‘WEEABOO JAPANESE’: EXPLORING ENGLISH-JAPANESE LANGUAGE-MIXING IN ONLINE JAPANESE POPULAR CULTURE FANDOM

A THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER IN PHILOSOPHY

IN

LINGUISTICS

AT MASSEY UNIVERSITY, ALBANY, NEW ZEALAND

ROWAN ELIZABETH ARBUTHNOTT GARDINER

2019

ABSTRACT

The complexities of a globalised modern society pose methodological and theoretical issues for linguistic research in areas such as Language Contact, Language-Mixing, and Sociolinguistics, due to the commodification and transmission of language and language features resulting in new language interactions. The boundary between definitions of language borrowing and code-switching is currently a matter of increased interest, particularly in terms of research identifying cases of language use involving unskilled participants.

This study investigates and linguistically analyses the presence of Japanese language features within English language contexts that are produced by members of online discussion forums who are fans of Japanese popular culture, and for whom fluency in Japanese language is not assumed nor required for participation. Corpus linguistics techniques were employed on data gathered from two online sources in order to identify what linguistic features were present and establish their extent according to frequency. These same corpora were qualitatively analysed to establish community attitudes towards English-Japanese language mixing and what these results indicated in terms of policing and community norms, and overall what both the quantitative and qualitative results meant for how the language phenomena could be defined according to current theoretical paradigms.

The results showed that the most frequent word class was nouns, and the semantic domains found were mostly related to Japanese fandom concepts that were topical to the forum, such as specific interests, clothing and fashion, food, media related terms, and religion and cultural terms. Most instances were single-word insertions, and where the few multi-word segments occurred they were specifically in reference to a negative stereotype within the community (weeaboo). This stereotype also indicated language policing was a factor affecting language use, and the results showed that while the Japanese language had high status, language use was socially restricted to specific situations and extents of use.

The language phenomenon is described as mostly language borrowing behaviours, but as the words retain a high level of knowledge of related assignations and also occur concurrently with a few code-switching type behaviours, the usage-based approach where both elements are considered different aspects of the same continuum is seen as a preferable theoretical paradigm.

ACKNOWLEDGEMENTS

A heartfelt thanks to my thesis supervisors, Dr Mary Salisbury, Dr David Ishii, and Dr Arianna Berardi-Wiltshire, who all gave me invaluable input, support, and encouragement throughout the whole master's thesis process.

I would also like to thank my mother, Anne Rundle, for her encouragement and editing and proofreading, and my father, Stuart Gardiner, for his encouragement and help with statistics.

Finally, I would like to respectfully acknowledge the anonymous members of the online communities I gathered my linguistic data from.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

TV Tropes ("Gratuitous Japanese", n.d.), a website dedicated to cataloguing tropes found in media, defines Gratuitous Japanese as something that is a feature in the *Anime*[1] Fanfiction Community. It varies in intensity, and thus in acceptableness, from words that are often considered to be familiar to even casual *anime* viewers, such as honorifics, kinship terms, and a variety of interjections and expressions, that are intended to help create atmosphere or flavour, to entire sentences or paragraphs in Japanese of varying degrees of accuracy. Further than that, the website states, is Fangirl Japanese which not only involves large amounts of Japanese text in their fanfiction, but also in their everyday life. ("Gratuitous Japanese", n.d.) It is the everyday life usage that is the focus of the current study, and is represented in the form of written language data gathered from the Cosplay and Elegant Gothic Lolita board (/CGL/) on the online image board website boards.4channel.org, and the now defunct Gratuitous Japanese Troper Tales webpage that used to be attached to TV Tropes. Commenters, called writers in this thesis, shared their real-life usage, interactions, and attitudes towards the trope.

The particular language contact situation that precipitated the use of language in question is interesting as it is due to exposure to pop culture media from a different culture and internet interaction rather than face to face contact as a minority within a different language majority community. Members of the community come into contact with the Japanese language through exposure to media imported from Japan, and then engage in interaction over the internet and in person with other English-speaking fans who have been similarly exposed to Japanese. This case is also not about local language contact along borders or migrant populations, or about participating in the global economy as studies on English use by other language spaces often find. Therefore, the community fits within LePage and Tabouret-Keller's (1985) comments that the speakers in question have easy access to the group through the internet; are motivated to be a part of the community by their appreciation of Japanese culture and media which is then monitored by the other members; and have the ability to change their linguistic behaviour to match that which is observed in the community; showing popular culture as a sociolinguistic space involving both collaboration and contention.

When referred to as such within this study, Anglophone Japanese Culture Fandom refers to a community that is global and has interactions in online spaces but involves English-based

[1] The Japanese term for animation, adopted in other countries to refer specifically to Japanese animation.

communication (whether as L1 or L2[2]) and covers Japanese popular culture products such as media (for example *anime*, *manga*[3], J-dramas, J-pop) and fashion, as well as interest in other aspects of both traditional and modern Japanese culture and language. Japanese popular culture and resultant fandoms have been explored in many studies from different disciplines (Jenkins, 2006), but the current study aims to concentrate on the linguistic characteristics. Williams (2006, p. 246) notes that "…it is not possible to essentialise *anime* fandom", as the popular culture phenomenon extends far beyond simply being a genre, and *anime* fans' experiences can vary wildly through *what* they experience as well as *how* they, as individuals, experience it. This does not, however, preclude claims being made about the linguistic and social behaviours and attitudes of those in the fandom, just that the scope and variability should be accounted for with clear descriptions of context and that the area is still a rich vein for future research within a number of different research disciplines. Of course, there is far more to what we might call Japanese Culture Fandom beyond animation, such as music, live action media, and fashion, but *anime* fandom is often used as a shorthand to reference the wider range, and this can be used as evidence to show that while there are different sub-communities within the fandom, in terms of topics and modality or location, there are trends that can be generalised across the whole. The focus of the current study is mostly on a section of the *cosplay* and Japanese fashion community, which is a more specific and niche area than *anime*.

When conducting research, the importance of acknowledging the particular standing of the researcher is stressed, as "there is no 'view from nowhere,' no gaze that is not positioned" (Irvine & Gal, 2000, p. 36), and the results of a study are typically informed by the theoretical and methodological positioning, and other potential biases, of the researcher.

The current study is not approached from a position of bilingual competence (i.e. proficiency in both Japanese and English), which informs some of the methodological and analytical decisions such as not looking at the few non-*romaji* tokens. I am a native English speaker who is not fluent in Japanese and I have never undergone either formal or self-directed schooling to become so. However, as the linguistic phenomenon under investigation is not seen to be, nor going to be, analysed as either bilingual or multilingual behaviour where speakers are considered to have a degree of linguistic competence, this limitation is unlikely to have a negative impact or influence on the study. As such, the translations and glosses of examples in this thesis will be relatively

[2] First language or second language.
[3] The Japanese term for graphic novels or comics; adopted by other countries to refer to Japanese graphic novels and comics.

basic. A lack of linguistic competence is a feature of the language being analysed in this study, as will be established in the results.

On the other hand, I am currently, or have been in the past, familiar with a wide variety of Japanese popular culture products and have interacted with the wider Japanese Culture Fandom community, including the particular section from which the data has been gathered (Cosplay and Japanese fashion). Thus, I consider myself to have both ethnographic insight into general behaviours and attitudes, and previous knowledge of much of the vocabulary and language features discovered.

Specifically, the gathered language data will be examined through a corpus linguistics approach with added elements of sociolinguistic factors and ethnographic perspective. The study itself is shaped according to the following three research questions:

(1) To what extent is Japanese language used; what linguistic forms and features are present, and how are they integrated?
(2) What attitudes towards this language use are found within the community, and how do those attitudes possibly affect language use?
(3) How do the patterns of language use and attitudes found within this community relate and/or contribute to current theoretical paradigms of language mixing?

To do so, the Japanese language used by writers will be analysed according to linguistic features such as orthography, word class frequency, semantic domain, the modification or naturalisation patterns present, as well as analysis in terms of loanwords and multi-word segments, and identifying positive or negative views towards the language that members of the community hold and how they result in language constraints or policing.

The general theoretical paradigms and trends found in previous research in the area will be described in Chapter 2, followed by the research design and data gathering methodology in Chapter 3, which includes descriptions of the exact contexts the data is being gathered from. Chapter 4 will include discussion and comparison to other research as well as the analysis areas noted above, whilst Chapter 5 will explore more general issues brought up by the results as a whole in order to answer Research Question 3 by integrating the overview of the language phenomenon laid out in Chapter 4 with wider linguistic theoretical paradigms.

2 THEORETICAL BACKGROUND

Language mixing is a feature of many language contact situations where different people meet and their languages interact, and is one aspect of focus in the field of Contact Linguistics. The following chapter will explore the general facets of contact linguistics and language mixing, the theoretical paradigms that have been developed to describe and explain the language interactions, and outline the research gap in which the current study fits.

2.1 CONTACT LINGUISTICS AND LANGUAGE MIXING

Within Contact Linguistics there are many different terms that describe different features, extents and processes of language mixing, such as vernaculars, styling, mixed language varieties like creoles and pidgins that are widely considered to be distinct language varieties, code-switching, crossing and polylanguaging, and borrowing and structural changes. There is much overlap between the linguistic features of these definitions, and different approaches and methodologies that researchers use as appropriate (for the latter see Chapter 2.2).

According to Winford (2003, p. 5), "Its [Contact Linguistics] objective is to study the varied situations of contact between languages, the phenomena that result, and the interaction of linguistic and external ecological factors in shaping these outcomes." Relevant social factors that affect language use include the length and intensity of contact between language communities, their population sizes in comparison to each other, the power relationships that occur in their interactions, and the communicative functions (Winford, 2003). Related is also socio-political factors such as attitudes towards the languages in general as well as mixing that occur on an individual and group level and any other factors that could motivate the use of one or the other language (Winford, 2003).

As a result of social and political factors including attitudes, language mixing has a historically negative reputation. Code-switching studies have generally revealed that speakers are considered 'careless', 'slack', or 'lazy', and unwilling to spend the time to learn and speak 'proper language' (Wiese, 2015). This is reflected in many of the colloquial terms for mixed language behaviour, such as Gutter Spanish for Spanish-Romani, African American Vernacular English often referred to as *thug* or *gangsta* language, and the connotations of incompetence that Spanglish and Engrish have. The English-Japanese language mixing used by the Anglophone

Japanese pop culture fandom is also an example of this; it is often called Gratuitous Japanese or Weeaboo Japanese.

The use of the term weeaboo in reference to a stereotype of a fan of Japanese pop culture who is obsessed with Japan to the point of possibly wishing themselves to be Japanese was in fact an innovation originating from the website 4chan.org, a section of which is the source for the current study's corpus. Historically, conflict arose between the Japanese culture boards and the boards that were both outside of and often uninterested in *anime* subculture, alongside which rose the popularity of the insult Wapanese, a portmanteau of wannabe Japanese or possibly white Japanese ("Wapanese", n.d.). Due to the contentious nature of the word and its overuse, a word filter was instituted on the website to change wapanese to weeaboo; a word that originated in a Perry Bible Fellowship comic strip coined by Nicholas Gurewitch, which in the original context implies something that is deserving of punishment ("Weeaboo", n.d.). The use of the term weeaboo has been prolific enough that it is the basis for terms to describe obsession with other cultures; koreaboo as someone obsessed with Korean pop culture, and teaboo as someone obsessed with British culture.

The current study is likely to be considered a borderline example of language mixing because of the strong attitudes, which requires analysis through linguistic theory that will explore social issues of ownership and ethnicity rather than bilingual competence, whilst also exploring less traditional language contact situations such as that involved in pop culture, internet, globalisation and the transmission of language across distance.

2.2 Research Paradigms

Within the research on language itself, as well as language learning and language change, there is a lot of debate about the terminology for describing the type of linguistic phenomenon focussed on in this current study; therefore the area is of particular interest to modern linguistics, and analysis of new case studies and language situations is important in order to add to the theoretical discourse.

Le Page and Tabouret-Keller (1985) emphasised the need to approach mixed language varieties from the view of coherent linguistic behaviour within a group of individuals that each have their own inherent views towards social and linguistic interactions and stereotypes, rather than from the view of there being an underlying system present and as such groups needing to be separated according to distinct language boundaries. Many sociolinguists (for example

Blommaert & Backus, 2011; Jørgensen et al., 2011; and Rampton, 1995) now prefer to refer to speakers' repertoires, being the sets of semiotic linguistic resources, styles, registers and genres that the speaker knows, which allows discussion outside of previous assumptions about how fixed the link between language, community, time and place is (Blommaert & Backus, 2011; Blommaert, 2010). Important to note also is how variable and often fragmentary different individuals' repertoires are (Rampton, 1995; Blommaert & Backus, 2011).

Whilst earlier approaches to code-switching and other contact linguistic topics focussed heavily on identifying underlying grammatical structures and attempting to create models that account for grammatical constraints affecting language production , more recent studies approach from a sociolinguistic point of view, such as above, focussing on language as a behaviour rather than an entity. The areas of lexical semantics and lexicology in terms of language borrowing have generally moved from analysing words at an isolated level of the word towards corpus-based linguistic studies in which the context of the word, in relation to other words and to the corpus as a whole, is emphasised (Zenner & Kristiansen, 2014). Analysing word forms and how they relate or differ to 'correct' language or grammaticality norms is still useful as a descriptive tool, but is less satisfactory when used as an approach to wider linguistic theory on its own.

The advantage of examining the data in terms of linguistic repertoires instead of base languages, is that the creativity of speakers can be observed, be it creativity within the language rules or creativity that manipulates or ignores rules. This creativity is also judged and restricted by the community involved, which agrees with other literature on similar language situations (Le Page & Tabouret-Keller, 1985; Varis & Wang, 2011). The inter-community policing of language crossing creativity therefore indicates that it is very important to be aware of distinctions in the data between conflicting linguistic behaviour or ideologies; the conflicts being between the behaviour the speaker believes is the proper standard, the behaviour the speaker believes they are performing, and the behaviour the speaker is actually performing (Le Page & Tabouret-Keller, 1985). Methodologically this means it is important to gather information about the sociocultural politics and views within the community the language occurs as well as examining the language features that are actually being produced.

More recent research follows and expands on this direction for analysing discourses of language mixing. Rampton (1995) coined the term language crossing to refer to the linguistic behaviour of code-alteration whereby speakers have in their repertoires and knowingly mix features (or resources) from different resource sets traditionally considered to be different languages, and often having varying amounts of linguistic competence in the different sets. Therefore, Rampton

(1995, p.289) sees crossing as a linguistic paradigm that "integrate[s] the analysis of code-switching with the study of second language learning." Rampton (1995) also notes that the linguistic behaviour involves the speakers encountering issues of legitimacy and ownership they need to reconcile with due to the crossing of social and/or ethnic boundaries. Rampton specifically looked at the crossing and stylisation of young people in multi-ethnic urban areas of Britain, investigating the speech practices of people with migrant backgrounds and then the spread of those particular practices among peers of other ethnicities and non-migrants. The languages involved were English-based Creole as used amongst Anglo and South Asian youth, Punjabi used by Anglo and Caribbean youth, and the use of South Asian English by all three. Crossing has also been used as a concept to explore the use of English in both Korean and Japanese pop culture, such as music, television shows, movies, and advertising (Lee, 2005).

Recent research has been undertaken to explore and further crossing, for example by Jørgensen (2008) who developed the term polylanguaging to describe urban youth language behaviours that Jørgensen claims are far more common than Rampton (1995) indicated. Specifically, case studies of Turkish-Danish mixing in Denmark are used to describe situations in which language contact leads to speakers' acquaintance or attention to the language, and then to the integration of those new features into in-group interactions. Issues arising due to ideas of access, ownership and legitimacy are continued to be discussed, and the relationship between individual speakers and a particular language, or set of resources, is seen as a socio-cultural construction that is negotiable (Jørgensen et al., 2011). Those who are considered native speakers are able to claim legitimacy and ownership over the language, while non-native speakers can claim access and varying degrees of legitimacy etc. depending on acceptance of their learning from others (Jørgensen et al., 2011).

Jørgensen's polylanguaging concept was created to slot into the long held linguistic frameworks of monolingualism, bilingualism, and multilingualism (Jørgensen, 2008), filling the gap where the terminology does not cover language use by speakers who are not, and are not expected to be, even minimally fluent. Therefore, the concept is related back to linguistic competencies and types of linguistic behaviour, rather than focussing on a metaphor of crossing social/ethnic boundaries. Jørgensen thus reiterates the perspective of Le Page and Tabouret-Keller (1985), where it is more useful, in particular cases, to analyse linguistic behaviours at the level of linguistic feature rather than language, as these speakers use whatever language features they have at their disposal as they wish, although not free from the pressure of social language norms.

However, Møller and Jørgensen (2009, p.150) distinguish between the minimal and contentious language elements where users are not expected to be fluent, and borrowed language units, claiming that "when speakers know what assignations and values come with the word" it is polylingual behaviour, but when that knowledge is lost or unknown it is a loanword. But, in order to make this evaluation a number of elements need to be investigated, therefore research into language use should not discard even what look like typical loanwords out of hand, as they are inherently intertwined within the history and description of language contact and change.

As Rampton (1995, 2010) identified the need for reconciling the sub-discipline of code-switching research with second language learning research, Backus (2014), in work based on Turkish-Dutch language mixing in the Netherlands, aimed to reconcile borrowing and code-switching behaviours as related elements of a wider process of language change. A usage-based approach of corpus linguistics is proposed to relate analysis of language as it is produced to Cognitive Linguistics and mental representations of language, including ethnographic aspects, such as attitudes and meta-discussion of language in order to see what could be constraining language use. Whilst there are methodological issues still to be had in investigating language knowledge and language use in terms of degrees of entrenchment, the overall concept of a usage-based approach has many advantages, particularly in the case of the current research where English-Japanese language mixing is found in Anglophone Japanese popular culture fandom spaces. All contact effects are seen as "aspects of a general process of language change" (Backus, 2015, p19), on different time scales; some synchronic, some diachronic. Therefore, the usage-based approach accounts for the shifting nature of linguistic features in both directions, and how it can vary both at an individual level and a community level as repertoires and contact with linguistic features, or language learning processes, differ between people.

Gee (2005) coined the theory of affinity spaces where people form affiliations primarily through shared practices or a common endeavour, and at a distance, such as through the internet rather than face-to-face interaction, and secondarily through things such as a shared culture, gender, ethnicity, or face-to-face interaction, and as such language learning is facilitated by sociocultural aspects rather than just classroom activities; which describes the community that uses Gratuitous Japanese. Fukunaga (2006), through the lens of affinity spaces, examines foreign language literacy development through Japanese popular culture such as *anime* (Japanese animation), which has also been investigated by Williams (2005); second language Japanese learning is positively affected by viewing Japanese popular culture materials, both from the interest in Japanese language and culture it implies and as a learning and reinforcement resource. This is of interest to the current study as popular culture is the main influence the

language use in question. Williams (2005) noted that students had picked up language elements and cues from *anime* etc even before starting language schooling. *Japanese* language students who view *anime* in the original Japanese with English subtitles (as opposed to a version where English speaking actors are dubbed over the original voice tracks) are exposed to and therefore gain a linguistic advantage in word recognition, listening comprehension and pronunciation, and awareness of various linguistic features such as *anime* related concepts, proper nouns, types of address, gendered speech, a distinction between formal and ordinary speech, slang, and tone of voice, as well as cultural knowledge such as Japanese social settings and rules (Fukunaga, 2006). It also must follow that viewers of subtitled *anime* who are not learning Japanese are nevertheless also exposed to these linguistic features; although then the question becomes, how do non-learners (or learners in conversation with non-learners) use this knowledge, if indeed they do.

Fukunaga (2006) also mentions the use of the internet for *anime*-related activities, where the internet is used to gather information about *anime* and Japanese culture, and to construct social identities as members of the affinity groups in question. Varis and Wang (2010) examined the super-diversity present on the internet in terms of a Chinese rap artist and his online linguistic interactions, and found that while the internet is a space that allows for huge diversity of culture and language use, interactions in internet subcultures or communities are still self-controlled by self-, peer- and state-imposed norms. Kytölä (2012) and Leppänen (2008, 2012), in studies of English-Finnish language mixing in Finland (or Finnish-based online spaces), have shown that normativity and prescriptive attitudes are no longer restricted to coming from traditional authorities such as government, journalism and academia, but also occur between peers as they negotiate the hierarchies of online spaces.

