theater, puppetry, and dance		Pottery and lacquerware		Gemstones		Other crafts	
		RF	NF	RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	213	0.551	57	0.147	14	0.036	5	0.013	8	0.021
	base	145	0.375	37	0.096	11	0.028	4	0.010	6	0.016
BrE	tokens	239	0.617	59	0.152	64	0.165	2	0.005	5	0.013
	base	148	0.382	48	0.124	14	0.036	2	0.005	3	0.008
SgE	tokens	43	1.001	32	0.745	17	0.396	1	0.023	1	0.023
	base	28	0.652	23	0.535	8	0.186	1	0.023	1	0.023
PhE	tokens	21	0.486	70	1.618	9	0.208	0	0.000	0	0.000
	base	21	0.486	31	0.717	8	0.185	0	0.000	0	0.000
MyE	tokens	38	0.896	9	0.212	10	0.236	3	0.071	3	0.071
	base	19	0.448	5	0.118	2	0.047	3	0.071	1	0.024

HKE	tokens	17	0.420	35	0.865	10	0.247	3	0.074	0	0.000
	base	15	0.371	10	0.247	1	0.025	2	0.049	0	0.000
All	tokens	571	0.605	262	0.278	124	0.131	14	0.015	17	0.018
	base	376	0.399	154	0.163	44	0.047	12	0.013	11	0.012

		Attributes and qualities		**Business and finance**		Business and working culture		Finance	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	464	1.200	1,051	2.717	421	1.088	630	1.629
	base	371	0.959	761	1.967	333	0.861	428	1.106
BrE	tokens	364	0.939	1,948	5.026	815	2.103	1,133	2.923
	base	279	0.720	1,483	3.826	645	1.664	838	2.162
SgE	tokens	106	2.467	256	5.957	132	3.072	124	2.885
	base	77	1.792	186	4.328	100	2.327	86	2.001
PhE	tokens	137	3.168	180	4.162	119	2.751	61	1.410
	base	90	2.081	134	3.098	95	2.197	39	0.902
MyE	tokens	71	1.674	327	7.709	221	5.210	106	2.499
	base	55	1.297	227	5.351	161	3.795	66	1.556
HKE	tokens	17	0.420	500	12.361	342	8.455	158	3.906
	base	17	0.420	311	7.688	210	5.192	101	2.497
All	tokens	1,159	1.228	4,262	4.517	2,050	2.173	2,212	2.344
	base	889	0.942	3,102	3.288	1,544	1.636	1,558	1.651

		Cooking and beverages		Beverages		Food and cooking	
		RF	NF	RF	NF	RF	NF
AmE	tokens	4,085	10.561	297	0.768	3,842	9.933
	base	2,432	6.287	168	0.434	2,280	5.894
BrE	tokens	2,217	5.720	194	0.500	2,059	5.312
	base	1,531	3.950	125	0.322	1,425	3.676
SgE	tokens	3,843	89.425	259	6.027	3,641	84.724
	base	1,898	44.166	144	3.351	1,790	41.652
PhE	tokens	1,051	24.301	48	1.110	1,027	23.746
	base	547	12.647	25	0.578	532	12.301
MyE	tokens	1,398	32.956	62	1.462	1,352	31.872
	base	769	18.128	41	0.967	737	17.374
HKE	tokens	1,929	47.688	267	6.601	1,690	41.780
	base	1,062	26.254	112	2.769	967	23.906
All	tokens	14,523	15.392	1,127	1.194	13,611	14.426
	base	8,239	8.732	615	0.652	7,731	8.194

		Currency and units of measurement		Currency		Units of measurement		**Education**	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	773	1.998	741	1.916	32	0.083	182	0.471
	base	390	1.008	363	0.938	27	0.070	63	0.163
BrE	tokens	1,042	2.688	1,009	2.603	33	0.085	33	0.085
	base	438	1.130	415	1.071	23	0.059	22	0.057
SgE	tokens	977	22.734	934	21.734	43	1.001	49	1.140
	base	395	9.191	362	8.424	33	0.768	31	0.721
PhE	tokens	254	5.873	253	5.850	1	0.023	33	0.763
	base	98	2.266	97	2.243	1	0.023	17	0.393
MyE	tokens	423	9.972	408	9.618	15	0.354	42	0.990
	base	174	4.102	165	3.890	9	0.212	18	0.424
HKE	tokens	351	8.677	335	8.282	16	0.396	4	0.099
	base	169	4.178	159	3.931	10	0.247	4	0.099

All	tokens	3,820	4.049	3,680	3.900	140	0.148	343	0.364
	base	1,664	1.764	1,561	1.654	103	0.109	155	0.164