Discussions on the sociolinguistics of globalisation, in particular by Blommaert (2010), also support the theory that a person's linguistic competence, within and between languages, should be described in terms of repertoires and resources, as language features shift meanings and functions when transferred globally rather than being static in the country of origin. Therefore, there has been a methodological shift to change the view of language as tied to a specific community, time and place and thus only having local functions, to one in which, given globalisation, language is mobile (Blommaert, 2010). Jørgensen et al. (2011) reinforce the idea that social values and meanings associated with particular features become highly negotiable when transferred globally; as is the situation in the Gratuitous Japanese community where English-speaking individuals gain Japanese language features into their repertoires through contact with Japanese popular culture.

Therefore, the recent frameworks note the fluid nature of language in terms of language use, and want instead to analyse in terms of "situated uses of linguistic resource[4]" (Androutsopoulos, 2011, p 294), shifting focus from language as a thing that exists to instead emphasise language as a behaviour. They also consider the power tensions that arise and are navigated within an affinity space or community in terms of ethnicity and ownership, but also due to the translocality or global nature of the current world, particularly in online spaces.

2.3 English-Japanese Language mixing

There are not many studies that have focussed on cases of English-Japanese language mixing where Japanese features are mixed into English, as opposed to English in Japanese (Miller, 1998; Lee, 2005; Moody, 2006) and some research on bilingual language from Japanese Americans in North America (Nishimura, 1995; Azuma, 1997; Morimoto, 1999), other than perhaps classroom settings in Second Language Teaching research.

Stott (2006) analysed articles found in a group newsletter written by native English speakers for other native English speakers participating in the Japan Exchange and Teaching (JET) Programme, investigating the frequency and type of usage of what they considered code-switches (intrasentential, intersentential, tag) and whether Japanese ability, age, and article topic were influencing factors. Japanese words, phrases, or quotations were found in 77.8% of articles and 88.9% of authors, and the suggested reasons for code-switching included being used for narrative devices like framing and for humorous effect and to show in-group membership and solidarity. Stott (2006) also noted the negative attitudes the English teachers had towards code-switching between Japanese and English by Japanese people (colloquially labelled Japlish), despite their own code-switching behaviours.

Stott (2006) advises it can be hard to define shorter instances of language alternation as either code-switching or lexical borrowing, as words borrowed by a particular group may not *yet* be recognised as loanwords outside of the group, or even by all members of the same group. Therefore, Stott is one of many recent researchers who suggest it is more useful to consider code-switching and borrowing as both part of a continuum.

Turner (2015) investigated code-switching in Japanese *anime* fandom-based fanfiction, using corpus linguistics to analyse whether there were differences in frequency of word classes

[4] Linguistic resources being linguistic elements such as sounds, words, and patterns, but also any other communicative and semiotic elements that are available to speakers (Blommaert & Backus, 2011).

between writers with some degree of formal Japanese language education and writers who did not. Interestingly, there was no significant difference which indicates there are different factors within the community that govern the extent and types of Japanese language used other than language ability. Turner (2015) also found that the usage of code-switches was often inconsistent, both in terms of switching freely between the Japanese form and the English equivalent and in terms of the number of code-switching occurrences varying across multiple texts from the same author.

An important aspect of Turner's (2015) research design was being able to directly investigate reasons and attitudes behind the language choices of all the authors who made up the corpus through a questionnaire. Positive influences on language use included believing that particular Japanese elements would give life to the narrative and characterisation and cases where meaning would be lost in translation. Negative influences included the fact that while familiar code-switches would be acceptable, unfamiliar terms would be considered confusing or disruptive, and that too many Japanese words may come across as showing off. Based on the information gathered, the writers appeared to be policing their own language use based on how they thought the community would view them, which restricted the Japanese elements used.

Following on from and influenced by the discussion of previous research and theoretical paradigms, the next section will cover the methodological approach and research design of the current study, including descriptions of the data sources and data gathering process, and how the resulting corpus was analysed according to language features.

3 METHODOLOGY

This chapter outlines the approach to collecting and analysing Computer Mediated Communication data, by explaining the sources the language data used in the current thesis was collected from, including describing the history and structure to serve as contextualisation for the findings and the rationale behind the choice of each source; establishing the parameters of the data gathering and preparation process; and explaining the tagging system created to sort the language data along with the tools employed for data analysis.

3.1 APPROACHING CMC DATA COLLECTION

Computer Mediated Communication (CMC), whilst historically referring to communication facilitated through the medium of a computer, also encompasses interactions using other communication technologies, such as mobile phones (Herring, Stein, & Virtanen 2013). CMC is predominantly a text based medium, but also includes audial and visual modalities; including in the ongoing incorporation of images, sound, and videos on Internet websites where emoticons, icons, images, music, and videos are either embedded within text for creative reasons or to clarify mood, or are in fact the main point of communication such as on websites like Youtube where text is the subordinate modality (Bodomo, 2010).

As the Internet is a wide, complex, and diversified space, and CMC varies strongly in terms of the available modalities and formats of communication across and even within different websites (Kytölä, 2012), it is therefore important in linguistic research to describe and contextualise the spaces where linguistic data is collected from as comprehensively as possible. Following from this, studying the language and context from an ethnographic, or insider's perspective, if possible, is useful in discovering knowledge about the background of both the writers and the website itself; and the connections both have to other online spaces (Lee & Barton, 2012). Ethnographic understanding of the space and data under examination is also of particular importance in being able to make well-informed ethical decisions on the usability and anonymization of the data (Kytölä, 2012). One of the advantages to studying CMC is the public availability and thus easy accessibility of a large amount of written linguistic data, including colloquial and spontaneous[5] language, but despite this public availability, the academic analysis

[5] Written language is generally seen by nature more planned than spoken language, but internet users can "write fast 'in the flow', without resorting to sources of help" as well as producing carefully constructed forms (Kytölä, 2010, p. 109).

of forum content involves the movement of that content from a colloquial space where it is expected to be viewed by individuals who are (even if not active contributors) expected to be aware of community rules, into an academic setting where it is being viewed by people who are, by default, outsiders.

A major approach in modern linguistics[6] to researching CMC is that of Corpus Linguistics, which aims to study language as it appears naturally in pre-existing spaces of communication, allowing for comprehensive analysis of language patterns in the specific context of where they occurred (McEnery & Hardie, 2012). Observed language, spoken or written[7], is gathered into large data sets, or corpora, which is then analysed with the help of corpus tools. Analysis of corpora is generally conducted both quantitively and qualitatively, as the frequencies and linguistic patterns found in a corpus need to be analysed in context with the aim of being able to interpret and explain the presence of those patterns (Biber, Conrad, & Reppen, 1998).

To facilitate a comprehensive qualitative and quantitative approach, alongside the incorporation of ethnographic insights from the researcher, (Androutsopolos, 2008) it is also useful to combine corpus data with attitudinal data if available (Zenner & Kristiansen, 2014), which can be gathered through the meta-discussion of writers as they discuss acceptable and unacceptable behaviours, including interactions where they directly police the behaviour or language of others.

3.2 SOURCES AND DATA COLLECTION

There are many high traffic websites where English-Japanese language mixing can be publicly observed; from social media sites such as Twitter (twitter.com) and Facebook (facebook.com), to forums and other sites like Tumblr (tumblr.com). However, many of these sites posed methodological issues for linguistic data collection, particularly in terms of creating a corpus. Websites that could have a snapshot of the entire site section for a specific time frame gathered, including all discussion and not just individual Japanese language instances, were prioritised as a potential site for data collection. Other criteria used to identify sites for collection were a high volume of website traffic and popularity; the content being worksafe[8]; featuring discussion related to Japanese culture fandom topics, indicating definitive exposure to and interest in

[6] Corpus linguistics, although not labelled as such at the time, was also a common approach in early linguistics, even without current day analysis tools (McEnery & Hardie, 2012).
[7] Or visual, in the case of Sign Language.
[8] No explicit or pornographic content; meaning content that is appropriate be viewed on a workplace computer without risk of censure.

Japanese products; being an English language website; and researcher familiarity with the website and website culture for ethnographic considerations (see Chapter 1).

The two websites chosen that fulfilled all these criteria are the popular (and controversial) imageboard '4chan' (boards.4chan.org), and the Troper Tales subsection of the large fan culture wiki[9] website TV Tropes (tvtropes.com). 4chan contains numerous sub-boards on topics divided into the broad categories of Japanese Culture, Video Games, Interests, Creative, Adult, and Other and has an anonymous userbase that functions without user accounts. TV Tropes contains registered writer contributed content that catalogues the many tropes[10] found in fiction, whilst the Troper Tales subsection was where many of the same writers shared their real-life examples or experiences with those tropes.

4chan is the primary source for data as it features organic language use as well as a small amount of linguistic meta-discussion, whilst Troper Tales instead features accounts of language use and, more importantly, a large amount of meta-discussion. Both sources feature content that has been publicly, and in many cases anonymously, posted to the internet, and are described in further detail in the following sections.

3.2.1 /CGL/

The main set of data gathered for the current study is from the English language image-based bulletin board website 4chan.org, which was originally based on the Japanese language image board Futaba Channel also known as 2 Channel or 2chan.net (4chan, 2016). According to 4chan.org, the website in its entirety has over 680 million-page views and over 22 million unique visitors per month; and according to Alexa Internet, a company that monitors web traffic data, its global rank is 219 (Alexa Internet, 2017)[11]. Originally, the website was dedicated to Japanese culture, but it has expanded over the years to now also encompass many other topics, including photography and other creative works; television and film; fitness; travel; and the paranormal.

Each topic has its own separate board, with its own moderators (called janitors), rules, and norms, and as such, although there are definite trends across the entire site, each board can be considered a different community[12]. Each board also varies in terms of their degree of worksafe-ness, as some boards are entirely pornographic or other explicit content, whilst others either

[9] A collaboratively constructed knowledge base website (Wikipedia, n.d.).
[10] A significant or recurrent theme (Oxford English Dictionary, n.d.)
[11] For context, Facebook's Alexa global rank is 3, Twitter's is 11, and Tumblr's is 47 (Alexa Internet, 2017).
[12] There is user crossover between boards, but it is not necessarily expected.

allow such content alongside other discussion or disallow it entirely. Whilst each board is moderated according to its own unique set of rules, there are also rules that universally apply to the entire website. For example, the website is only meant to be accessed by writers who are 18 years or older (although this is hard to enforce as the website is an anonymous one); there are no registered writers; and, aside from a small number of tripcode writers (a type of pseudo-registration that allows a username to be displayed on a person's posts[13]), posts are labelled as having been authored by Anonymous, although the site does log the number of distinct posters in each thread versus the number of posts.

The anonymous structure of 4chan lends a general sense of unaccountability to the community, and 4chan has been described by the media as a "…lunatic, juvenile community…" that is "…brilliant, ridiculous and alarming…" (Michaels, 2008, para. 5). 4chan is widely considered to be a significant hub in the creation of and distribution of internet memes[14], including rick-rolling[15], lolcats[16]' and even the term weeaboo itself (see Chapter 3). Whilst the anonymous nature means that theoretically writers could post content and creatively use language freely without direct backlash, it has instead developed a culture where broader social hegemonic structures are not exactly challenged or subverted, and fascist, racist, sexist, and ableist attitudes are a common theme, although to differing degrees across the different boards. 4chan serves as "…a constant reminder to fans on the Internet that their behaviour is being monitored, observed, and judged" (McGee, 2012, p.60), both to writers in other online communities as well as those on 4chan itself. Therefore, as 4chan is a space in which linguistic policing is present, the level of English-Japanese language mixing observed is hypothesised to be more restricted and thus less complex than what may be present in other fandom spaces; and the aim of the current study is to establish an analysis of what could be considered a baseline degree of English-Japanese language mixing in Japanese culture fandom spaces, to which more prevalent usage in other spaces can be related back to or measured against.

The current study has narrowed the focus to a single Japanese culture board, titled Cosplay and Gothic Lolita, styled as the acronym /CGL/. The board features discussion and image sharing on topics related to *cosplay* (or costume play) and sewing and prop techniques, as well as a wide variety of *j-fashion* styles, (or Japanese street fashion) including the titular Gothic Lolita. Whilst

[13] Usage of 'tripcodes' is generally seen as against the ideology of the board and is frowned upon.
[14] defined as "a piece of culture, typically a joke, which gains influence through online transmission" (Davison, 2012, p. 122).
[15] Linking to the song 'Never Gonna Give You Up' by Rick Astley under false pretences; tricking users who are expecting different content.
[16] Humorous pictures of cats with misspelt or grammatically incorrect captions.

the board is moderated against off-topic posts or threads, the strictness of such is at the discretion of the janitor. Keeping in mind the anonymity of board writers, and the wide range of different boards on vastly different topics, 4chan describes the general demographic of the website in Figure 1 below. Although there are no individual board statistics, according to ethnographic observation of the community, /CGL/ appears to have a far higher ratio of female writers than other boards. There are occasionally threads that focus specifically on male participants in *cosplay* or j-fashion, which is one indicating factor supporting that female writers are the majority, as opposed to the rest of the website.

> Age: 18-34
> Gender: ~70% male, ~30% female
> Location: United States (47%), United Kingdom (8%), Canada (6%), Australia (5%), Germany (4%), France (2%), Sweden (2%), Netherlands (2%), Poland (1.5%), Brazil (1.5%)
> Interests: Japanese culture, anime, manga, video games, comics, technology, music, movies
> Education: Majority attended or currently enrolled in college
>
> (4chan, 2016)

Figure 1: Demographic characteristics of 4chan users

/CGL/ is a 'blue board[17]', meaning it is worksafe and contains no explicit (i.e. pornographic) content, which was part of the reason that /CGL/ was chosen over other Japanese culture boards, such as /A/, the board for *anime* and *manga*, which does contain explicit content. /CGL/ was also chosen in order to examine the language use in Japanese culture topics outside of specifically *anime* and *manga* contexts as previous research has done (for example, Bloem, 2014; Turner, 2015), although those contexts are still present within /CGL/. Other advantages for examining /CGL/ language use include researcher familiarity as opposed to other boards or websites, which, as previously mentioned in Chapter 3.1 above, helps influence insight into the behaviours, norms, and functions of language use beyond the relatively small cross section of data gathered. Such ethnographic considerations, where the researcher approaches the study from within the community, are important in identifying and understanding often specific, nuanced meanings, and language values (Androutsopoulos, 2008).

The structural properties of the /CGL/ board is asynchronous, meaning writers post at their own convenience, and can contribute to both new and old conversations; multi-authored by mostly anonymous writers; public as there are no requirements for gaining access to the content; and interactive, as writers contribute content and engage in discussions with each other. There is a

[17] Specific to 4chan; worksafe boards are indicated with a blue background theme.

finite number of threads that can be present on the board at one time (circa 150) and these are organised by current activity, with the threads of the most recent posts being moved to the top of the list (see Figure 2 below for an image of the board structure). When a new thread is created the least active thread that is on the bottom of the list 'drops off the board' to make space for newer threads. Threads also have an imposed size limit, and those that reach 300 posts or 150 images stop being moved to the top of the list when new posts are made until they too drop off the board. 'Finished' threads are stored in a short-term archive hosted on the board before being deleted. Thus, there is a high thread turnover, although /CGL/ is considered a slower board in comparison to others on 4chan, with threads tending to remain for days as opposed to only hours on faster boards.

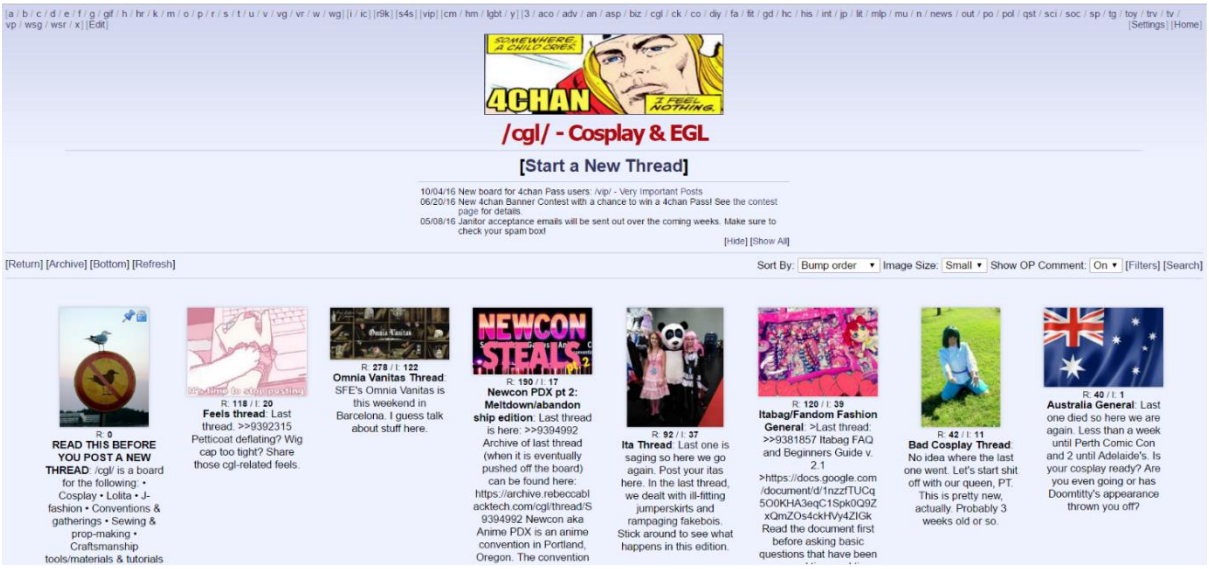


Figure 2: Partial screen capture of the top of the 4chan.org/cgl thread 'catalog' page

Figure 3 below details the general format of threads on 4chan. Each post is headed with name[18], date, time, and post number metadata, with the first post including the thread title. Whilst other forums may employ a quote function where a writer can include the text of the post they are replying to in their own post or tag the specific writer(s) they wish to converse with, 4chan uses numbers to link posts with their replies and vice versa (see Figure 3 below). Each post's number is shown on the top right corner, and replies to a specific post begin with that number in-text (indicated with rectangles as shown in Figure 3), achieved by either clicking on the post number and responding to a reply box pop up, or by copy and pasting it into an already open reply box. A post that has been replied to is also indicated by the presence of the post numbers of the replies in the bottom left corner (indicated with ovals as shown in Figure 3).

[18] Unless the writer employs a 'tripcode' the name on the post defaults to *Anonymous.*

Figure 3: Diagram of the /CGL/ thread reply structure[19], showing post numbers and how they link together

The previously noted 4chan forum structure of threads being deleted rather than indefinitely archived may have posed a methodological issue in terms of data collection, if not for the existence of a third-party archive system (archive.rebeccablacktech.com/cgl/) specifically set up to save /CGL/ threads in chronological order, from the date the original post was created. It is from the third-party archive that the data for the current study was gathered.

All threads that were started within the date range of May 1st to 15th 2015 were accessed through their individual pages on archive.rebeccablacktech.com/cgl/ and the data for each thread copied and pasted into a separate text (.txt) document, as per McLellan (2005, as cited in Kytölä, 2012) and Turner (2015), in Notepad++. For convenience purposes, threads with fewer than five replies were not saved because they were likely to have complicated, and unnecessarily prolonged, the collection process; and also because these threads were either deleted by forum moderators soon after creation due to the violation of rules (as the moniker janitor implies, discretion in 'cleaning up' off topic or obvious trolling[20] posts is used); or received little to no attention from /CGL/ community members and quickly dropped off of the board. In either of these above two cases, these small threads were not considered particularly useful

[19] Screen capture from the mobile website version.
[20] Where someone is pretending to be sincere in order to cause disruption (Hardaker, 2010).

representations of the community, by either the researcher or the community itself. Cases of short threads that received little attention from users included questions and image or anecdote requests that were ignored, as requests that do not also include some content contribution are, somewhat paradoxically, looked down upon; or told by other writers to ask again in an already present thread rather than create an entirely new thread. As the 4chan website framework limits the number of threads that can run concurrently, and /CGL/ is considered to have a slower rate of posting compared to other boards, there is a community preference to limit extraneous threads that include subjects that do not lend themselves to long lasting discussion or content contribution.

The data saving itself was methodologically simple, perhaps in comparison to other types of forums, as each thread's content, in its entirety, was already in a separate page, rather than across multiple pages that would need to be saved separately or compiled.

To further narrow the amount of data to a manageable amount for the purposes of this study, the 222 threads saved were sorted randomly into a final corpus of 100 threads. Each text document included the web address information in the name, and was labelled according to whether the thread was predominantly *cosplay* focussed or j-fashion focussed; or left unlabelled when it was a mixture, such as with threads discussing a convention where both j-fashion and *cosplay* were going to be worn by different members. The distribution of j-fashion, *cosplay*, and unspecific threads after being culled to the final 100 threads resulted in relatively even numbers, with 35 *cosplay* threads, 31 j-fashion threads, and 34 unspecific threads. However, the total word counts across these three categories differ, with the *cosplay* threads combining to 90698 words; the j-fashion threads to 138286 words; and the mixed threads to 174988 words.

To prepare the text documents for corpus analysis the remaining forum structure noise, such as headers and footers left over from hyperlinks, were manually deleted. Post numbers and post date and time information were also deleted using the find and replace function in Notepad++[21] to achieve a more accurate word count figure, as many wordcount tools count number strings as words. However, the removal of this type of forum structure noise was completed after the tagging process, meaning a copy of each thread document was able to be saved with the post numbers and dates intact alongside the corpus tags for analysis purposes where it was useful to retain information related to the flow of discussion via comment and reply order.

[21] Consequently, any numbers present in-text in thread comments were also deleted.

As previously mentioned, 4chan is an image board site, and, in the case of /CGL/, many of the images posted featured text. However, due to methodological issues, whilst the pictures were originally saved alongside the text, this multimodal aspect of the forum has not been retained in the data analysed, because counting all text, including large blocks of English text and Japanese orthography to add to the total word count of the corpus, would have been difficult. The images included meme response pictures, as exampled in Figure 4 below, with tokens of language produced by members of the 4chan community (or other online communities) and images created to help other members decipher Japanese orthography, which would have been useful additions to the corpus. Other images featured tokens produced by mainstream sources, such as scans of Japanese fashion magazines, which, whilst not necessarily prudent to include in the corpus, are still important to note the presence of, as it illustrates the community members either interacting with, or at least being exposed to, the Japanese language.

Figure 4: A Japanese language variation of a common meme posted in response to trolling

Kytölä (2012) critiques McLellan's (2005, as cited in Kytölä, 2012) argument for converting writer discussions on forums into a text-only file which then allows for the easy use of corpus tools because it affects or entirely removes the multimodal nature of the source forum or website. Both the images, which include memes that are an integral part of the conversation where included, and the "…interactional, conversation-like features" (Kytölä, 2012, p. 124) as indicated by post reply structures, have been removed in the preparation for corpus analysis, but given the focus of the study on individual language tokens and the amount of data collected, this is a necessary limitation. And, as previously mentioned, copies of the /CGL/ forum data with all multimodal structure intact have either been saved separately from the text files or are accessible in some way, although the scope of the current study does not include a detailed analysis of specific writer reply interactions.

Overall, the resulting corpus gathered from 4chan.org/CGL/ consisted of 403,972 words.

3.2.2 TROPER TALES

TV Tropes (tvtropes.org) is a large wiki website dedicated to cataloguing tropes found in many forms of media and where site writers collaborate to create a comprehensive, although generally with an informal and humorous tone, description of a trope as well as presenting gathered examples of those tropes as found in various media sources. TV Tropes' Alexa Internet (2017) global rank is 1924.

One of the tropes described and discussed is Gratuitous Japanese. According to TV Tropes, Gratuitous Japanese refers to a typified mixing of Japanese and English by the *Anime* Fanfiction Community, where Japanese is inserted into *anime* based English language fanfiction[22] to varying degrees of success by writers who are generally perceived to have low Japanese language competence. According to TV Tropes, the inclusion of Japanese honorifics and kinship terms, and a small number of routine phrases and exclamations that are considered likely to be familiar to even a casual viewer of *anime* or *manga*, is acceptable, but more obscure terms or large sections of text are less so. An even more disdained subsection of the trope is the use of the language in everyday life, online or offline, as opposed to only in fanfiction writings The English-Japanese language mixing behaviour described here, although seen in a generally negative light with many strong value associations displayed, is the same linguistic behaviour that is the focus of the current research, and therefore, discussion of Gratuitous Japanese on TV Tropes is a useful source of data. Furthermore, Lee and Moody (2012) argue that in examining a group's attitudes towards a subject, for example popular culture, it is necessary to take into account definitions and discussions found in sources that matter to that group, even if they are sources that would not be considered by academic circles to be official.

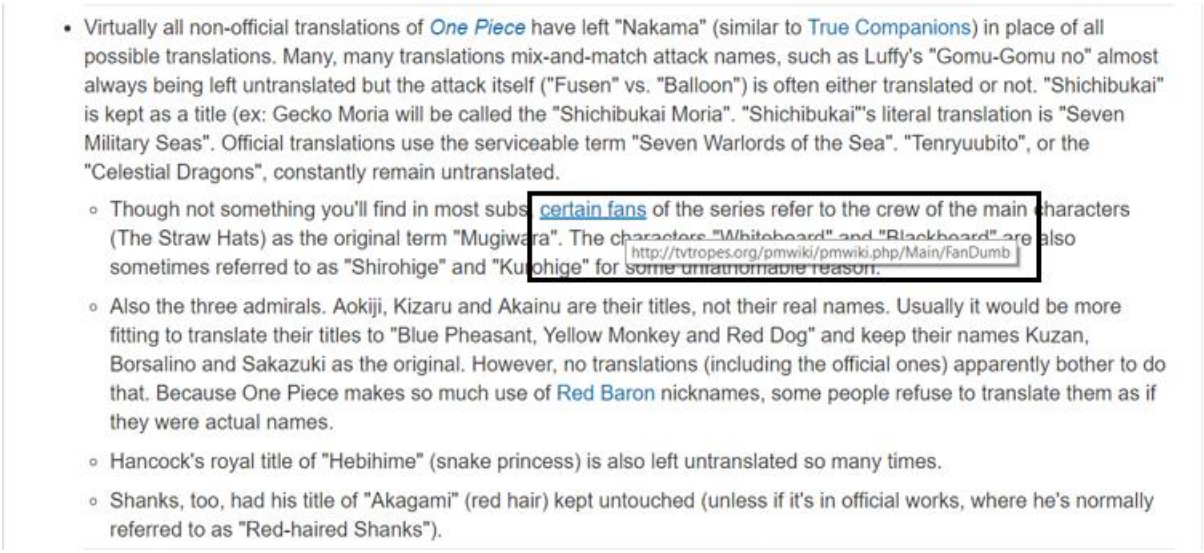
Troper Tales was originally a subsection of the TV Tropes website where writers, called tropers, could share their own experiences of the tropes in real life, and which include their opinions on both those experiences and the trope itself. Troper Tales was considered controversial in that the discussions of real life tropes by the site writers quickly became increasingly unbelievable and inappropriate, creating a community reputation that discouraged new members; and which other sections of the wider TV Tropes community felt reflected negatively on the site as a whole (InfiniteZenith & NinjaClown, 2013). Whilst the site was in use for several years, in 2011, due to conflicts about the content, and precipitated by Google removing their advertising due to the presence of not worksafe[23] materials, Troper Tales was moved offsite (InfiniteZenith &

[22] Fanfiction refers to fan written stories based on media such as television shows and books.
[23] Discussion of mature or sexual themed tropes.

NinjaClown, 2013). There is a caveat on the usage of this data given its age, although it is not old enough to be obsolete. Troper Tales no longer has a presence on the internet in website form because the host website the Troper Tales content was moved to was completely shut down in 2012. However, some pages from the site were saved and subsequently uploaded to an online library (scribd.com) in a combined text document format, including the Gratuitous Japanese section.

It is important to note that the content saved from the now defunct Troper Tales site was titled "Shit That Didn't Happen" (shitthatdidnthappen.txt) by the writer who created the document, which ties into the culture and development of the website that led to it firstly being moved off TV Tropes, and subsequent removal from the Internet.

The format of the website is asynchronous in that writers add content when they wish this content is not time- or date-stamped; it is slow moving as the allowable contributed content is heavily restricted as with other wiki type sites, although Troper Tales was less restrictive as it was less formal than the greater TV Tropes site; the content is multi-authored by registered writers, but contributions are added as collaborations to a larger text where writer names are not directly attributed; and interactive only to a degree as writers can expand on, or relate to another writer's contribution as a sub-point (see Figure 5 below), but not have ongoing conversations. Another feature that is unique to the website is indicated by the black box in Figure 5 below, which shows how writers embellish their contributions with words or phrases that are turned into hyperlinks that link to other trope pages, making a subtle but pointed commentary as an aside. Figure 6 below also shows how that feature is preserved in the text document remnant of Troper Tales, with the hyperlink itself lost but the trope name remaining alongside the originally embellished text within the double square brackets '[[ ]]'.



- Virtually all non-official translations of *One Piece* have left "Nakama" (similar to True Companions) in place of all possible translations. Many, many translations mix-and-match attack names, such as Luffy's "Gomu-Gomu no" almost always being left untranslated but the attack itself ("Fusen" vs. "Balloon") is often either translated or not. "Shichibukai" is kept as a title (ex: Gecko Moria will be called the "Shichibukai Moria". "Shichibukai"'s literal translation is "Seven Military Seas". Official translations use the serviceable term "Seven Warlords of the Sea". "Tenryuubito", or the "Celestial Dragons", constantly remain untranslated.
  - Though not something you'll find in most subs certain fans of the series refer to the crew of the main characters (The Straw Hats) as the original term "Mugiwara". The characters "Whitebeard" and "Blackbeard" are also sometimes referred to as "Shirohige" and "Kurohige" for some unfathomable reason.
  - Also the three admirals. Aokiji, Kizaru and Akainu are their titles, not their real names. Usually it would be more fitting to translate their titles to "Blue Pheasant, Yellow Monkey and Red Dog" and keep their names Kuzan, Borsalino and Sakazuki as the original. However, no translations (including the official ones) apparently bother to do that. Because One Piece makes so much use of Red Baron nicknames, some people refuse to translate them as if they were actual names.
  - Hancock's royal title of "Hebihime" (snake princess) is also left untranslated so many times.
  - Shanks, too, had his title of "Akagami" (red hair) kept untouched (unless if it's in official works, where he's normally referred to as "Red-haired Shanks").

Figure 5: Screenshot of the TV Tropes webpage to show content text structure

```
* I have a friend who frequently uses "baka" as an adjective, and then[[HypocriticalHumor calls me out whenever I
speak a foreign language because "this is America".]]
** I regret ever teaching my nooby friends ''baka''. They used it too much and IT DROVE ME CRAZY. Of course, I do
sometimes mumble to myself in Japanese. I used to say ''kuso'' alot but it just morphed into "ksch" eventually.
*** Hey, I do that "ksch" thing too! And yes, it also came from "kuso".
```

Figure 6: A sample example from the shitthatdidnthappen.txt document

Due to the nature of the Troper Tales page, the language found is all reported speech rather than naturally occurring in written conversation. The writers contributed discussion on where and when they had encountered, or they themselves used the language, and their opinions on the appropriateness of that language. While it is interesting to be able to see the types of speech that writers claim they use in comparison to the actual language reality that is found in another data source (such as the /CGL/ corpus, in this study) with the caveat, of course, that it is not a comparison across the same writers, the main usefulness, and thus purpose within the current study, of the Troper Tales Gratuitous Japanese discussions as a source of linguistic data is in discovering and analysing the attitudes towards language use that are prevalent in the general community, taking advantage of meta-discussion by community members that is already available online rather than elicited with a questionnaire.

Because the comments on Troper Tales are in response to an already contentious ideal of stereotypical language use, writers are particularly frank in revealing their negative as well as positive or neutral attitudes, which is of particular importance and interest for research into the language behaviour attitudes. Kytölä (2012) noted, as a caveat on the usefulness of questionnaires, that a person is unlikely to report negative or particularly hateful attitudes in surveys or interviews but may exhibit direct hate-talk about certain language use when

encountered online. In the case of Troper Tales, the discussion found seems to fit somewhere in between a formal questionnaire and meta-discussion in response to language directly encountered.

As the main corpus of language data collected for the current study is from a forum with an anonymous writer base with few identifiable writers (4chan.org), and also because of the disjointed nature of the gathered threads[24], it would only be possible to gather *general* language attitudes with a questionnaire as opposed to specific reflections by the writers featured in the data as there would be no way to guarantee those who answer a questionnaire are the same as those whose linguistic behaviour is observed in the corpus, unlike, for example, Turner's (2015) study where specific writers of fanfiction texts were the focus, and who were more easily able to reflect back on those specific texts. This methodological disadvantage of the /CGL/ data is therefore intended to be minimised by the inclusion of the Troper Tales data set, which consists almost entirely of meta-discussion on English and Japanese language mixing, to complement a smaller amount of language attitude data found in the /CGL/ corpus.

Meta-discussion on both the use itself and the situational appropriateness or acceptability of specific language behaviours is an important contextualisation cue for researchers and their analysis of those language behaviours, situating observed language use within the wider context of the community's attitudes and ideologies (Kytölä, 2012). 4chan and Troper Tales are not connected or specifically related websites, but as described, they both have a somewhat similar contentious reputation, and both sites fit into the wider online community of Japanese popular culture fandom. An aim of the current research is to present English-Japanese language mixing as a type of linguistic behaviour occurring in the wider sphere of the Anglophone Japanese fandom affinity group and establishing the general trends of such, rather than presenting it as a feature tied to a single website community.

The Troper Tales document was also prepared for use as a corpus by being cleared of extraneous formatting (such as the asterisks '*' shown in Figure 6 that replicated the bullet point structure in Figure 5) and tagged, resulting in a final word count of 11251 words. However, as previously mentioned, the scope of the current research has limited analysis to identify language attitudes rather than commenting on the linguistic makeup of the specific Japanese language tokens.

[24] 100 random threads on different subjects chosen from 222 gathered within a specific time frame, and a likely different writer base presence across those threads.

3.3 TAGGING AND ANALYSIS SYSTEM

The tagging system created for, and from, the language data found in the /CGL/ corpus was superficially modelled off Turner (2015) (although modified to encompass a wider range of features) who in turn based their system off the tags used by Leech and Nicholas (as cited in Turner, 2015) in the British National Corpus and the CLAWS part-of-speech tagger tool ("CLAWS part-of-speech tagger", as referenced in Turner, 2015). The similarities between the current study tags and Turner's tags, and where they diverge are outlined in Appendix 1. Each tag in the current study entered into the corpus document was attached to the language token with an underscore symbol '_' and consists of a unique symbol or letter combination that was created for both functionality and descriptiveness requirements.

The resulting tag set has a hierarchy involving four distinct levels of tag groupings: *Orthography*, *Part of Speech*, *Feature*, and *Attitudes* (see Appendix 2 for a comprehensive list of the final tag set, including descriptions, explanations and examples). The hierarchy was to facilitate different levels of analysis as well as the accurate counting and separation of the tagged data as many tokens in the corpus required multiple tags, as seen in Examples (1) and (2) below, which would hinder the ease of frequency analysis. Each token of Japanese language found was first tagged with a single tag from each of the first two tag groupings (Orthography and Part of Speech), whilst the other two sets (Feature and Attitudes) were only used where applicable.

(1) >kawaii_^_JPAD_JL desuuuu_^_JPV, sugoi_^_JPIE!!!_JF _DS

(2) *Bad enough that this is the usual 'kawaii_^_JPAD_JL maido_^_JPN_JMPR_JF' outfit I see at the con*

Further tag grouping outside of the four main levels (Orthography, Part of Speech, Features, Discussion) was established through the tag naming conventions to link related tags together. Similar to Turner (2015) the tags were created to be as buildable as possible for searchability within the corpus. In Figure 7 below, '*' indicates that there are related tags that build upon that base; for example, whilst the Noun word class is split into both nouns (_JPNO) and attributive nouns (_JPNA) in the final tag set, they can also be searched for and analysed as a unit if required (_JPN).

_J*	All Japanese related tags
_JP*	All parts of speech
_JPH*	All types of honorifics
_JPN*	All types of nouns
_JN*	All Japanese name types
_JM*	All Japanese modification features
_E*	All English related tags
_EM*	All English modification features
_D*	All Discussion tags

Figure 7: Break down of the tag set and how it builds on itself

Due to the interlinking nature of many of the tags, and the hierarchical structure within the tag set created to counter any data skewing that would be the result of tokens having multiple tags and thus multiple counts, the tagging system had to be established before the main text tagging took place, which is in contrast to Turner's (2015) tagging where tags were created as each element was discovered in the texts, rather than directing the discovery of code-switches according to previously defined tags. However, as the first step of the tagging process was to identify the presence of Japanese language tokens (with one of the three straight forward orthography tags) before any judgement on the word class or possible feature was made, the resulting familiarity with the word classes found before formal tagging was not seen as a limitation.

The orthography tags labelled the Japanese language tokens as being written in the Japanese alphabets of *kanji* or *kana*, the Roman alphabet (*romaji*), or if it was instead a Kaomoji (Japanese style emoticons created through punctuation symbols). Tokens in *romaji* are the focus of this thesis and are thus the only tokens that are categorised further, partially due to the language competence limitations of the researcher[25] (see Chapter 1), but mainly due to a specific interest in the prevalence of this form of alphabet, as the usage of the Roman alphabet requires less Japanese language competence than the Japanese specific *Kanji* or *Kana* alphabets. The symbol

[25] I cannot comment on the lexical or grammatical forms of Japanese words in the *kanji* or *kana* as I cannot read them.

'^' was chosen as the form for the *romaji* tag, because it stands out in large blocks of text, and the few natural instances of the symbol found in the corpus as the *kaomoji* '^_^' were manually excluded from the frequency count.

The Part of Speech tags, the only other compulsory tag aside from an Orthography tag, were, as per Turner (2015) mostly[26] defined based on their function in the English sentence rather than how they would be classified in a Japanese dictionary, as most tokens were integrated according to English grammar. For example, the tag _JPV also includes Japanese nouns used as verbs, such as in Example (3) below, where the noun *waifu* (wife) is used as a verb to indicate the action of choosing a favoured character or person as a fictional wife.

(3) *10/10, would waifu_^_JPV_JMPR*

Japanese names, including those of real people or characters, place names, and *anime*/*manga* series or Japanese company names, were also tagged as they were present in the data, but are easily able to be separated from the other data during analysis. The _JN tag was separated into _JNC for character, person, or place name, and _JNS for series name as *anime* or *manga* series names often have English translated versions that are frequently used interchangeably with the Japanese name, which was a potential area for analysis.

The tagging itself was done manually as an automated tagging tool that recognised Japanese words that were often altered in some way alongside English words was not available, and the results would have required extensive proofing due to ambiguity in how the words were being used. Manual tagging also engendered familiarity with the distinct words and linguistic patterns that were encountered in the corpus. The feature of manual tagging also influenced the tag names as they needed to be both descriptive and simple enough for remembrance and ease of input; for example, _JPHS spells out the shorthand Japanese Part [of speech] Honorific Suffix and _JMPL spells out Japanese Modification Plural.

Some of the decisions for which tags were used in certain instances needed to be made on a case by case basis. Tagging for Japanese plurals (_JMPL) was sometimes hard to keep consistent because of the presence of loanwords and collective nouns. The implication of Japanese plurals is discussed in Chapter 4. Compound words were often inconsistent in form, and as such appeared as both a single word and separate words. For tagging purposes, the single forms were separated, such as in Example (4) below, where the fashion style name *himegyaru* [princess gal]

[26] See Appendix 1 for exceptions.

is tagged as two tokens. *Hime* [princess] is taken as modifying *gyaru* [gal] rather than *himegyaru* being an entire unit because *gyaru* is also present by itself or in other compound iterations.