		Electronics and technology		**Entertainment and leisure activities**		Comics and cartoons		Leisure activities	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	202	0.522	4,674	12.083	4,039	10.442	863	2.231
	base	142	0.367	2,877	7.438	2,412	6.236	660	1.706
BrE	tokens	103	0.266	1,938	5.000	1,004	2.590	961	2.479
	base	59	0.152	1,329	3.429	626	1.615	733	1.891
SgE	tokens	69	1.606	2,797	65.085	2,149	50.006	833	19.383
	base	46	1.070	1,275	29.669	988	22.990	354	8.237
PhE	tokens	41	0.948	2,078	48.046	1,579	36.509	764	17.665
	base	25	0.578	792	18.312	648	14.983	243	5.618
MyE	tokens	23	0.542	1,242	29.279	897	21.146	311	7.331
	base	18	0.424	640	15.087	456	10.750	192	4.526
HKE	tokens	24	0.593	474	11.718	213	5.266	281	6.947
	base	14	0.346	277	6.848	94	2.324	177	4.376
All	tokens	462	0.490	13,203	13.993	9,881	10.472	4,013	4.253
	base	304	0.322	7,190	7.620	5,224	5.537	2,359	2.500

		Games		Travelling and tourism		**Fashion**		Accessories		Clothes	
		RF	NF	RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	384	0.993	50	0.129	205	0.530	3	0.008	193	0.499
	base	278	0.719	21	0.054	133	0.344	3	0.008	125	0.323
BrE	tokens	179	0.462	52	0.134	280	0.722	9	0.023	260	0.671
	base	104	0.268	26	0.067	195	0.503	9	0.023	178	0.459
SgE	tokens	601	13.985	378	8.796	184	4.282	0	0.000	177	4.119
	base	182	4.235	81	1.885	107	2.490	0	0.000	100	2.327
PhE	tokens	289	6.682	8	0.185	45	1.040	0	0.000	44	1.017
	base	118	2.728	5	0.116	35	0.809	0	0.000	34	0.786
MyE	tokens	64	1.509	94	2.216	56	1.320	0	0.000	56	1.320
	base	31	0.731	19	0.448	36	0.849	0	0.000	36	0.849
HKE	tokens	93	2.299	61	1.508	100	2.472	1	0.025	96	2.373
	base	26	0.643	24	0.593	64	1.582	1	0.025	60	1.483
All	tokens	1,610	1.706	643	0.681	870	0.922	13	0.014	826	0.875
	base	739	0.783	176	0.187	570	0.604	13	0.014	533	0.565

		Hairstyles		Footwear		**Government, administration, and politics**		**Greetings, interjections, and gestures**		**Historical periods**	
		RF	NF	RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	0	0.000	9	0.023	585	1.512	36	0.093	83	0.215
	base	0	0.000	5	0.013	462	1.194	31	0.080	66	0.171
BrE	tokens	0	0.000	11	0.028	911	2.350	20	0.052	163	0.421
	base	0	0.000	8	0.021	750	1.935	20	0.052	68	0.175
SgE	tokens	0	0.000	7	0.163	173	4.026	12	0.279	93	2.164
	base	0	0.000	7	0.163	112	2.606	12	0.279	66	1.536
PhE	tokens	0	0.000	1	0.023	130	3.006	7	0.162	74	1.711
	base	0	0.000	1	0.023	103	2.381	5	0.116	39	0.902
MyE	tokens	0	0.000	0	0.000	299	7.049	12	0.283	28	0.660
	base	0	0.000	0	0.000	214	5.045	10	0.236	22	0.519
HKE	tokens	0	0.000	3	0.074	382	9.444	4	0.099	83	2.052
	base	0	0.000	3	0.074	217	5.365	4	0.099	20	0.494
All	tokens	0	0.000	31	0.033	2,480	2.628	91	0.096	524	0.555
	base	0	0.000	24	0.025	1,858	1.969	82	0.087	281	0.298

		Manufacturing and materials		**Military and bushido**		**Nature**		Environment	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	106	0.274	678	1.753	3,997	10.333	2,545	6.579
	base	44	0.114	439	1.135	2,338	6.044	1,427	3.689
BrE	tokens	115	0.297	609	1.571	3,061	7.897	2,116	5.459
	base	54	0.139	375	0.967	1,908	4.922	1,320	3.405
SgE	tokens	70	1.629	191	4.444	1,412	32.857	609	14.171
	base	42	0.977	113	2.629	748	17.406	286	6.655
PhE	tokens	37	0.855	179	4.139	898	20.763	363	8.393
	base	16	0.370	110	2.543	368	8.509	212	4.902
MyE	tokens	52	1.226	76	1.792	1,319	31.094	625	14.734
	base	12	0.283	52	1.226	536	12.635	314	7.402
HKE	tokens	9	0.222	124	3.065	792	19.580	391	9.666
	base	9	0.222	48	1.187	448	11.075	212	5.241
All	tokens	389	0.412	1,857	1.968	11,479	12.166	6,649	7.047
	base	177	0.188	1,137	1.205	6,346	6.726	3,771	3.997