(4) *I wear gyaru_^_JPN_JC, hime_^_JPAT gyaru_^_JPN_JC to be precise, so I get a lot of varied comments*

Linguistic corpora are predominantly used for quantitative analysis, however as Biber and Conrad (2001) note that they can also be used for qualitative research by examining instances that frequency analysis has noted an interesting finding in context. The purpose of tagging for features other than part of speech was to maximise the usability of the corpus as a qualitative resource by allowing interesting instances of language and language discussion to be identified in the first read through of the corpus and made searchable for future analysis. The attitude tags, which were only used where applicable, highlight discussion in the corpus that is of interest for the analysis of community ideologies, such as meta-discussion on Japanese language use (_DM: Discussion Metalinguistic) or instances of language in-text that introduces or reinforces stereotypes found within the community (_DS: Discussion Stereotype).

One language identification decision that consciously differed from two of the few other studies examining English-Japanese language mixing[27] (Stott, 2006; Turner, 2015) was the inclusion of loanwords; defined as such by their presence in the Oxford English dictionary. Stott (2006) and Turner (2015) entirely excluded loanwords from their data as their studies were focused specifically on code-switching, and loanwords present in official dictionaries can generally be acknowledged as being far closer to the borrowing side of the language mixing continuum than the code-switching side. The current study, in contrast to Stott and Turner, is focused on examining all Japanese language behaviour observed, and determining what language classifications can be applied afterwards. Loanwords are, however, acknowledged with the feature tag _JL to be able to filter them out when desired for analysis.

3.4 TOOLS

The tags were manually added to each token identified in Notepad++ according to the analytic typology discussed above and catalogued in Appendix 2. Notepad++ is free text editor that allows for easy opening and navigation of large numbers of files in different tabs within the same window, as well as editing across those multiple tabs. Notepad++ contains a search function, as well as search and replace, that can be used within one document or across all open document

[27] specifically, where English is the matrix language

tabs, as well as allowing plugins, such as Combine and TextFX which were employed in the tagging and analysis of the corpus.

NPP Combine (v.1.0.108; SCOUT-Soft, n.d.) combines all documents open in tabs in Notepad++ into one new document, with an option for the original file names to be included to title each section of the new document or have blank lines as the divisions, the latter of which was employed. The Combine plugin was required to prepare for another plugin, TextFX (v.0.2.6; n.d.). TextFX is a word count plugin that was used to establish the word count of the /CGL/ and Troper Tales corpora. The plugin, as opposed to some other word counters, counts even multiple tags attached to a word as part of that word, rather than counting it as separate word. It does also, however, count number sequences and symbols unattached to other words as words, requiring the removal of post numbers and date and time information from the /CGL/ corpus, and the removal of structural asterisks in the Troper Tales document, as was detailed in Chapter 3.2.

4 RESULTS

The aim of the current research is to investigate and linguistically analyse the usage of Japanese language features within English sentences in data gathered into a corpus of written language from an online Japanese pop culture fandom, as well as identifying attitudes that may be present in the community towards that particular form of language use.

The first section of this chapter covers Research Question 1: To what extent is Japanese language used; what linguistic forms and features are present and how are they integrated. In order to answer these questions, the data was analysed in terms of orthographical, lexical, and semantic aspects of the words found. The second section explores the sociolinguistic factors informing language use through qualitative analysis of language attitude data gathered, in order to answer Research Question 2: What attitudes towards this language use are found within the community, and how do those attitudes possibly affect language use?

4.1 LANGUAGE FEATURES

The corpus data set collected from the Cosplay and Gothic Lolita forum on the image board website 4chan, comprising of text from 100 threads from 1/5/2015 to 15/5/2015, came to a total of 402449 words. Within the corpus, 3420 tokens of Japanese language and Japanese style kaomoji were found, making up 0.85% of the overall total corpus (see Table 1 below). As will be discussed further below, the main focus of the research here is the *romaji* tokens, the alphabet script which makes up the vast majority of the data at 99.39% of all tokens (see Table 2 below), and which brings the corpus percentage to 0.84%.

However, for context, if all instances of Japanese names and place names are removed (see Appendix 3 for a list all _JNC and _JNS words), the total corpus percentage becomes 0.74%, and in removing the tokens that are considered loanwords according to the Oxford English Dictionary the number drops further to 0.22%. The current study did not aim to restrict to one aspect of language mixing (i.e. focussing solely on loanwords or code-switching), and thus all Japanese tokens were tagged, including _JNS, _JNC and _JL, as they are still aspects of Japanese language that are important to note as being used. Overall, the presence of Japanese words is a small one, not even breaking 1% of the corpus. However, especially given the website is meant to be exclusively English language based, the presence of this 0.85% is indeed worth investigating, and further analysis does reveal patterns; the language use is not random.

Table 1: Percentage occurrence of Japanese in CGL corpus overall

Tag combination	**Description**	**Frequency**	**% of total CGL corpus**
_K, _JK and _^	*All tokens*	3420	0.85
_^	*Romaji tokens only*	3399	0.84
_^ minus _JNS and _JNC	*Romaji tokens minus Japanese series and character names*	2964	0.74
_^ minus _JNS, _JNC and _JL	*Romaji tokens minus Japanese series and character names and Japanese loan words*	897	0.22
Corpus Word Count		402449	100

4.1.1 ORTHOGRAPHY

The first observation that can be made about the language data is the form any Japanese language tokens take, such as the orthography. The Japanese language has two different orthographic scripts: *kanji* which is a meaning-based logographic script based on Chinese scripts that is generally used for content words; and *kana* as a sound-based syllabary that splits into *hiragana* for function words and *katakana* for loanwords, although that is somewhat of a reductive distinction (Doerr & Kamagi, 2014). This difference from English language script therefore requires literacy to be an active learning process as it is not likely to be picked up through casual language encounters. However, there is another script used in Japan called *romaji* which is the standardised romanization of the Japanese orthography (transliteration into Roman letters) generally used for text directed at foreigners or for computer input (Gottlieb, 2010).

The collected data is predominantly in *romaji* script – the English alphabet – with only 13 occurrences from the total 3420 Japanese tokens being in one of the Japanese *kana* scripts, which is only 0.38% of those tokens (see Table 2 below) in comparison to 99.39% being written in *romaji.*

Table 2: Frequency of the orthography tags in the CGL corpus

Tag	Frequency	%
_^	3399	99.39
_K	13	0.38
_JK	8	0.23
Total	**3420**	**100**

As commented on above, literacy is separate from verbal language, and the Internet is predominantly a written modality (refer to CMC in 3.1) where written language and stylisation is reinforced through exposure, meaning that a prevalence of *romaji* words indicates that the interaction occurring is between people who do not find *kana* or *kanji* a necessary part of communication. The language interaction in the corpora is therefore between Anglophones of varying language learning relationships with Japanese, instead of communication between Japanese-speakers, bi- or multilinguals, or L2 language learners actively furthering their Japanese language skills. This supports the idea that the expected fluency of the intended audience is not especially high.

Swapping between alphabets in CMC requires swapping between different keyboards on the writers' electronic devices, which, were it deemed necessary and they had the required literacy, would not be impossible. However, in this case the usage of *romaji* is seen as a feature of the language type beyond typographical issues.

Following orthography, the next section continues the lexical description of the data, investigating the spread and frequency of word classes.

4.1.2 WORD CLASSES

As previously mentioned in Chapter 3, the Japanese tokens present in the corpus were identified and tagged according their word class, and whether they represented other interesting features. See Appendix 2 for a compilation and description of all the parts of speech that were identified, as well as other features identified to be of interest for analysis.

The word class frequency results were calculated two ways. Firstly, the overall frequency of the different parts of speech as they were used in the corpus, with percentages calculated according to the total number of Japanese *romaji* tokens (see Table 3 below); and secondly, the frequency of word class according to the word list constructed from the data (see Table 4 below). The

former shows the word classes of the words that are used more frequently, and those that are used less frequently, whilst the latter is sorted according to the actual range of different words used. For example, nouns (including attributive nouns and pronouns) make up 69.11% of the times a Japanese word is used in the corpus, but they make up 84.62% of the constructed word list.

Table 3: Frequency of Part of Speech Tags in the CGL Corpus

Tag	**Raw Frequency**	**Percentage of Total Romaji Tokens**
JPHT	6	0.18
JPHS	102	3.00
JPV	330	9.71
JPNO	1890	55.60
JPNA	446	13.12
JPPA	8	0.24
JPIE	2	0.06
JPSS	15	0.44
JPPR	2	0.06
JPAD	162	4.77
JNS	76	2.24
JNC	360	10.59
Total Romaji Tokens	**3399**	**100**
Corpus Word Count	**403972**	**0.84**[28]

[28] Percentage of total *Romaji* tokens in corpus

Table 4: Distribution of Japanese Words According to Word Class in the CGL Corpus

Word Class	Frequency	Percentage
Noun	108	84.4
Honorific	6	5.6
Adjective	3	2.3
Verb	3	2.3
Particle	3	2.3
Interjection	2	1.7
Sound Symbolism	2	1.7
Pronoun	1	0.9
TOTAL	128	100

The difference is due to the spread of the frequency of words appearing in the /CGL/ corpus heavily skewing towards a small range of words. Of the total 3399 *romaji* tokens in the corpus, there are only 128 different words used (see Appendix 4). Different spelling variations or abbreviated forms were counted as one word, e.g. *kigurumi* and *kigu* [onesie-style pyjamas, often animal themed] were counted as one word rather than two, as were *cosplay*, *cosplayers* and *cossupurai*. This indicates a consistency of use as the same words appear over and over again, whilst others are not so common, at least within this particular corpus. Of the individual words found, 84.62% (110 words) are nouns, attributive nouns, or pronouns (pronouns infrequently, however) (see Table 4). Honorifics are the next most prolific word class, but at the lesser presence of only 5.6% (6 words). The findings as presented in Table 4 also help contextualise the higher verb frequency found in the usage statistics; with the main verb identified as the English naturalisation *cosplayed*, a loanword (see Chapter 4.1.5) that was originally a noun.

As discussed above, Nouns (tagged _JPNO) are the largest word class at 55.6% of all token instances and 84.4% of the individual words. This result is expected according to other research in Contact Linguistics (Cannon, 1994) and therefore nouns will be discussed further in terms of semantics (see Chapter 4.1.3), modification (see Chapter 4.1.4), and loanwords (see Chapter 4.1.6) in later sections.

Attributive nouns (tagged _JPNA) which account for 13.12% of the tokens, were included separately from nouns partially based on their inclusion in Turner's (2015) research, but also as when they are used attributively they modify an English noun, which could be linked to

naturalisation and change in meaning as a concept or thing expressed in English is modified (according to English conventions) to be related to Japanese concepts.

Verbs (tagged _JPV) appeared at 9.71%, which, while a small percentage, is far more frequent than reported in Turner (2015). In the current study, the presence of verbs is skewed by the inclusion of words used as verbs according to the English language. For example, *cosplay*[29], a word coined within Japanese (although from English elements) as a noun, is then modified by English grammatical affixing into a verb, *cosplayed* or *cosplaying*, and as such there is not a huge large of Japanese verbs like *desu* actually present. There is an argument for not including these words, but methodological boundaries are necessary, and in the case of the type of language being studied, it is not surprising there are instances of language use that are hard to categorise neatly; language is complicated, particularly when mixing and adapting words, so this is merely a reflection of that.

Adjectives (tagged _JPAD) occurred at 4.77% within the corpus and made up 2.3% of the word list. The main adjectives found were *kawaii* [cute], *ita* [painful to look at], and *sugoi* [awesome]. *Ita* was also used as a noun in the corpus, compare the two uses in Example (5), and was tagged either as an adjective or a noun accordingly.

(5) *How do you know there is a thing called ita_^_JPN_JC but not know what it takes to be ita_^_JPAD_JC?*

Honorifics, including use as both a title (tagged _JPHT) and as a suffix (tagged _JPHS), accounted for 3.18% of the instances of Japanese language tokens. Separately, they occurred more often in suffix form at 3% than titles at 0.18%. The most common suffix was *-chan*, which functions as a "diminutive everyday suffix for children" (Loveday, 1986, p45), and thus indexes a cute, youthful meaning. In the CGL corpus, it is suffixed to a non-name noun to create a descriptive label, such as in Example (6) and (7), which is of particular use in an anonymous forum where individual writers are not identified and tied to their posts by usernames. This is also a convention used in Japan to refer to workers when their name is not known, i.e. *shopkeeper-san*, although generally with the more general suffix *-san*.

(6) *Sidestep. Just say no. If they ask why you don't want to partake, explain. If they think you're a meanie-bully-chan* [honorific suffix]*, get the fuck out that comm because you're hanging out with snowflakes*.

(7) *Poutine-chan* [honorific suffix]*? Do you still want to take me on a poutine tour?*

[29]Which, as a loanword, was not included in Turner (2015).

Honorifics, title form as well as suffixes, do not translate directly from Japanese to English because of the way they are used in Japan. English language does not have the same politeness systems, so even in translating kinship terms literally, such as *oniichan* into big brother' English language speakers do not address each other in the same way as Japanese speakers do. For example, saying "Hey, bro" or "Hey, sis" to a sibling is still functionally different from the situations and contexts and extents to which Japanese speakers use honorifics. *Senpai* has a meaning that can be translated (older classmate or more experienced colleague), but there is no situation in an English-speaking context in which it has an equivalent, because English language has a far more basic honorific system.

Whilst there is no structural need to use Japanese suffixes in an English language conversation, reasons for speakers to wish to do so include wanting to indexicalise Japanese-ness and recreate a Japanese context with little fluency, ingroup situations where attempting to show off language knowledge, and signalling solidarity with others in the same fandom.

Instances labelled as Sound Symbolism (tagged _JPSS) were found to be a very small percentage of the *romaji* tokens, but their presence is interesting as they indicate some knowledge of onomatopoeia conventions that are unique to Japanese culture[30]. Particles (tagged _JPPA), for example *ka, ne*, are infrequent (0.24%) and are present only within the few multi-word phrases within the /CGL/ corpus (see Chapter 4.1.6). Interjections (tagged _JPIE) were the least frequent of the word classes, making up only 0.06% of all tokens. Whilst usage was very low, community knowledge of words within that category is likely higher, which may be suggested by the far high percentage (25.35%) of Interjections found in the self-reporting Troper Tales corpus[31] (see Appendix 5). The two (0.06%) pronoun instances, *watashi,* were found in a discussion about the choice of Japanese text to be used on a t-shirt design rather than being used within a phrase, indicating a higher level of fluency from that particular writer than other examples.

The spread of word classes is only one aspect to the lexical categorisation of the language use. Next is a description of the morphology of the words found, in terms of content words versus function words, and how the words are grouped according to topic or semantic domain.

[30] Sound effects are conventionalised in manga.

[31] A full comparison of both corpora was not methodologically sound for the current study, but this result was of interest (see Appendix 5 for word class frequencies across both corpora).

4.1.3 SEMANTIC CATEGORISATION

Investigating the occurrence of content and function words[32] is a common part of code-switching and borrowing research, although generally part of a grammatical approach (Muysken, 1995). In the current study, rather than using the results to account for underlying structure and rules of language, it is a point of lexical description. Another aspect of lexical description is the semantic domains of the words used, and how they relate to each other and the topics within the corpus. Previous research has indicated that, in terms of content words such as nouns, certain semantic domains are more likely to be borrowed and incorporated as loanwords (Cannon, 1994; Tyson, 1993). Backus (2014), on the topic of 'borrowing', notes that identifying the semantic domains of the words used reveals information about both diachronic and synchronic language contact situations, and therefore is important to investigate in studies aiming to describe or comment on language contact.

Semantic domains are used as a concept for grouping words by their meanings rather than any deeper examination of the linguistic theory (Gliozzo, 2006). It is useful as a categorisation tool to be able to relate the words found back to the topics of discussion that are present in the corpus, and to identify patterns of relations together. The topics of discussion in the corpus are narrow as the corpus is constructed of language data taken from one place where there are rules about what is considered 'on-topic', thus it is not surprising that many of the Japanese words used relate specifically to concepts that are discussed in a way that tends towards the borrowing side of the language mixing continuum, although code-switching also is characterised as being triggered by topic changes (Backus, 2014).

Cannon (1994) listed descriptive categories, or semantic domains, in their description of Japanese loanwords found in English, such as weaponry, martial arts, culinary, business/currency, arts, sports, religions, health/disease, furniture/household, and trademarks. One of the dimensions Stott (2006) used to correlate their data the topic of the articles within their corpus, including Japanese culture, music, travel overseas, life/experiences in Japan, and life/experiences in home country as well as others; although the semantic domains of the words themselves were not explored. The CGL corpus features discussion of Japanese media as filtered through costuming and crafting topics, and topics related to fashion styles originating in Japan.

[32] Content words contain semantic content, such as nouns, adjectives, while function words create the grammatical meaning of the sentence.

The purpose behind investigating what types of semantic domains are present and beginning to discern the relationships between them and the topics of discourse within the corpus is not to create an exclusive taxonomy but to instead be descriptive to help identify patterns.

Firstly, the Japanese lexicon uncovered in the current research was considered in terms of whether the words were mainly content or function words. Content words were the most prolific, and could be divided into technical terms, general terms, or ritual terms, according to what was found in the data. Within the technical terms, there were the following semantic domain groupings: specific interests or hobbies, for example *shinai* [bamboo sword]*, figma* [action figure] *, dakimakura* [body pillow printed with the image of an *anime* character]; food, for example *gohan* [rice]*, omurice* [rice omelette]*, onigiri* [rice ball]*, sushi* [filled rice roll wrapped in seaweed]; clothing items, for example *hakama* [wide pleated trousers]*, gi* [traditional style top]*, kimono* [traditional style robe]*, furisode* [robe with long sleeves for unmarried women]*, yukata* [casual summer robe]*, sukumizu* [swimsuit]*, shimapan* [teal and white striped underwear]; fashion styles or terms, such as *ganguro* [lit. 'black-faced girls; style typified by tanned skin, bleached hair and bright clothes]*, yamanba* [more extreme version of *ganguro*]*, akamoji-kei* [mainstream fashionable style]*, ouji* [little prince style]*, nanchatte* [school uniform style]*, seifuku* [sailor style school uniform]; media related terms and genres, for example *anime* [animation]*, manga* [graphic novel]*, bara* [muscled man; a genre focussed on muscled men]*, otaku* [obsessed with popular culture]*, yuri* [lesbian media]*, yaoi* [gay media]*, moe* [a type of character or feeling[33]]; and religious or cultural terms such as *miko* [priestess]*, kamisama* [god]*, tenshi* [angel]*, Shinigami* [death god]*, gaijin* [foreigner]*, seppuku* [ritual suicide]. The general terms found include less specific words such as *kokoro* [heart], *kawaii* [cute], *baka* [idiot], *himitsu* [secret], *sugoi* [awesome], *waifu* [wife], *husbando* [husband], *neko* [cat], *megane* [glasses], *shiro* [white], *kuro* [black]. Other content words can be categorised as ritual words, for example *ganbare* [good luck] and *moshi moshi* [hello]. There was a smaller amount of function words, which were categorised as grammatical items such as *desu* [copula verb to be] and the particles *ne, yo, ka* or as terms of address such as kinship terms and honorifics, such as *onii-chan* [older brother]*, nee-chan* [older sister]*, -senpai* [older classmate or colleague]*, -chan* [addressing a small child or young girl]*, -tan* [cuter version of *-chan*], *-kun* [addressing a boy or junior at work].

Some of these general words also overlap with the technical category due to words that are popularised by specific use in media and thus have developed specific connotations beyond

[33] A hard to define term; see Galbraith (2009) for discussion of the meaning.

media merely being the source of language transmission, and the words found are also generally used in the specific context of directly discussing Japanese cultural products due to the moderated nature of the CGL forum. *Megane*, means eyeglasses but is also used as a descriptor in Japanese media to describe a particular character stereotype which has connotations beyond just wears glasses, implying personality traits such as cool and collected and quiet and possibly shy, but also vaguely sinister. *Waifu* and *husbando*, are *wasei eigo*[34] words that have been semantically narrowed to refer to a character you love and consider to fictionally be your wife or husband but can also be used as a cute way to refer to literal spouses. *Shiro* and *kuro* are respectively the terms for black and white which are also used to label specific substyles of the Japanese Lolita fashion, i.e. an all-white or all-black coordinated outfit.

The specificity or narrowed contexts of words discussed above could also occur in stylised ways specific to a particular group rather than to how it was originally used, such as the development of a meme within the fandom. For example, *desu* is a very common structural word in Japanese, but why might it be relevant to the Anglophone community aside from repeated exposure, and more so than other grammatical words? Because, in this case the community developed their own cultural response to it, as a meme, which elevated the connotational meaning to something that is attractive to be borrowed and used. Otake (2010) claims that the stronger the impact a cultural image has on the audience, the more likely that language term will be adopted as opposed to a translation.

As a precursor to discussing language modification, it is interesting to note that in November 2015, after the time period of data collection, the /CGL/ board, as well as other 4chan boards had word filters reinstated after a change in website ownership. While many web-forums implement word filters to remove curse words and slurs, on 4chan, a website known for its irreverence, it is slang abbreviations or words related to various internet memes that are considered overused or annoying that are changed, many to Japanese words. For example, *smh* [shaking my head] is changed to *baka* [idiot], *tbh* [to be honest] is changed to *desu* [copular verb, roughly translated as 'to be'], and *fam* [short for family], used to address close friends rather than literal family) is changed to *senpai* (honorific used for one's senior in school or work). The word choice is not random; well-known English slang is swapped with well-known Japanese words that also have a similarity in meaning.

[34] Words that are coined in Japan from borrowed English elements. See Chapter 4.1.5.

The presence of these filters would have skewed the data from 'natural' language usage had they been in place during the data collection window, and it should be noted, will likely skew research based on future data collection, but the fact that these filters insert Japanese words is interesting in of itself. The filters offer background contextualisation information about the web-forum affinity space; it is a linguistic space that is policed by moderators as well as peers within the community, and shows that some Japanese words are common or popular enough to have a meme-like status.

The reason this meme-like usage of the words is interesting is the semantic change of the meanings it implies; while, in this particular usage of the language, writers will be still aware of the origin of the words, and, to whatever degree, the original meaning of the words, the new usage is now entwined with specific examples of English online slang. And, in fact, the Japanese terms are now the preferable and less objectionable forms that both help construct the identity of the forum by employing language play that relates to the userbase of Japanese culture enthusiasts and also continues the implication that the words being replaced (*tbh, fam, and smh*), which have been overused to distraction. The Japanese words have been semantically narrowed to the colloquial meanings of the original English slang words, while transporting and exaggerating indexicalised meanings.

There are also examples of this 'meme-ification' of Japanese words within the data collected for this study. In Example (8) below, from the Troper Tales data, *desu*, the Japanese copula verb to be is used ungrammatically as a name for comedic effect, and was hyperlinked to a 'Memetic Mutation' or meme page detailing the DESUDESUDESU meme[35] that has contributed to the elevation of the word *desu* to far more than an innocuous grammatical feature. The same is occurring in Example (9), where *desu* is reduplicated and employed as if it were a different verb in a joking facsimile of wedding vows. The Japanese words are being used in a joking manner; indexical meanings are still associated, but functionally, or grammatically, the sentences are nonsense. The resulting use or meaning of the words is for language play rather than conveying the literal meanings.

(8) *Now you can head back to the supa supa kawaii Gratuitous Japanese page! Sayonara, Desu[[[[MemeticMutation DESUDESUDESU]]]]-chan! Arigato!*

(9) *I now pronounce you: waifu and waifu. You may now desu desu*

[35] In the anime Rozen Maiden one of the characters has a verbal style in which she reduplicates *desu*.

The relationship between language mixing processes and slang and jargon will be discussed further in Chapter 5.

4.1.4 MODIFICATION AND NON-MODIFICATION

Investigating how the language tokens are either modified, categorised as English modification, or not, categorised as Japanese modification, according to English language structure is another area pertaining to the question of how the language is integrated.

English modification as a feature refers to instances where the Japanese words are fully integrated into English morphological structure, which includes affixing like verb suffixes, *cosplayed and ita-ing*, adjective suffixes, *kawaiiest*, and plural suffixes, *kimonos*, in order to change the grammatical function. Cannon (1994) found this suffixation to be a common trend in loanwords borrowed from Japanese into English. Out of the 3399 tokens, there are 730 instances of English modification. On the other hand, non-modification refers to Japanese language features that have been retained from the original words or language, such as Japanese plurals, Japanese abbreviation, and Japanese pronunciation spellings.

In most cases plurality in Japanese language is inferred through context rather than inflectionally as it is in English. Regular plurals in English are indicated through the phonetic variations of the plural suffix -s, while a smaller category of words take irregular plural endings, including the zero morpheme -∅ where there is no phonetic content as in words like ‘sheep’ and ‘fish’. The latter is found to be an option for pluralisation of Japanese loanwords alongside the more typical -s. In the following Examples (10) and (11), *kimono*, an item of traditional Japanese dress, is used with different pluralisation styles by two different users, and Example (12) shows that some members of the community are aware of the lack of plural marking in Japanese and that this has led to a question of typographical conventionalisation.

(10) *Used/slightly flawed kimono are so cheap on eBay and other second-hand markets that it's not cost-effective to make your own.*

(11) *Anyone know a good guide for making traditional kimonos?*

(12) *Y'all know the plural of "kimono" is just "kimono", right? the 's' isn't necessary... In Japanese yes, not sure as a loan word...*

Length and frequency of use was posited as a factor influencing which Japanese words took -s and which kept the Japanese zero morpheme, but studies (Cannon, 1984) have not revealed a pattern indicating this, or any other explanation for how plurals are integrated, although Bloem

(2014) speculates the reasoning behind fandom members not using English plural forms choose to do so for authenticity.

The tagging and analysis of these plurals was sometimes hard to keep consistent due to the difference between count nouns and mass nouns, or more specifically that a word may have both aspects in use to differentiate between the borrowed concept in general and in specific. For example, *cosplay* can be a countable item in referring to a specific costume while also being used in a broader sense as a concept or activity; the difference between '*Cosplay* is…' and 'My *cosplay* is…'

The Japanese pattern of abbreviation and blend creation involves the first two morae[36] of each word being combined into a shortened form. Members are aware of how media series names are abbreviated in Japan (tagged JMSH), such as *Utapri* for *Uta no Prince-sama* and *Tourabu* for *Touken Ranbu*, as well as *wasei eigo* blend words for popular culture products like *purikura* [picture capture] and *cosplay* [costume play] and often replicate it. Members of the community are therefore exposed to knowledge of the process through particularly accessible examples. English has a similar blending process, but the first morae or syllables of each word are not always used[37].

Japanese pronunciation adaptations of English words (tagged JMPR) appeared in Stott's (2006) study of code-switching employed in newsletter articles written by native English speakers living in Japan. More specifically, they are English words that have been borrowed into Japanese and modified to match the phonological structure of Japanese. As will be discussed further in Chapter 4.1.5, Japanese has a high rate of borrowing lexical items from English, which therefore creates methodological issues for labelling borrowings or code-switches in the current study. Stott (2006) argued that these words should be considered code-switching despite their English origin, as the writers made the specific decision to write the modified form of the word when, as native English speakers, they could have written the English form instead. Stott also claimed that these words encode different ideals or stereotypes relevant to the Japanese context in comparison to their originator counterparts. The example Stott gives is that of the difference between Japanese *sarariiman* (salary-man) and Western businessmen, where the Japanese version of the word as well as perhaps implying the Japanese ethnicity of the man, also encodes

[36] A mora is a small phonological constituent relating to 'weight'; a short vowel or light syllable is typically one mora, while a long vowel or heavy syllable is typically two morae (Fabb, 2001).
[37] In the blend of actors Brad Pitt and Angelina Jolie's names the beginning of the former and the end of the latter is used to create 'Brangelina'.

the stereotype of a constantly suit-clothed man who works long hours and goes to karaoke bars after work.

Many of the instances of Japanese pronunciation adaptations of English words found in the current study have a similar effect in that they encode a particular meaning specific to the Japanese culture fandom. One of the most evident examples of this are the nouns *waifu* and *husbando*. According to the website Know Your Meme ("Waifu", n.d.), colloquial usage of the word in Japan predates the usage by the Japanese culture fandom. The word, borrowed into Japanese from English, is generally used in Japan to refer to someone's literal wife, although it also came to be used to refer to a fictional two-dimensional character an *otaku* wants to be or considers to be their wife. The point of exposure of this modified form of the word to audiences outside of Japan is generally thought to be a quote from the 2002 *anime Azumanga Daioh,* where a character replies *"Mai Waifu"* to a question about a woman in a photograph. Consequently, the term became popularised in online Japanese culture fandom, in terms of its metaphorical meaning. Originally, according to an archived copy of frequently asked questions on 4chan.org's *anime* board /A/, *waifu* was not necessarily considered to be a gender specific term, which is interesting in terms of the degree of distance between the Japanese version and the original English version that it implies, with *waifu* being considered by the online Japanese culture fandom as a specific concept removed from the original literal meaning of the word. However, later a masculine version of the word was coined. Whether *husbando* to mean a favourite male character was coined by the Anglophone Japanese culture fandom community following the conventions of Japanese pronunciation to match *waifu*, or whether it was coined in Japan for that meaning (as opposed to the literal meaning) by *otaku* and then exposed to the Anglophone community is harder to verify, but the implication found in online discussion of the meaning of the word seems to be that it is the former case.

(13) *I need to get one of their Dress Sabers, so I will fite u for mai waifu* [wife].

(14) *He's my husbando* [husband] *and I haven't even seen a good cosplay of him. Why live.*

(15) *The guy at the booth was probably her husband.*

(16) *Oh my God so fucking cute 11/10 would waifu*

Every instance of *waifu* and *husbando* found in the research data, such as in Examples (13) and (14), refers to the meaning of a favourite fictional character, specifically the character one would want to marry if they were able to. When wife and husband are used instead, as in Example (15), each case specifically refers to the literal meaning and not the metaphorical one. Thus, *waifu* and *husbando* have semantically narrowed meanings from how the words are used in Japan. As *waifu* is thus considered a specific concept important to Anglophone Japanese culture fandom,

Example (16) shows language creativity where it can be used as verb describing the act of considering someone or something a *waifu*.

Another Japanese pronunciation pattern word frequently used in the /CGL/ corpus is *burando* [brand], used to refer to the big-name Japanese brands in Japanese Lolita street fashion, as opposed to smaller Japanese brands and non-Japanese independent brands. *Burando* is another loan word from English in Japanese, present in the 1980s Japanese fashion movement *DC Burando,* meaning designer and character brands, involved brands with very recognisable style aspects easily tied to individual designers (Web Japan, n.d.). With Lolita fashion being a niche fashion, and particularly one with an Anglophone internet space on an anonymous image board, amongst other places, which is known to be very judgemental and harsh, brand clothing is an important status symbol, particularly in regards to issues of authenticity given the distance between the fashion's origins in Japan and the Anglophone community.

In Example (17) below *burando* is used as a standalone noun, rather than paired with an attributive noun or other explanatory factor, and refers to a range of specific brands for a specific fashion style. In the CGL corpus there are 19 instances of *burando* being used comparison to 202 instances of the word brand. The latter is used in Example (18) twice, once to refer to the big-name Japanese Lolita brands in general as a mass noun similar to *burando* in Example (17), and secondly to still refer to Japanese Lolita brands but in the singular as separable brands. Brand as a term is used interchangeably with *burando* but is also used in contexts outside of the specific Japanese Lolita brand context, while *burando* is not. Thus, *burando* also has a narrowed meaning, but the separation between *burando* and brand is far less pronounced than that found with *waifu* and wife, as brand contains the *burando* meaning while wife does not contain *waifu*.

(17) wtb [want to buy] *a bag sax or cream or possible even a warm brown. has to be burando* [brand] *but it doesn't matter which one*

(18) Talk about your new coords, how much brand you have, how shitty other people's coords are. elitism talks, how ita [painful to look at] *your comm is because they don't wear the most popular brand*

Given the word brand can be used, and is more frequently used, to connote the 'Japanese Lolita brand' meaning, it would then follow that using *burando* is a marked choice by the writer, to add some sort of extra emphasis. It perhaps emphasises the high status of the brands by emphasising the Japanese-ness of them. Usage of the word also perhaps is a playful, self-aware and slightly self-mocking critique of the writers' own views on the Japanese brands denoting high status, in a community such as /CGL/ where enthusiasm and enjoyment of a particular

Japanese item needs to be tempered with self-awareness so as to not seem like weeaboo behaviour.

There are a few of the above style words which have far less specific or narrowed meanings and are instead innovations employed by the writers to encode a general Japanese-ness, but also a sense of playfulness or light-heartedness. These words also tend to be ones that could be considered more in terms of innovative language created by the Japanese culture fandom community of practice as opposed to being terms that are borrowed and repeated as *waifu* and possibly *burando* would be, although not in all cases. An exception would be *sankyuu* ('thank you') and *baibai* (bye bye), as in Example (19), which are used in Japan, but do not have a specifically changed or narrowed meaning when used by Anglophone Japanese culture fans.

(19) His favorites are "Baibai" [bye bye] and "Sankyuu~" [thank you].

(20) Lolita? Looks like lazy meido [maid].

(21) hot topic is so rori [loli] guiz!! ^-^~~~

(22) >wear 34B
>measure self using /r/abrathatfits
>28G, naisu naisu [nice nice]
>order 28G bra from recommended list

(23) >Seemed like some interpersonal dramu [drama] was going on.

Meido [maid] in Example (20) above refers to a style of maid costume popularised by themed restaurants in Japan called maid cafés. Maid and *meido* are used interchangeably in the /CGL/ corpus, with the English form more common. *Rori*, as shown in Example (21) is the shortened form of the Japanese pronunciation of the fashion style gothic lolita, and in this case is used in a mocking way to invoke the stereotype of a person who would use that language sincerely. In Example (22), *naisu naisu* [nice nice] is used in a comment about finding a correct bra size, and therefore is outside of a specifically Japanese context although the interaction is still located within a forum focussed specifically on Japanese topics. *Dramu* [drama] in Example (23) is perhaps styled as such with mocking or self-deprecating reference to the Japanese pronunciation modification convention in the Anglophone community, as opposed to attempting to mimic Japanese phonology sincerely. From the above examples it can be seen that Japanese pronunciation modification is a linguistic innovation style present in the repertoire of the members of the Anglophone Japanese culture fandom alongside other internet language innovation conventions such as misspelling for effect, which can be employed by the speaker for language play opportunities. Stott (2006) suggested that Japanese pronunciation transliteration of English words is often used for comedic effect or parody, whilst Turner (2015)

found that the Japanese versions of the English names seemed to be used for characterisation and as a diminutive.

It is important to mention, given the emphasis within the current study on the crossing of sharply felt ethnic and cultural boundaries, that the replication of Japanese pronunciation of English words used by the Anglophone Japanese culture fandom could be seen as mocking Japanese accents or their attempted approximations of English words. Stott (2006, p. 32) notes that native English speakers in Japan often have negative views towards the Japanese-English codeswitching employed by native Japanese speakers that one speaker referred to as Japlish. More in depth discussion of the racial issues that have arisen with the rise of Anglophone Japanese culture fandom is covered in Chapter 5.1, but, briefly, the language usage in the community appears to be either used somewhat sincerely invoking both Japanese as a high status language and solidarity with other community members, or mockingly, including self-deprecation, in regards to the negative stereotype of Anglophone fans who are too involved in or enthusiastic about Japanese culture fandom to the point of 'riding a fine line between homage and racism' (Saturday Night Live, 2011) as opposed to specifically making fun of the language practices of actual Japanese people. However, the latter interpretation cannot be dismissed, particularly when taking into account issues of ethnicity, colonisation, and globalisation.

Iwabuchi (2002) claims that it is one thing to note that popular culture influences play and behaviour globally and influences the perception of Japan as a cool nation because of cool cultural products, but quite another thing to say this perception is actually associated with a realistic appreciation of Japanese life styles or ideas. In terms of linguistics, Hill (1999) asserts that language styling in which speech elements of (minority) community are replicated indicate a wider process of subtle racism even while individual utterances could be considered playful or benign, and that in the wider context of colonisation and racism "it is very hard for a member of a dominant group to simultaneously 'use' and yet signal, 'this is not mine, it is truly yours, and I honor you by adopting it.'" (Hill, 1999, p.554)

The Japanese pronunciation feature (tagged _JMPR) is interesting because it is multimodal in a way as it is spoken language explicitly replicated in written form. It indexicalises Japanese-ness more so than the Roman alphabet must do for the writers; but not necessarily in the more formal or 'correct' way that the usage of *katakana* for *wasei eigo* or loanwords in Japanese would indicate. It is an extra level of Japanese-ness that is accessible to writers who do not know and therefore cannot employ Japanese orthographies.

The above sections on shortening and pronunciation patterns also indicates awareness of the patterns of Japanese morphology and how Japanese speakers modify the English words they borrow that even non-speakers and non-learners have gained from exposure to Japanese language.

4.1.5 LOANWORDS

Japanese is an avid borrower of words from English; Moody & Matsumoto (2012) state that "there is essentially no sector of Japanese culture of society that does not regularly use loanwords from English…", and that usage in popular culture sectors is especially high. One of the main driving forces behind language borrowing is widely considered to be the need to fill lexical gaps and label culturally specific concepts or new technologies (Weinreich, 1953), however, many of the English loanwords borrowed into Japanese are used in place of acceptably equivalent terms (Honna, 2008), indicating there are other factors in play influencing language change. The proliferation of English loans in Japanese has led to the presence of *wasei eigo* words, in which English elements are used within Japan to create new compounds or expressions that are still more closely related to Japanese contexts than what the original elements (Miller, 1998). These words are then reintroduced to the Anglophone community through the transmission of Japanese pop culture, which then, in turn, may be borrowed. The current research has included these instances as Japanese language tokens, despite the originally English elements, as those elements have been filtered through Japanese sensibilities for Japanese purposes, but they are clearly described as such in the current research, alongside the other loanwords that are included. In an increasingly globalised world, language contact and language change are increasingly complicated, and the most important thing is to examine and describe the context as clearly as possible, particularly from an approach that views borrowing behaviours as a diachronic aspect of a constantly fluctuating continuum where code-switching behaviours are the synchronic aspect (Backus, 2014, 2015).

Out of the 128 unique words found in the CGL data (see Appendix 4), 29 of them (22.66%) are loanwords according to the Oxford English Dictionary. Out of those 29 words, 28 are nouns (96.55%) and one is an adjective (3.45%). One word is also used as a verb in the data, *cosplay,* and is thus included in the verb count statistics in previous sections but refer to previous discussion on word class designation decisions. In terms of usage frequency (see Appendix 6), seven out of the top 20 most frequently appearing Japanese words are mainstream loanwords, and account for 89.26% of the usage in the top 20. Six out of the top 20 words are also *wasei*

eigo words. The overwhelming percentage of nouns agrees with other studies on loanwords and Japanese loanwords in English (Cannon, 1994).

Most of the loanwords used within the /CGL/ corpus are technical terms for specific Japanese things, for example *katana* [Japanese sword], *sushi* [food item], *hiragana* [Japanese orthography], as well as specific cultural concepts, for example *seppuku* [ritual suicide], *otaku* [obsessive nerd]. Within the Oxford English dictionary definitions these words are also specified to be Japanese style, or the meaning is prefaced with "in Japan" ("Otaku", n.d.), as opposed to less specific, and possibly older and more entrenched, loanwords such as *tsunami* [tidal wave] which only mentions Japan in the etymology information. *Hentai* [pervert] is narrowed to be a technical term for a specific Japanese product or media genre from the wider meaning present in Japan of someone or something that is perverted. *Kawaii* [cute], originally an adjective, can also function as noun, where the meaning is conceptualised as a specific cultural movement that invokes the quality of being cute. *Anime* and *manga* are specialised to mean a specific product (i.e. Japanese animation and Japanese comics) and become words in English that differentiate Japanese media from general animation, while in Japan the words are general and are also used to refer to foreign animation.