		Fauna		Flora		Fungi		**Science and medicine**	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	208	0.538	1,177	3.043	67	0.173	1,558	4.028
	base	141	0.365	726	1.877	44	0.114	657	1.699
BrE	tokens	296	0.764	621	1.602	28	0.072	1,483	3.826
	base	144	0.372	420	1.084	24	0.062	617	1.592
SgE	tokens	204	4.747	572	13.310	27	0.628	204	4.747
	base	137	3.188	305	7.097	20	0.465	91	2.118
PhE	tokens	86	1.988	433	10.012	16	0.370	107	2.474
	base	44	1.017	100	2.312	12	0.277	67	1.549
MyE	tokens	452	10.655	225	5.304	17	0.401	172	4.055

	base	109	2.570	102	2.405	11	0.259	73	1.721
HKE	tokens	167	4.129	204	5.043	30	0.742	468	11.570
	base	96	2.373	119	2.942	21	0.519	169	4.178
All	tokens	1,413	1.498	3,232	3.425	185	0.196	3,992	4.231
	base	671	0.711	1772	1.878	132	0.140	1,674	1.774

		Alternative medicine		Biology		Chemistry and biochemistry		Physics		Psychology	
		RF	NF	RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	1,378	3.562	5	0.013	5	0.013	0.421	3	0.008	7
	base	563	1.455	2	0.005	5	0.013	0.204	2	0.005	6
BrE	tokens	1,402	3.617	5	0.013	8	0.021	0.163	4	0.010	6
	base	579	1.494	2	0.005	4	0.010	0.077	3	0.008	1
SgE	tokens	191	4.444	0	0.000	10	0.233	0.070	0	0.000	0
	base	84	1.955	0	0.000	4	0.093	0.070	0	0.000	0
PhE	tokens	105	2.428	0	0.000	0	0.000	0.023	0	0.000	1
	base	65	1.503	0	0.000	0	0.000	0.023	0	0.000	1
MyE	tokens	172	4.055	0	0.000	0	0.000	0.000	0	0.000	0
	base	73	1.721	0	0.000	0	0.000	0.000	0	0.000	0
HKE	tokens	449	11.100	0	0.000	2	0.049	0.247	6	0.148	1
	base	163	4.030	0	0.000	2	0.049	0.049	1	0.025	1
All	tokens	3,697	3.918	10	0.011	25	0.026	0.254	13	0.014	15
	base	1,527	1.618	4	0.004	15	0.016	0.122	6	0.006	9

		Parapsychology		**Seafaring, vehicles, and transportation**		**Society**		Groups and organizations	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	1	0.003	98	0.253	3,143	8.125	131	0.339
	base	1	0.003	61	0.158	1,988	5.139	99	0.256
BrE	tokens	0	0.000	59	0.152	2,585	6.669	175	0.451
	base	0	0.000	43	0.111	1,812	4.675	82	0.212
SgE	tokens	0	0.000	100	2.327	1,119	26.039	44	1.024
	base	0	0.000	43	1.001	604	14.055	27	0.628
PhE	tokens	0	0.000	14	0.324	870	20.116	28	0.647
	base	0	0.000	9	0.208	516	11.931	21	0.486
MyE	tokens	0	0.000	15	0.354	753	17.751	41	0.967
	base	0	0.000	8	0.189	435	10.255	21	0.495
HKE	tokens	0	0.000	15	0.371	698	17.256	42	1.038
	base	0	0.000	10	0.247	361	8.925	19	0.470
All	tokens	1	0.001	301	0.319	9,168	9.717	461	0.489
	base	1	0.001	174	0.184	5,716	6.058	269	0.285

		Professionals and hobbyists		Ethnicity and residency		Status and rank		Titles, forms of address, and names	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	1,127	2.914	169	0.44	1,127	2.914	643	1.662
	base	812	2.099	83	0.21	812	2.099	253	0.654
BrE	tokens	1,437	3.707	27	0.07	1,437	3.707	208	0.537
	base	1,102	2.843	26	0.07	1,102	2.843	132	0.341
SgE	tokens	342	7.958	26	0.61	342	7.958	250	5.817
	base	234	5.445	23	0.54	234	5.445	106	2.467