Loanwords, defined as such due to their presence in an English language dictionary[38], make up a high percentage of the words found in the /CGL/ data, which was expected due to the usage of the words which are single word instances for the most part, mostly nouns that are semantically technical terms or specifically related to the fandom, and the overall culture of the forum, which includes language and behavioural policing, resulting in self-deprecation with the knowledge that their interests are something that need to be defended and their interactions toned down. Therefore, a number of the commonly used nouns not currently considered to be loan words according to the metric of being in a mainstream English language dictionary could be considered to be on the loanword or borrowing end of the language mixing continuum within the community, both due to the level of usage which indicates a high level of entrenchment which is the basis in a usage-based linguistics approach (Backus, 2015) and the majority of terms being closely semantically related to Japanese specific items or concepts (as investigated in Chapter 4.1.3). Mahootian (2012) emphasises that the unit of language under analysis should be contextualised language, not an abstract unit of language, which is especially important here

[38] The Oxford English Dictionary.

as description, context, and usage information is perhaps more useful of a metric than labelling words as loanwords or non-loanwords, especially given a repertoires and resources approach.

4.1.6 MULTI-WORD SEGMENTS

The Japanese language use that most resembles traditional code switching are the few instances of multi-word segments of Japanese. There are only seven examples found in the /CGL/ corpus (see Appendix 7), but their presence nonetheless indicates longer segments are available for use to at least some degree. These longer segments also feature Japanese function words, such as particles, the use of which, grammatically or ungrammatically, is rare in the corpus.

The grammatical structures used in the multiword phrases are very simple grammar and are all similar sentences structures, such as the use of the *desu* formula in Example (24). *Desu* is the polite form of the non-past tense[39] copula verb, essentially meaning 'is' or 'to be', expressing an equivalent or similar relationship (reference). In Japanese, the predicate 'noun + copula verb' can be a sentence by itself without a subject explicitly stated, such as in Example (25) below found in the CGL corpus.

(24) *Yesterday I was in a cycling store with my SO [significant other]. My SO went into the changing room and the sales person helping out my SO approached me and asked, "Ninhonjin desu ka?"* [Are (you) Japanese?] *(Are you Japanese in Japanese.) Sales assistant is obvious whitey and that startled me. I'm Chinese and I get this lot, but never in Japanese. He probably was just curious, but it made me immediately think of guys with huge yellow fever whom I'm weary of. I live in a place without much Asians soooo yeah.*

(25) *>kawaii desuuuu, sugoi!!! [*(subject) is cute, (subject is) awesome]
>I bet he/ she's not even a real fan
>generally being unnecessarily loud, obnoxious and squealing
>Yaoi paddles, chan memes and 'doing it ironically'
>'sluts'
>leaving rubbish

(26) *I'd rather go to a panel with someone who has something to teach instead of being kawaii desu yo neeee* [(it) is cute, isn't (it)? (with slightly different, perhaps stronger, implication than 'kawaii desu ne')]

Structurally, Example (26), in comparison to (25) where the phrase is alternated intersententially rather than inserted intrasententially, is a Portmanteau Sentence, categorised by a mixture of elements from different sources (Nishimura, 1995). English and Japanese have different word orders, SVO (subject, verb, object) and SOV (subject, object, verb), meaning that intrasentential

[39] Japanese does not differentiate between present and future tenses (reference)

alternation or insertional code-switching style situations often occur with a certain degree of redundancy as phrasal elements are repeated. However, Nishimura (1995) claims that these sentences are structurally acceptable with the final element of the English sentence serving as the initial element of the Japanese sentence, resulting in a double verbal element in both languages; a SVOV or VSV structure where the first V is English and the second Japanese. Subject deletion is grammatical in Japanese meaning this element does not need to be incorporated a second time, allowing for the flow of elements to be undisrupted.

The multi-word phrases were generally employed in order to draw on or refer to a stigmatised community stereotype for mocking purposes, which can be said due to a number of features present in Examples (25) and (26). Reduplication of vowels indicates an emphasised and over the top type of speech, as does multiple exclamation marks. *Kawaii* and *sugoi* are most likely used as they are well known adjectives that could be seen as typical of weeaboo vocabulary and heighten the excitement the writer is, sarcastically, portraying. Example (26) also perhaps indicates that members of the community are aware to some degree that including multiple particles (*yo* and *ne*) means adding more feeling to the phrase in Japanese (Nishimura, 1995).

The number of tokens found overall as a percentage of the whole corpus is small, as is the count of individual words, but while the language phenomenon is not overly complex in terms of integration, it is of particular import as both the strong language attitudes found within the community and the indirect nature of the language contact make it an interesting area of study, which will be discussed in the following section.

4.2 Language Attitudes

The nature and boundaries of language is a recent area for in depth research and debate with concentration on terminology and theoretical paradigms, but has long been a point of contention in the field of Linguistics. Beyond any linguistic features such as lexicon and grammatical structure, language is an inherently political concept. “Language is a dialect with an army and a navy” is a famous saying attributed to the linguist Max Weinrich in the 1940s (Ahearn, 2012, p.122), indicating the importance of positioning language studies within historical and current political contexts by exploring the sociocultural dimension. Whilst language labels and boundaries are often set by governments and are often the result of geopolitical processes with traditional language gatekeepers such as language policy creators, dictionaries, journalists, and academia; in many cases Weinrich’s army and navy can also refer to ordinary people acting as gatekeepers and who micro-manage the acceptability of certain

language tokens in niche speech communities or an affinity space that is characterised by an informal register. Kytölä (2012) notes this as characteristic of online communities, where normativity and prescriptive attitudes are no longer restricted to coming from traditional authorities but also occur between peers as they negotiate the hierarchies within the online space.

The following results record attitudes towards language use gathered through a qualitative analysis of the /CGL/ and Troper Tales corpora, with particular focus on the latter as it was chosen to predominantly be an attitudes resource. The general trends are identified and discussed, separated into 'Negative Attitudes' and 'Positive or Neutral Attitudes'. The Troper Tales pages within the corpus consist of anecdotal evidence given by writers on the topic of their real-life examples of 'Gratuitous Japanese', a label for English-Japanese language mixing that has a clear value judgement clearly built in; awareness of there being issues of *legitimacy* tempering acceptableness of the language use is known by each commenter, although, as is indicative of a wider trend of difference in the community, these views are internalised and affect language use to different degrees. The general trends in attitudes gleaned from within the corpus are clearly laid out in Appendix 8 with examples and will be discussed in more detail in the following sections, first exploring attitudes that were positive or neutral and then negative ones. The aim is to answer Research Question 2: What attitudes towards this language use are found within the community, therefore how those attitudes possibly affect language use.

4.2.1 Positive and Neutral Attitudes

An attitude expressed within the corpus was considered positive if there seemed to be status or enjoyment gained from the language behaviour. The positive nature is viewed from the point of view of the one expressing the opinion; whilst others may view that language usage in a negative way when they encounter it, but the writer is sharing as if it were a positive anecdote or something they enjoy. An attitude was considered neutral if the writer made no specific stance clear towards their language report; this was in comparison to later examples (see Chapter 4.2.2) where something similar would be broached with a justification, implying the writers know it is possible the community will view it negatively.

The first general trend is that knowledge of Japanese language seems to be high status, and as such writers want to show off their knowledge in order to attain this status. This is not surprising given the fact the community is made up of fans of Japanese cultural products, but it does also imply that in general, Japanese language is also a product or commodity to be enjoyed or sought

after, rather than merely being the vehicle for the popular culture products. In Example (27) below, the writer is actually being critiqued for the examples they have given to claim Japanese language knowledge, and they are dismissed as trying to brag. The writer's knowledge is deemed insufficient by other participants so their attempt to gain status through language fails, but it indicates that Japanese language knowledge is somewhere status can be negotiated. In a similar but more generalised vein, another trend is participants showing off what they think are unique or interesting language behaviours (see Example (28)), which supports the idea that language, whether foreign language or terminology based or just in general, is overall an area in which power is enacted.

> *(27) [This troper has a decent knowledge of Japanese…]*
> *Sorry, but this is NOT 'decent knowledge'. When you fully master both Kana systems, fully understand at least 1000 Kanji, read and understand written text from a textbook, and are able to exchange a few replies with a tutor, you will have the right to say you know the basics. Decent Japanese knowledge means you can read a daily newspaper, and buy your daily food, and generally find your barings in an average town in Japan. So shut up and stop bragging. You're only insulting people who actually bother learning the language.*
>
> *(28) Though, for some strange reason, his speech style seems to be…stoic-ish.*

There are many cases where the use of Japanese language is deemed acceptable, and perhaps even encouraged. For example, where there is no direct translation for the concept, or translation would lose significant meaning or important connotations in the process. The original language of the popular culture is Japanese, and thus those Japanese words have connotations specific to the culture that can be lost within a translation using an English word. Those connotations may be, in some ways, more specific to the non-Japanese consumers as their encounter with the words are more directly associated with the popular culture product than in Japan where there are many other contexts in which the words are used (excepting terms coined specifically for the product). In Japan the Japanese language is the normal language of communication and thus terms, while inherently being Japanese, don't index 'Japanese-ness' in the way they do for the non-Japanese consumers; the language itself is more default, for example the term *anime* refers to all animation including non-Japanese animation, whilst outside of Japan *anime* is specifically Japanese animation. Example (29) below shows a discussion of a Japanese media related term, *tsundere,* which is a common character archetype originally developed within Japanese popular culture. Words such as *tsundere* are seen more as loanwords necessary to convey specific cultural concepts or technical terms.

(29) *Tsundere is actually beginning to enter this troper's vocabulary. Since stealing from other languages is pretty much the entire basis of English, I figure it's fair.*

- *There really isn't an English equivalent for the word anyway.*
- *Yeah, it's not quite so gratuitous in that particular case*

As specific words are always accepted, so are certain contexts in which Japanese language is allowed or engenders no negative flack; the community will allow it.

In Example (30) the speaker refers to using the honorific address terms *sempai* and *kohai* in conversations with other Japanese language learners or Japanese speakers, meaning in a Japanese language context where the terms are part of the language structure and necessary for politeness. On the other hand, they are implying that outside of this context it would be superfluous and inappropriate.

(30) *I only use sempai/kohai with other J-leaners/Japanese people*

In the cases where examples of own language use are given without justification or disclaimer such as in Example (31), it is viewed as a neutral attitude because the writer does not come across as pre-emptively defensive, as in the examples from the negative attitudes section. However, the topic of the corpus is Gratuitous Japanese, so they are not necessarily saying that the behaviour is not gratuitous but instead that the gratuitous nature is implied.

(31) *This Troper finds himself using [[Gintama]] "Zura ja nai, Katsurada!" as a swear, does this also count as a [[Gosh Darn It to Heck]]?*

4.2.2 Negative Attitudes

An attitude identifies in the corpus was considered negative if it seemed to be disparaging the language type or language writer, either self-referential or directed towards others. Given the negative stereotype of a weeaboo and the topic of discussion for this corpus source being explicitly labelled as gratuitous, negative attitudes were expected to be present and more prolific than positive or neutral attitudes. While mainly from the Troper Tales corpus, a number of the attitudes also resonated within the /CGL/ corpus, particularly where interactions involving negative stereotypes related to language use were tagged as evidence for qualitative discussion.

An attitude that is common to both the current research and to internet language in general is reference to the behaviour occurring at a young or past age, either in the self or others. In Example (32) below the writer mentions the behaviour in his past, claiming that while he believed it was 'cool' at the time, now he is older he considers it embarrassing behaviour that he has since left in the past. May be considered a synonym for how long someone has been in

the community and thus having more knowledge of the community rules as well as the general life knowledge and maturity that comes with actual age. Presuming youth or inexperience (such as being labelled a noob[40]) is a common insult or shut down tactic in general internet culture.

(32) *This troper suffered this fresh off his trip to Japan when he was still a [[OldShame]] failtacular teenager who thought he was cool randomly injecting Japanese swears into conversations. Now he leaves poorly pronounced Japanese to his Otaku friends, who now look as silly as he once did. They'll grow out of it…*

References to the linguistic behaviour or the person producing it as being annoying or unacceptable were also found, indicating that the language is objectionable in some way rather than being widely accepted or routine, such as in Example (33). This gives evidence for there being a negative stereotype present in the Japanese Popular Culture Fandom, of which certain usage of Japanese language is a feature. Example (34) is from the /CGL/ corpus, where a writer lists a number of what they consider undesirable behaviours that may be encountered at a Popular Culture Convention, one of which concerns language. Example (34) seems to be referring to a stereotypical phrase that is perhaps overused, and thus has a number of attitude connotations attached to it within the wider community beyond its basic translation (for further analysis of the phrase *kawaii desuuuu, sugoi!!!* see Chapter 5.1.6).

(33) *This troper knew a whole group of girls that did this. At first it was, like, cool, you guys like anime, too but it quickly got annoying, especially since all of them though it was so clever that they could do this.*

(34) *[What's the one thing at a con that immediately flags warning signs about a person for you?]*
>kawaii desuuuu, sugoi!!!
>I bet he/ she's not even a real fan
>generally being unnecessarily loud, obnoxious and squealing
>Yaoi paddles, chan memes and 'doing it ironically'
>'sluts'
>leaving rubbish

Fluency or linguistic competence was mentioned by writers, often to justify a writer's usage of Japanese, or, as in Example (35) below, to enforce a boundary for other writers. The writer is aware of the language politics within the community, and has explicitly positioned themselves in line with them, perhaps so as not to over-reach in their own behaviours and thus be open to attack; stating their level, or lack thereof, of fluency in Japanese before another writer can exploit it as a weakness. For example, a troll might attack the spelling and grammar of a post,

[40] Slang for someone who is a newcomer and therefore does not have certain skills or knowledge.

and through which attack the writer's notion of face and identity they have only been able to construct through written word (Hardaker, 2013).

(35) *If you speak a word of Japanese in everyday conversation (unless completely appropriate in context) [[BerserkButton]] you'd better be fluent. (Note that I'm not fluent, but I don't insert random words in sentences ever, either.)*

The attitude in Example (34) also highlights an important issue considered in the theoretical paradigms such as crossing (Rampton, 1995) and polylanguaging (Jørgensen, 2008); fluency is a contributor to whether a writer has ownership over the language, or at the very least access to using it. Thus, it is a view within the community that if someone uses Japanese and is indeed fluent to a socially acceptable degree, that justifies their usage of the language, and if someone is not fluent in Japanese, they should not attempt random language mixing.

The writers also felt the need to defend the use of the language they were using in general. By doing so, the writer is stating that their language behaviour is not due to ignorance of the community attitudes but instead despite them. They are aware that the language may not be considered acceptable by many of their peers, but other factors have influenced them to use it anyway. In Example (36) the writer is claiming that the language use is actually situationally appropriate, but also that 'fun' has outweighed other influences.

(36) Top troper also defends his use of "Yare yare daze" as being situationally appropriate, muttered under his breath most of the time, and really fun to say.

Alongside fluency, current pronunciation was considered important, although this is not an issue in written speech. Mispronunciation, one of the features mentioned in Example (37), is in particular seen as typical of weeaboos, who are thought of as ignorant or disrespectful. Incorrect pronunciation, particularly in cases where there is seen to be no proper attempt made, can indicate a disregard for the wider presence of Japanese language and Japanese culture beyond a shallow experience of media. This is perhaps particularly notable in Anglophone spaces, due to the colonial history of English-speaking countries and how globalisation interacts with appropriation.

(37) This troper used to speak like this. Now she knows a group of girls who do this all the time, and it is one of the most annoying things she's ever heard. The super-squeaky voices, the gratuitous honorifics, the terrible pronunciation…She is sincerely sorry she ever acted that way.

Finally, the terms weeaboo and *otaku* are frequent descriptors within the /CGL/ and Troper Tales corpora. As previously noted, they are terms used to indicate a person has a high degree of interest in Japanese popular culture, but whilst also connoting a set of stereotypical obsessive

behaviours that have a negative effect on the person's behaviour and personality. People may describe themselves with these terms in order to emphasise or label their interest (indicating solidarity with an ingroup), but positive usage does not preclude the negative associations. Instead, the self-labelling of these terms can be seen as being either self-deprecating or ignorant. *Otaku* has a milder connotation in Anglophone Japanese culture fandom than it does in Japan, perhaps as the word has semantically narrowed to indicate a fan of Japanese media rather than obsessive fandom in general, but the knowledge of the roots of the word is still part of the community. Mostly, people wish to avoid being labelled as a weeaboo, and may self-monitor their writing to avoid it as in Example (38).

> *(38) This troper has the tendency to shout "YATTA!" when something good happens to her, and is also prone to saying "itadakimasu" before she eats, but other that those two phrases, that is the extent of her use of Gratuitous Japanese, as she does not want to be seen as a weaboo.*

A common theme in the rather cruel descriptions of typical weeaboo behaviour found by McGee (2012) is that of boundary violations, including physical elements like 'at, dirty, smelly and social elements like loud and inappropriate. The violation of linguistic and ethnic boundaries must also be noted, as shown in the attitude data found in this research. Linguistic policing is just as frequent as policing of other social behaviours.

The previous sections have identified that there are significant language attitudes towards specific behaviours, including but not limited to linguistic behaviours, and therefore that they have an effect, to some degree, on which linguistic behaviours are acceptable and thus likely to be allowable in the community. Studies of code-switching have long established that mixing language behaviours are generally disapproved of and thought to be lazy from both an internal and external perspective (Gardner-Chloros, 2009), and the language behaviours described within crossing and polylanguaging definitions have similar views but with an extra factor of disapproval in terms of a lack of accepted ownership of the language (Rampton, 1995; Jørgensen et al, 2011). Therefore, with data taken from an online space where strong attitudes, along with other common internet attitudes and powerplay dynamics, exist, the level or extent of Japanese language use can be considered to be a basic restricted level that sits closer to language borrowing on the language change continuum. Therefore, the words are likely to be ones that are familiar to a wide range of people within similar affinity spaces, aside from very specific technical terms from the more niche topics of discussion which would be important vocabulary terms to them but not as well known outside of that subsection. This indicates the fact that language usage is policed within the community, which will be explored in the following section.

4.2.3 LANGUAGE POLICING

One of the particularly interesting features of the current study is the global status of the Japanese language and Japanese culture. Language policing in this case raises more issues related to ethnicity and appropriation (i.e. racism) than, for example, the use of English, a widely spread globalised language tied to a historically, and still presently, colonising force. Intracommunity debates on correct terminology and meanings of specific terms, and the general trends of attitudes towards the use of Japanese language elements in otherwise English language contexts is evidence for the fact that language use is not exempt from constraints or gatekeeping within the community; some aspects are accepted and even encouraged, while others are discouraged or actively policed.

Rampton and Charalambous (2010, p.5) note that crossing behaviours, or language mixing in which speakers have little claim to ownership of the language variety, "may strengthen boundaries, undermine them, or assert their irrelevance." The different factors influencing and the results themselves of language use uncovered in the current research indicates that language and power intersect in society; language is a tool for wielding power, and language is an area of contention wherein power is wielded. Ahearn (2012) considers the concept of hegemony to be particularly important to understanding the intersection of language and power, in which domination does not only occur based on violence or economic control but rather on political, cultural and institutional influence.

As language moves through time and space, and as human actors enact and experience conflicts, negotiations, and resolutions that create these historical connections, language is revealed to be a social and political object (Blommaert, 2010) alongside, or indeed because of, its general function as a communicative object. People have various explicit and unconscious views on language, both positive and negative, which can result in the regulation of society through the regulation of language itself, from more extreme situations as controlling accessibility to an entire language to specific people or attempts at total erasure to intracommunity conflict over the number of loanwords accepted in a niche speech community's dialect. Therefore, for any attempts at describing a community or affinity space, and any research done on that space, it is important to not only examine what language is used, but also what is not used in terms of what seems to be explicitly or implicitly disallowed. In Kytölä's (2012) longitudinal study of English use on a Finnish football forum found that while some usage of playful non-Standard English was accepted, a number of users who were deemed to be using 'bad English' unintentionally ended up being ridiculed to the point of leaving the forum, or at least no longer participating.

As the /CGL/ corpus is deemed to be particularly constrained by negative opinions on language use (see Chapter 3.2.1), the instances of Japanese language found could perhaps be considered ones that are universally accepted, and, to relate it to classifications and theories in the wider research area, more likely to be on the borrowing end of the language change continuum (see Chapter 5.2). The most common Japanese words found, or even all of them due to the policing discouraging unnecessary language use, can be considered loan words within the community even though they are not considered such in mainstream English, as defined by their presence or non-presence in official English-language dictionaries.

The analysis of language attitude trends can be combined with the analysis of the language features in order to comment on how the language is used. Stott (2006) saw the linguistic behaviour being used to show in-group membership or solidarity, for narrative effects such as for framing or humour, and for discussion of Japanese concepts or to show respect for Japanese culture and people. Leppänen et al (2009) found English being used by Finnish youths to function as a resource for enabling expertise in situations like console game playing where the pop culture language is originally English or to be used to construct the linguistically, culturally and socially hybrid identities found in modern activities like hip-hop music or fanfiction. Turner (2016) found factors such as translation issues and common usage within the fandom to influence language use[41] in *anime* fanfiction, as well as to give expected flavour to the story and for identity creation.

Similar purposes were found in the current research. While most of the language use leaned more towards using correct terminology (see Chapter 4.1.3), there were also instances where the language was used to create a playful tone, such as in Example (39) where the word is spelt in a way to come across as cute or coy, perhaps referencing language patterns from Japanese *anime*. Other instances, such as Example (40), use Japanese language in a mocking way, drawing reference to the weeaboo stereotype in order to bolster their sarcasm. Other functions found in the corpus were using Japanese for creating an identity and in-group solidarity, as well as signalling knowledge where knowledge is a commodity.

(39) *Hi-mi-tsu~* [Secret]

(40) *oh anon, didn't you know marvel-lolita was so kawaii desu ne???* [it's cute, isn't it?]

[41] According to questionnaire answers answered by the fanfiction authors (Turner, 2015).

The next section will delve into how the current case relates to wider linguistic theory, such as what the extent and form of the language features indicate in terms of language learning and the language change continuum.

5 FURTHER DISCUSSION

Whilst the previous chapter included discussion of the results in terms of Research Questions 1 and 2, the following sections will discuss what those results mean in terms of current theoretical paradigms in Linguistics, and answering Research Question 3: How do the patterns of language use and attitudes found within the data relate and contribute to current theoretical paradigms of language mixing? The extent of the language found and the sociocultural features that position the language in the wider community and global landscape indicate the following points of interest: language as a set of resources and as a commodity, and how foreign language learning can be related to other types of language learning and innovation; and how borrowing and code-switching behaviours are diachronic and synchronic aspects of a language change continuum.

5.1 LANGUAGE AND LANGUAGE LEARNING

The intersection of language and power was broached in Chapter 4.2.3, detailing the intra-community language policing that appeared to be present as the result of language attitudes about the acceptability of use of Japanese language elements. It was found that 'correct' terminology was important, according to historical or community agreed upon norms of meaning, particularly in terms of technical concepts, and that Japanese language knowledge was given high status. However, it was also indicated that there were correct situations in which language was acceptable to be used and not so much outside of that, and those who did not heed the community norms were likely a weeaboo, which denotes youth, ignorance, and obnoxiousness.

Language, in its position as a tool for wielding power as well as being the target of power negotiations, is part of the symbolic capital that Bourdieu (1977) claims can be used interchangeably with material capital in the marketplace of human interaction, and that language is a commodity. Heller (2010) believes that whilst this has always been true, the importance and value of language as a resource within the marketplace has grown due to the political economic conditions of a late capitalism society, such as increasing globalisation. One of the results of this is that the perception of language has shifted to seeing language as a marketable commodity of its own from being mainly a marker of ethnonational identity (Heller, 2003).

In the case of the current study, language becomes untethered from the original ethnonational identity to a certain degree, and then, due to being the language of the popular culture products, the language becomes an intrinsic part of the many aspects of popular culture that people enjoy and use to interact with and construct their identities.

As language is commodified and transmitted in modern society, sociolinguistics now prefers to consider language in terms of a speakers' repertoire, which is made up of the variable and often fragmentary sets of semiotic resources, styles, registers and genres the speaker knows. This allows for analysis beyond previous assumptions about any fixed links between language, community, time and place (Blommaert & Backus, 2011; Blommaert, 2010). Whilst this concept has been at the heart of sociolinguistics for some time, recent direction has taken it from being linked to a speech community to being linked to the individual (Blommaert & Backus, 2013).

Viewing language as a repertoire filled with various language related features therefore affects how language learning processes are viewed. The variable and fragmentary nature of a person's repertoire indicates that their language learning experiences or exposure to language can also be variable and fragmentary. Blommaert and Backus (2013) detail different levels of language learning which can result in different levels of competency in a speaker. Relevant to the current study is the minor forms of language learning. Language Encounters involve generally moments or times in a speaker's life that they have encountered and learned small bits of language, such as age-group slang learning, temporary language learning, single-word learning, and recognising language. Whilst no active competence is acquired in many of these types, they are still encounters with language that have left some, even if only small, mark on an individual's repertoire.

One does not need to be in a bilingual community or be actively learning a language in order to gain experience of the patterns and features of a foreign language, although those are spaces where decent language immersion occurs; instead, even if only at a minimal level, it can occur with immersion in foreign language popular culture. A participant in Williams (2005, p.168) claimed "Even my roommate knows enough Japanese that [bad translations] bother her, and she's never taken a class", and Turner (2015) found that there was no significant difference between Learners and Non-Learners in their English-Japanese language mixing behaviours in *anime*-based fanfiction.

In terms of popular culture, Jenkins (2006) describes the idea that knowledge is hive mind based in online communities, where not everyone is required to know all the knowledge and instead

it is shared communally. This is interesting in terms of Blommaert and Backus' (2013) moving away from repertoires based on the speech community as a whole to those of individual speakers. Popular culture knowledge can also be tied to the social factors restricting language use as fans are motivated in their interactions with each other by a pleasure in sharing knowledge as well as in the knowledge itself (Jenkins, 2006). Another aspect of the Anglophone Japanese fandom community is that there are speakers of differing Japanese competence interacting with non-learners, and many of the non-learners are aware of and often participate in discussions of language meaning, etymology, and policing of acceptable language.

The repertoire and usage-based approach to language learning and language knowledge, when considered in terms of power and commodification and popular culture, indicates that foreign language learning and exposure is not dissimilar in process to other language processes of innovation, such as fanspeak, in-group language, jargon, slang, and memes. This ties globalisation and other sociolinguistic elements of the modern world with language change processes, and adds support to the idea of language change being a continuum that can vacillate back and forth at an individual and community level.

5.2 Language Change Continuum

Backus (2014) states that because a loanword is a foreign-language word that is accepted to some degree amongst the lexicon of the borrowing language, it follows that the qualifier loanwords is frequency. A code-switch involves shifting to linguistic features from another language mid-utterance or mid-discourse, in which there is awareness of the different or foreign nature of the shift. Whilst not full accepted, modern language contact and language mixing research is tending towards agreeing that these two factors are not exclusive of each other. In particular it is acknowledged that there is little reason or need for synchronic distinction between them (Myers-Scotton, 2002), whilst Backus (2014) asserts that a synchronic analysis of a word in a corpus cannot reveal much about the degree to which it is integrated. This is a move from the structural and nativisation aspects of integration that is investigated in older structural approaches to wanting to know about cognitive representation, such as the degree of entrenchment in the speaker's repertoire and how conventional it is considered in the speaker's speech community.

If, in the process of language change, foreign language words can be introduced and then become established as loanwords, then it stands to reason that there are different stages

involved, and that at any given point in time different words or linguistic elements may be at different points along the way.

In the current study, the distance, both literal and metaphorical, the speakers have from native speakers could perhaps mean considering the language phenomenon as lexical borrowing rather than code-switching, especially as many speakers have little knowledge of Japanese outside of what is commonly used in the community. However, as there are instances of switching for specific discursive effect with full knowledge of the Japanese origin (see Chapter 4.1.6) in the same space as borrowing, it therefore makes sense to reconcile language borrowing and codeswitching as different time dimensions of a single process. This view also allows for the fluctuation of use of particular linguistic elements (Backus, 2014) as language learning is not necessarily a permanent process, and it is reliant on usage and reinforcement; hence linguistic theory preferring the terminology language learning rather than language acquisition (Blommaert & Backus, 2011).

The question of fluency can be considered a feature impacting language use and change that also has social and political results, rather than a defining characteristic of a language type definition. The English-Japanese language mixing shown in this study is not by any means skilful bilingual behaviour (code-switching), nor is it employed as a strategy to make up for deficiencies in language knowledge (Second Language Teaching).

However, it must be noted that a limitation of the current study in terms of a language change continuum is that the scope is not longitudinal (see Chapter 3.2) and instead is a synchronic cross-section of the language use in a specific community space, and thus does not address language change over time. Other linguistic elements discussed, relating to ideas of language exposure and attitudes, do seem to allow that a usage-based approach that views language change and mixing as intrinsically linked is useful.

With the development of the usage-based approach proposed by Backus (2014), which includes this language change continuum, there still needs to be an emphasis on an awareness of the social, ethnic, and language boundaries that are often felt very keenly by speakers, and that these issues can often lead to a struggle within a speaker or a community of how to reconcile those boundaries and ideas of ownership and legitimacy beyond the linguistic analysis abstraction level. To clarify, the social and political influences on language use in this approach is seen as the driving force rather than underlying structural and grammatical constraints, but it also needs to be made clear that speakers of varying levels of power and influence in the

linguistic marketplace have their own views about what language boundaries exist and why, whether or not there is any truth to it.

6 CONCLUSION

This study aimed to explore the Japanese language usage in online Anglophone Japanese popular culture fandom, through the quantitative and qualitative analysis of linguistic data gathered into a corpus from two online forums. Every instance of Japanese language was tagged according to its word class, as well as some other features for qualitative analysis.

The first Research Question asked about the extent to which Japanese language is found, and what linguistic forms and features were present. The study focussed on instances of *romaij* orthography as it was the large majority of words found and also necessary due to researcher limitations. The most frequently used word classes were nouns and attributive nouns, which meant that most of the words were content words that could be further categorised according to technical, general, and ritual usage. The minimal instances of function words were grammatical particles and terms of address. The words could also be categorised into a few semantic domains, including specific interests or hobbies, food, clothing and fashion, media related terms, and religion and cultural terms. The extent or integration was looked at in terms of English suffixation, Japanese zero morpheme plurals, Japanese abbreviation patterns, and words that seemed to replicate Japanese pronunciation patterns. Loanwords and multi-word segments were also investigated, and link to answering Research Question 3.

The second Research Question asked what attitudes towards the language use were found in the community and how they may affect language use, and this was answered from qualitative analysis of a corpus where writers commented on their real-life experiences with the trope of Gratuitous Japanese. The attitudes found included how, although Japanese language knowledge was seen as high status, the language use could be considered annoying or obnoxious if used outside of acceptable contexts, the qualifiers for which seemed to include the audience and situation in which it occurred as well as the fluency level of the speaker. These attitudes, along with the other results, showed that as language is used to enact power tactics in a community, power is also enacted on the language itself. The presence of intra-community language policing was interesting because of what it meant for global patterns of power as it was not speakers who have ownership over Japanese doing the policing, but other members of the fandom community.

The third Research Question asked how the patterns of language use and attitudes found in the previous questions relate to current linguistic theory in the field of language mixing. Specifically, the English-Japanese language mixing found had more features in common with the borrowing

side of the language change continuum as opposed to than the code-switching side, although there were multi-word segments present in which a switch occurred for sarcastic effect by styling as weeaboo and all the assignations that come with that stereotype. A current linguistic theory calls to view borrowing as the diachronic element of language change while code-switching is a synchronic event on the same continuum of language change. The current study supports this as the separation of loanwords from other language mixing runs into issues when a decent degree of origin and assignation knowledge is still present.

Although the Anglophone online Japanese culture fandom should still be considered rather isolated and niche as a community or affinity space, as opposed to mainstream culture, and the focus of the current study an even smaller sub-section, the current study nevertheless has implications for the wider breadth of factors that code-switching and language crossing theoretical paradigms need to encompass in an age of increasing globalisation and superdiversity. The internet is a space that allows for new types of language and culture interactions. Rather than physical contact and interaction, such as with migrant minority communities interacting with majority communities or along geographical borders, giving rise to lexical borrowings or code-switching strategies, now languages can, and do, travel over distances, facilitated through technology and popular culture exports and resulting in almost a separate thing from the ethnonational origin. Further, the internet allows for people of varying sociocultural and geographical backgrounds who have been exposed to that same language to form a community of practice in the online space, which then creates and reinforces its own language norms and policies. The key point here is that the new language community, aided and shaped both by the dimension of the online space and by the form of language contact, takes the imported language and interacts with it on its own terms, through the lens of cultural differences and distance, for its own purposes, including the occurrence of code-switching type behaviours. These are informed by that particular situation to, in many instances, manifest somewhat differently to those in more traditional types of language contact cases.

An avenue for future research would be to investigate the repertoires of Japanese culture fandom members in terms of what language they know rather than just what they use. Creating a vocabulary test based on the language data found in the current study for participants to complete would allow a researcher to gain more insight into the depth of Japanese knowledge participants have. As previously established in this thesis, the word list in Appendix 4 can be considered to be a base level of language knowledge within the fandom community, as their presence in the corpus indicates a lack of controversy over their usage, although some of the terminology is specific to less common cultural products.

Further semantic analysis would be an interesting issue for future research, such as a comparison of the usage of common Japanese words with their near-equivalent English versions to investigate the process of semantic narrowing and broadening. For example, *cosplay* is used 1597 times in the /CGL/ corpus in comparison to costume having 305 hits, and *kawaii* is used 74 times compared to cute with 648 hits. It is hard to comment on this in the scope of the current study as the /CGL/ corpus contains discussion specifically moderated to be within a Japanese culture fandom topic, and thus all words are generally used within that context. Whilst *cosplay* also includes discussion of non-Japanese fandoms and costuming, it does not overcome the above limitation. Backus (2014) believes a limitation of corpus studies is the lack of variability of texts and thus situations, which limits the generalisations that can be made about language use. It would also be interesting to investigate how many of the Japanese tokens can be considered untranslatable words with cultural concepts or innovation unique to Japan, and how this relates to the borrowing and code-switching debate. This feature was originally considered for the current study, but it was hard to keep consistent, particularly with the other features being tagged for.

7 REFERENCES

Ahearn, L. (2012). *Living language: An introduction to linguistic anthropology.* Chichester, West Sussex, United Kingdom: Wiley-Blackwell.

Alexa Internet (2017). Retrieved from https://www.alexa.com/siteinfo

Androutsopoulos, J. (2011). From variation to heteroglossia in the study of computer-mediated discourse. In C. Thurlow, & K. Mroczek (Eds.), *Digital Discourse: Language in the New Media* (pp. 277-298). New York, NY: Oxford University Press.

Azuma, S. (1997). Lexical categories and code-switching: A study of Japanese/English code-switching in Japan. *The Journal of the Association of Teachers of Japanese,* 31(2), 1-24.

Backus, A. (2014). A usage-based approach to borrowability. In E. Zenner, & G. Kristiansen (Eds.), *New Perspectives in Lexical Borrowing* (pp.19-39). Boston, MA: Walter de Gruyter.

Backus, A. (2014). Towards a usage-based account of language change: Implications of contact linguistics for linguistic theory. In *Questioning Language Contact* (pp. 91-118). Leiden, The Netherlands: Brill.

Backus, A. (2015). A usage-based approach to code-switching: The need for reconciling structure and function. In G. Stell, & K. Yakpo (Eds.), *Code-switching Between Structural and Sociolinguistic Perspectives* (pp. 19-38). Berlin, Germany: Walter de Gruyter.

Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: much more than bean counting. *TESOL Quarterly, 35(2),* 331-336.

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

Bloem, W. (2014). Japanese fanspeak in the Anglophone manga and anime fan culture. (Unpublished masterate thesis). Leiden University.

Blommaert, J., & Backus, A. (2011). Repertoires Revisited: 'Knowing language' in Superdiversity. Working Papers in Urban Language & Literacies No. 67. Available at https://www.kcl.ac.uk/ecs/research/research-

centres/ldc/publications/workingpapers/abstracts/wp067-repertoires-revisited-knowing-language-in-superdiversity

Blommaert, J., & Backus, A. (2013). Superdiverse repertoires and the individual. In I. de Saint-Georges, & J. Weber (Eds.), *Multilingualism and Multimodality: Current Challenges for Educational Studies* (11-32). Rotterdam, The Netherlands: Sense Publishers.

Blommaert, J. (2010). *The sociolinguistics of globalization.* Cambridge, United Kingdom: Cambridge University Press.

Bodomo, A. (2009). *Computer-Mediated Communication for Linguistics and Literacy: Technology and Natural Language Education.* Hershey, PA: IGI Publishing.

Bourdieu P. 1977. The economics of linguistic exchanges. *Information (International Social Science Council), 16(6)*, 645–68.

Cannon, G. (1984). Zero plurals among the Japanese loanwords in English. *American Speech, 59(2)*, 149-158.

Cannon, G. (1994). Recent Japanese borrowings into English. *American Speech, 69(4)*:373-397.

Doerr, N., & Kamagi, Y. (2014). Power of language idelogies: Challenging the notion of foreign loanwords in Japanese-as-a-foreign-language classroom. *Writing Systems Research, 6(2),* 148-166.

Fabb, N. (2001). Linguistics and literature. In M. Aronoff, & J. Rees-Miller (Eds.) *The Handbook of Linguistics* (pp. 447-465). Oxford, United Kingdom: Blackwell Publishing Ltd.

Fukunaga, N. (2006). "Those anime students": Foreign language literacy development through Japanese popular culture. *Journal of Adolescent and Adult Literacy, 50(3),* 206-222.

Galbraith, P. W. (2009). Moe: Exploring virtual potential in post-millennial Japan. *electronic journal of contemporary japanese studies*, *5.* Retrieved from https://www.japanesestudies.org.uk/articles/2009/Galbraith.html

Gardner-Chloros, P. (2009). *Code-switching.* Cambridge, United Kingdom: Cambridge University Press.

Gee, J. (2003). *What video games have to teach us about learning and literacy.* New York, NY: Palgrave Macmillan.

Gliozzo, A. (2006). Semantic domains and linguistic theory. In *Proceedings of the LREC 2006 workshop "Toward Computational Models of Literary Analysis",* Genoa: Italy.

Gottlieb, N. (2010). The romaji movement in Japan. *Journal of the Royal Asiatic Society, 20(1)*:75-88.

Hardaker, C. (2013). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research, 6 (2),* 215-242.

Heller, M. (2003). Globalization, the new economy and the commodification of language and identity. *Journal of* Sociolinguistics, *7(4)*, 473–92.

Heller, M. (2010). The commodification of language. *The Annual Review of Anthropology, 39*, 101-14.

Herring, S., Stein, D., & Virtanen, T. (2013). *Pragmatics of Computer-Mediated Communication.* Berlin, Germany: Walter de Gruyter.

Hill, J. (1999). Styling locally, styling globally: What does it mean? *Journal of sociolinguistics, 3(4),* 542-556.

Honna, N. (2008). *English as a multicultural language in Asian contexts: Issues and ideas.* Tokyo, Japan: Kurioso Publishers.

InfiniteZenith & NinjaClown (2013). *Tv Tropes and the rise of pseudo-intellectualism.* Retrieved from https://infinitemirai.wordpress.com/2013/08/14/tvtropes-and-the-rise-of-pseudo-intellectualism/

Irvine, J., & Gal, S. (2000). Language ideology and linguistic differentiation. In P.V. Kroskrity (Ed.) *Regimes of Language: Ideologies, Politics, and Identities* (pp. 35-84). Santa Fe, CA: School of American Research Press.