PhE	tokens	293	6.775	7	0.16	293	6.775	96	2.220
	base	207	4.786	4	0.09	207	4.786	46	1.064
MyE	tokens	314	7.402	9	0.21	314	7.402	134	3.159
	base	234	5.516	8	0.19	234	5.516	61	1.438
HKE	tokens	467	11.545	10	0.25	467	11.545	98	2.423
	base	258	6.378	8	0.20	258	6.378	32	0.791
All	tokens	3,980	4.218	248	0.26	3,980	4.218	1,429	1.515
	base	2,847	3.017	152	0.16	2,847	3.017	630	0.668

		Other types of people		**Sports and martial arts**		Martial arts		Sports	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	923	2.386	3,657	9.454	3,657	9.454	0	0.000
	base	572	1.479	2,112	5.460	2,112	5.460	0	0.000
BrE	tokens	647	1.669	3,802	9.809	3,691	9.522	111	0.286
	base	384	0.991	2,227	5.745	2,156	5.562	71	0.183
SgE	tokens	162	3.770	1,061	24.689	1,058	24.619	3	0.070
	base	99	2.304	493	11.472	492	11.449	1	0.023
PhE	tokens	294	6.798	694	16.046	694	16.046	0	0.000
	base	178	4.116	400	9.249	400	9.249	0	0.000
MyE	tokens	170	4.008	729	17.185	704	16.596	25	0.589
	base	88	2.074	378	8.911	363	8.557	15	0.354
HKE	tokens	45	1.112	599	14.808	593	14.660	6	0.148
	base	33	0.816	289	7.145	286	7.070	3	0.074
All	tokens	2,241	2.375	10,542	11.173	10,397	11.019	145	0.154
	base	1,354	1.435	5,899	6.252	5,809	6.157	90	0.095

		Tradition, mythology, and religion		Mythology		Religion		Tradition	
		RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	1,850	4.783	4	0.010	1,788	4.622	59	0.153
	base	916	2.368	3	0.008	884	2.285	30	0.078
BrE	tokens	731	1.886	1	0.003	694	1.790	39	0.101
	base	432	1.115	1	0.003	416	1.073	17	0.044
SgE	tokens	381	8.866	7	0.163	281	6.539	97	2.257
	base	225	5.236	4	0.093	183	4.258	41	0.954
PhE	tokens	266	6.150	6	0.139	256	5.919	7	0.162
	base	124	2.867	5	0.116	116	2.682	6	0.139
MyE	tokens	178	4.196	0	0.000	166	3.913	12	0.283
	base	99	2.334	0	0.000	91	2.145	8	0.189
HKE	tokens	385	9.518	0	0.000	366	9.048	19	0.470
	base	159	3.931	0	0.000	149	3.684	10	0.247
All	tokens	3,791	4.018	18	0.019	3,551	3.764	233	0.247
	base	1,955	2.072	13	0.014	1,839	1.949	112	0.119

		Values and concepts		**Weaponry**		**Writing, language, and bookbinding**		Bookbinding		Writing and language	
		RF	NF	RF	NF	RF	NF	RF	NF	RF	NF
AmE	tokens	512	1.324	169	0.437	442	1.143	0	0.000	442	1.143
	base	213	0.551	114	0.295	199	0.514	0	0.000	199	0.514
BrE	tokens	468	1.207	77	0.199	108	0.279	0	0.000	108	0.279
	base	191	0.493	59	0.152	82	0.212	0	0.000	82	0.212

SgE	tokens	141	3.281	33	0.768	169	3.933	0	0.000	169	3.933
	base	62	1.443	16	0.372	77	1.792	0	0.000	77	1.792
PhE	tokens	58	1.341	62	1.434	423	9.780	0	0.000	423	9.780
	base	41	0.948	40	0.925	87	2.012	0	0.000	87	2.012
MyE	tokens	40	0.943	41	0.967	37	0.872	0	0.000	37	0.872
	base	33	0.778	16	0.377	28	0.660	0	0.000	28	0.660
HKE	tokens	61	1.508	13	0.321	20	0.494	0	0.000	20	0.494
	base	31	0.766	6	0.148	16	0.396	0	0.000	16	0.396
All	tokens	1,280	1.357	395	0.419	1,199	1.271	0	0.000	1199	1.271
	base	571	0.605	251	0.266	489	0.518	0	0.000	489	0.518

		Other objects	
		RF	NF
AmE	tokens	1	0.003
	base	1	0.003
BrE	tokens	4	0.010
	base	2	0.005
SgE	tokens	0	0.000
	base	0	0.000
PhE	tokens	0	0.000
	base	0	0.000
MyE	tokens	0	0.000
	base	0	0.000
HKE	tokens	3	0.074
	base	3	0.074
All	tokens	8	0.008
	base	6	0.006