Irvine, J., & Gal, S. (2000). Language ideology and linguistic differentiation. *Language Ideology and Linguistic Differentiation*, 35-84.

Iwabuchi, K. (2002). *Recentering globalisation: Popular culture and Japanese transnationalism.* Durham, United Kingdom: Duke University Press.

Jenkins, H. (2006). *Fans, bloggers, and gamers: Exploring participatory culture.* New York, NY: New York University Press.

Jørgensen, J. (2008). Polylingual languaging around and among children and adolescents. *International Journal of Multilingualism, 5(3)*:161-176.

Jørgensen, J. N, Karrebaek, M. S., Masden, L. M, & Møller, J. S. (2011). Polylanguaging in superdiversity. *Diversities, 13(2)*:23-37.

Kytölä, S. (2012). Multilingual Web Discussion Forums: Theoretical, Practical and Methodological Issues. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse.* New York, NY: Routledge.

Lee, C., & Barton, D. (2012). Multilingual texts on Web 2.0: The case of flickr.com. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse.* New York, NY: Routledge.

Lee, J. S. (2005). Discourses of fusion and crossing: Popular culture in Korea and Japan. (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.

Lee, J. S., & Moody, A. (2012). Sociolinguistics and the study of English in Asian popular culture. In J. S. Lee, & A. Moody (Eds.), *English in Asian popular culture* (pp. 1-12). Hong Kong: Hong Kong University Press.

LePage, R., & Tabouret-Keller, A. (1985). *Acts of identity: Creole-based approaches to language and ethnicity.* Cambridge, United Kingdom: Cambridge University Press.

Leppänen, S. (2008). Playing with and policing language use and textuality in fan fiction. In I. Hotz-Davis, A. Kirchhofer, & S. Leppänen (Eds.), *Internet Fictions* (pp. 62-83). New Castle, United Kingdom: Cambridge Scholars Publishing.

Leppänen, S. (2012). Linguistic and generic hybridity in web writing: The case of fan fiction. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse.* New York, NY: Routledge.

Leppänen, S., & Piirainen-Marsh, A. (2009). Language policy in the making: An analysis of bilingual gaming activities. *Language Policy, 8(3)*, 261-284.

Leppänen, S., Pitkänen-Huhta, A., Pirrainen-Marsh, A., & Peuronen, S. (2009). Young people's translocal new media uses: A multiperspective analysis of language choice and heteroglossia. *Journal of Computer-Mediated Communication, 14(4)*, 1080-1107.

Loveday, L. (1986). *Explorations in Japanese sociolinguistics.* Amsterdam, The Netherlands: John Benjamins Publishing Company.

Mahootian, S. (2012). Repertoires and resources: Accounting for code-mixing in the media. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse.* New York, NY: Routledge.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: method, theory and practice.* Cambridge, United Kingdom: Cambridge University Press.

McGee, J. (2012). Discipline and post: Foucault and "Weeaboo horror stories" on the internet. *愛知淑徳大学論集. グローバルカルチャー・コミュニケーション研究科篇, 4,* 49-61.

Michaels, S. (2008, March 19). *Taking the Rick.* Retrieved from https://www.theguardian.com/music/2008/mar/19/news

Miller, L. (1998). Wasei eigo: English "loanwords" coined in Japan. In J. H. Hill, P. J. Mistry, L. Campbell (Ed.), *The life of a language: Papers in linguistics in honor of William Bright* (pp. 123-139). Berlin, Germany: Mouton de Gruyter.

Møller, J., & Jørgensen, J. (2009) From language to languaging: Changing relations between humans and linguistic features. *Acta Linguistica Hafniensia, 41(1)*:143-166.

Moody, A., & Matsumoto, Y. (2012). Lu-go and the role of English loanwords in Japanese: The making of a 'pop pidgin'. *Asian Englishes Today: English in Asian Popular Culture*, 103-126

Moody, A., Matsumoto, Y. (2012). Lu-go and the role of English loanwords in Japanese: The making of a 'pop pidgin'. In Lee, J. S., & Andrew Moody (Eds.), *English in Asian Popular Culture* (pp. 103-126). Hong Kong: Hong Kong University Press.

Moody, A. (2006). English in Japanese popular culture and J-pop music. *World Englishes, 25(2),* 209-222.

Morimoto, Y. (1999). Making words in two languages: A prosodic account of Japanese-English language mixing. *International Journal of Bilingualism, 3(1),* 23-44.

Muysken, P. (1995). Code-switching and grammatical theory. In L. Milroy, & P. Muysken (Eds.), *One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching* (pp. 177-198). Cambridge, United Kingdom: Cambridge University Press.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes.* Oxford, United Kingdom: Oxford University Press.

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

Otake, M. (2010). English loanwords from Japanese: A survey of the perceptions of American English speakers. *Tokyo Seitoku University Bulletin, Faculty of Humanities and Applied Psychology*, *17*, 17-32.

Otaku, n, n.d. Retrieved from https://www-oed-com.ezproxy.massey.ac.nz/view/Entry/246749?redirectedFrom=otaku#eid

Rampton, B., & Charalambous, C. (2010). Crossing: A review of research. Working Papers in Urban Language & Literacies No. 58. Available at https://www.kcl.ac.uk/ecs/research/research-centres/ldc/publications/workingpapers/the-papers/58.pdf

Rampton, B. (1995). *Crossing: Language and ethnicity among adolescents.* London, United Kingdom: Longman.

Rampton, B. (2010). From 'multi-ethnic urban heteroglossia' to 'contemporary urban vernaculars'. Working Papers in Urban Language & Literacies No. 61. Available at https://www.kcl.ac.uk/ecs/research/research-centres/ldc/publications/workingpapers/abstracts/wp061-from-multi-ethnic-urban-heteroglossia-to-contemporary-urban-vernaculars

Saturday Night Live. (2011, December 10). *J-Pop talk show: Japanese culture enthusiasts* [Television series]. Retrieved from https://www.youtube.com/watch?v=2DwpXmTSojA

SCOUT-Soft. (n.d.). NPP Combine (v.1.0.108) [Computer software]. Retrieved from http://www.scout-soft.com/combine/

Stott, N. (2006). A study of code-switching in newsletter articles written by native English speakers residing in Japan. *福岡県立大学人間社会学部紀要*, *14(2),* 31-50.

TextFX. (v.0.2.6) (n.d.). Retrieved from https://sourceforge.net/projects/npp-plugins/files/TextFX/

Turner, M. (2015). Bocchan, it's time to wake up: A corpus-based analysis of how any why non-learners code-switch in Japanese fandom-based fanfiction. (Unpublished masterate thesis). University of Helsinki.

Tyson, R. (1993). English loanwords in Korean: Patterns of borrowing and semantic change. *El Two Talk, 1(1)*, 29-36.

Varis, P., & Wang, X. (2011). Superdiversity on the internet: A case from China. *Language and Superdiversities, 13(2),* 71-83.

Waifu. (n.d.). Retrieved from https://knowyourmeme.com/memes/waifu

Wapanese. (n.d.). Retrieved from https://www.urbandictionary.com/define.php?term=Wapanese

Web Japan. (n.d.). *Fashion: From court dress to designer brands.* Retrieved from https://web-japan.org/factsheet/en/pdf/e34_fashion.pdf

Weeaboo. (n.d.). Retrieved from Know Your Meme: https://knowyourmeme.com/memes/weeaboo

Wiese, H. (2015). "This migrants' babble is not a German dialect!": The interaction of standard language ideology and 'us'/'them' dichotomies in the public discourse on a multiethnolect. *Language and Society*, *44 (3)*:341-368.

Wiki. (n.d). Retrieved from https://en.wikipedia.org/wiki/Wiki

Williams, K. (2006). The impact of popular culture fandom on perceptions of Japanese language and culture: The case of student anime fans. (Unpublished doctoral dissertation). University of Texas, Austin.

Winford, D. (2003). *An introduction to contact linguistics.* Oxford, United Kingdom: Blackwell.

Zenner, E., & Kristiansen, G. (2014). Introduction: Onomasiological, methodological and phraseological perspectives on lexical borrowing. In E. Zenner, & G. Kristiansen (Eds.), *New Perspectives in Lexical Borrowing* (pp.1-18). Boston, MA: Walter de Gruyter.

8 APPENDICES

Appendix 1: Comparison of the current study's tags to previous research's tagging systems

Gardiner (2017) Tagging System		**Turner (2015) Tagging System**		**Stott (2006) Tagging System**	
Tag	Description	Tag	Description	Tag	Description
K	Japanese orthography				
JK	Japanese kaomoji				
^	Romaji				
JPHT	Japanese part of speech: honorific title	**JHTI**	Japanese honorific title		
JPHS	Japanese part of speech: honorific suffix	**JHSU**	Japanese honorific suffix		
JPV	Japanese part of speech: verb	**JPVE**	Japanese verb		
		JPAD	Japanese adverb		
JPNO	Japanese part of speech: noun	**JPNO**	Japanese noun		
JPNA	Japanese part of speech: attributive noun	**JPAT**	Japanese attributive noun		
JPPA	Japanese part of speech: particle	**JPPA**	Japanese particle		
JPIE	Japanese part of speech: interjection or expression	**JPIE**	Japanese interjection/expression		
JPSS	Japanese part of speech: sound symbolism				
JPPR	Japanese part of speech: pronoun	**JPPR**	Japanese pronoun		
JPAD	Japanese part of speech: adjective	**JPAJ**	Japanese adjective		
JNS	Japanese name: series	**JMNV**	Japanese version of the canon; includes abbreviations		
JNC	Japanese name: character, person, or place name		Japanese version of a character name from the canon		
JNE	Japanese name: Anglophone based company, event, or username				
JL	Japanese loan word				
JF	Japanese phrase				
JAC	Japanese anime catchphrase				
JMSH	Japanese modification: shortening				

Gardiner (2017) Tagging System		Turner (2015) Tagging System		Stott (2006) Tagging System	
Tag	Description	Tag	Description	Tag	Description
JMPR	Japanese modification: pronunciation	**JMNV**[42]	Japanese version of a character name from the canon		
JMPL	Japanese modification: plural	**JGPL**	Japanese plural		
EM	English modification	**EGPL**	English plural	**Intrasentential: Word internal**	Changes within sentence or clause: mixed with English affixes
		EGGE	English genitive (possessive *'s*)		
DS	Discussion: stereotype				
DM	Discussion: metalinguistic				
				Intrasentential: Single point	Changes within a sentence or clause: a single instance
				Intrasentential: Multi-point	Changes within a sentence or clause: multiple instances
				Intersentential	Changes over sentence or clausal boundaries
				Tag	Parenthetical comments or formulaic phrases

[42] Turner's (2015) tag JMNV correlates with _JMPR as well as _JNC as one use of JMNV by Turner involves the Japanese modification or pronunciation of English names in the source text being replicated in fanfiction.

Appendix 2: Japanese token tag set with notes and examples

Tag	Shorthand	Note	Example	English translation
	Orthography tags			
_K	Japanese orthography	Text in kanji or kana (Japanese writing systems) as opposed to romaji (roman letters)	LOL "bilingual anon," it says キュート**_K** not キューン。**_K** Are you not "da real nihonjin_^_JPN???"_DM	*LOL "bilingual anon," it says **cute** not **kyuun** Are you not "da real **Japanese person**???"*
_JK	Japanese kaomoji	Japanese style emoticons (faces created with punctuation symbols)	Someday... **;_;_^_JK**	*Someday **[crying face]***
_^	Romaji	Tokens written in romaji, which is the Romanization of the Japanese orthography (into Roman letters) generally for text directed at foreigners or for computer input	Bad enough that this is the usual **'kawaii_^_**JPAD_JL **maido_^**_JPN_JMPR_JF' outfit I see at the con.	*Bad enough that this is the usual '**cute maid**' outfit I see at the con.*
	Part of Speech tags			
_JPHT	Japanese part of speech: honorific title	Japanese honorific used as an unattached title, or kinship terms	I failed **senpai**_^**_JPHT**	**senpai:** honorific suffix or title used to show respect to those who are the speaker's senior (in age, experience, or position) at school or in the workplace
_JPHS	Japanese part of speech: honorific suffix	Japanese honorific attached as a suffix	They could have said/ done something about **midlifecrisis-chan**_^**_JPHS**	**-chan:** honorific suffix generally used as an affectionate diminutive
_JPV	Japanese part of speech: verb	Words that are classified as verbs or adverbs in Japanese, as well as words used to function as verbs in English	…approached me and asked, "Ninhonjin_^_JPN **desu**_^**_JPV** ka_^_JPPA?"_JF	*…approached me and asked, "**are** you Japanese?"* **desu:** copula verb often translated as 'to be'
_JPNO	Japanese part of speech: noun	As classified in English	Im trying to find sewing patterns for **seifuku**_^**_JPN**_JMPL	*Im trying to find sewing patterns for **sailor uniforms***
_JPNA	Japanese part of speech: attributive noun	Nouns used to modify another noun, as classified in English	I love Meta's **kimono_^_JPNA**_JL prints	
_JPPA	Japanese part of speech: particle	As it would be classified in Japanese	…I have the Otome_^_JNP_JC **no_^_JPPA** sewing_JNC book…	*…I have the **sewing book of Otome [style]**…* **no:** particle to turn a noun into a modifier
_JPIE	Japanese part of speech: interjection or expression	Interjections and other formulaic expressions, including greetings and swear words	Suck it up and do it. **GANBARE!!**_^**_JPIE**	Suck it up and do it. **Good luck/You can do it!!**
_JPSS	Japanese part of speech: sound symbolism	Japanese ideophones describing sounds and other sensory perceptions	I totally Read that as **kyuun**_^**_JPSS**.	**kyuun:** symbolic of the clenching of one's heart when confronted with something cute, or falling in love

Tag	Shorthand	Note	Example	English translation
_JPPR	Japanese part of speech: pronoun	Words that are classified as pronouns in Japanese	or at the very least omitting the **watashi _JPPR**	**watashi:** first person pronoun (note: the context of the usage is discussion of an in progress art design with Japanese text on it)
_JPAD	Japanese part of speech: adjective	As classified in English	y'all are **ita_^_JPAD**_JC as fuuckk	**ita:** slang term used in Japanese Lolita fashion to mean an outfit or person that is 'painful to look at' (from the Japanese adjective/interjection '**itai**' meaning 'painful' or 'ouch')
_JNS	Japanese name: series[43]	Tokens that are existing Japanese series names	I miss the days of good **Dangan Ronpa_^_JNS** cosplay_^_JPN_JMSH_JL.	**Dangan Ronpa:** A Japanese video game franchise
_JNC	Japanese name: character, person, or place name	Tokens that are the Japanese names of characters, real people, or place names	I cosplayed_^_JPV_JMSH_JL_EM **Mikasa_^_JNC** and **Makoto_^_JNC** from Free!	Characters from the anime series Free! Eternal Summer
	Feature tags			
_JNE	Japanese name: Anglophone based company, event, or username	Names using Japanese that are created outside of Japan by Anglophone based companies, etc.	(and **Con-nichiwa**_^_JPN**_JNE** and **Kikori_^_JPN**_JNE)	**Con-nichiwa** and **Kikori [Con]** are two anime conventions in Arizona, USA
_JL	Japanese loan word	Words from Japanese that are classified as loanwords in English per their presence in the Oxford English Dictionary	I wouldn't buy brand for a **cosplay**_^_JPNA_JMSH**_JL** anon	**cosplay:** Japanese style contraction of the English words 'costume' and 'play', referring to dressing up as a character from pop culture, originally coined by Nobuyuki Takahashi as コスプレ
_JF	Japanese phrase	A Japanese phrase segment, longer than one word	>**kawaii**_^_JPAD_JL **desuuuu**_^_JPV, **sugoi**_^_JPIE!!!**_JF** _DS	***[subject] is cute, awesome!!!***
_JAC	Japanese anime catchphrase	Phrases that are specific to characters in anime and manga	She occasionally will use gratuitous Japanese with other people when she is quoting a character (such as quoting Nozomu_^_JNC and saying "**zetsubou shita!_JAC**").	*She occasionally will use gratuitous Japanese with other people when she is quoting a character (such as quoting Nozomu and saying "**I'm in despair!**").*
_JMSH	Japanese modification: shortening	Japanese pattern of abbreviation and contraction where the first two morae of each word are combined for a shortened form	**TouRabu**_^_JNS**_JMSH** is layers central, but some of their outfits are summer-appropriate.	Shortening of the name of a Japanese game, **Touken Ranbu**

[43] JNC and JNS are separate as there are often English translation versions of series names that are used interchangeably with the original Japanese name

Tag	Shorthand	Note	Example	English translation
_JMPR	Japanese modification: pronunciation	Modification of written English words to mimic Japanese pronunciation conventions	The guy in the middle back RUINED my **husbando**_^_**JPN**_JMPR	**husbando:** partially modified version of English 'husband', meaning a character the speaker has a particular love for and thus imagines them as a fictional spouse
_JMPL	Japanese modification: plural	Lack of plural as per Japanese grammatical convention	Used/slightly flawed **kimono**_^_JPN_JL**_JMPL** are so cheap on eBay and other second-hand markets…	**kimono:** traditional Japanese garment
_EM	English modification	Modification/nativisation of words per English grammatical conventions, such as pluralisation and verbal suffixes.	I was thinking about **cosplaying**_^_JPV_JMSH**_EM**_JL for the release of the new movie…	
	Attitudes tags			
_DS	Discussion: stereotype	Discussions or mentions in text establishing the presence of a negative stereotype within the community/affinity space	People are very different from what they're like at a con. Except **weeaboos _DS**, that shit is 25/8/366	**weeaboo:** a non-Japanese person who is overly obsessed with Japanese culture (anime in particular) in what is considered an obnoxious and 'uneducated' way
_DM	Discussion: metalinguistic	Discussions in text about language use for qualitative analysis	y'all know the plural of "kimono_^_JNP_JL" is just "kimono_^_JNP_JL_JMPL", right? the 's' isn't necessary...**_DM**	Discussion about the way to pluralise a Japanese word

Appendix 3: All JNS, JNC, and JNE words found in the CGL corpus

Rika	Chouji	Mikisugi	Izaya	nana	urogan
Kaminas	Rei	Asuka	Motoko Kusangai	Aeka	Abe
Maki	TSUME	jotaro	tamamo	Yoko	Nagase
Suitengu	Kyon	Haruka	Miyuki	Marin	Peko
Rei Kurosawa	Rakuten	Junko	touko	Doremo	Monokuma
Mikan	Yamada	Koizumi	Syo	Togami	Nanami
Kuzuryuu	Mukuro	Souda	Gundam	maizono	Usui
mondo	Komaru	Masaru	koujaku	Shiroba	Okabe
aoba	Amane Nishiki	Miku	Nekopara	Gamagori	Monoko
Gakupo	Fujitora	Jiraiya	Yamamato	ayanamisatoru	Kira
Ryuko	Oikawa	Anzu	ichiroya	Komoni	Nui
Miki	Tadashi	Reimi	Yukiho	Nonon	satsuki
Hiro	Naruto	Kairi	Hinata	junketsu	Taokaka
Yamanbagiri	Satsuki	Zashiki Warashi	Ame Warashi	Tsubasa	senketsu
Yuuko	Sakura	Tokyo	Taigakun	Goku	Watamote
balloonketsu	ae-tan	Tenso	tao	harajuku	Oni
Akira	Rilakkuma	Korilakkuma	vanilla chan	AKIRA	Akemi Homura
%dokidoki	Watashi Sendou	Kageri	hNaoto	Kirakishou	Fuu
Masumi Kano	Misako	Daiso	mononoke	Hikaru	Konomi
shinjuku	Hakuoki	tenshi	Hibiki Kuze	Kanji	kamen
Kojima	Yukako	Ichigo	Kira	Josuke	yasu
Hei	Saitama	Ginko	Araki	hachikuji	pikachu
sayo	Araragi	Kyuubey	kobito	Chibiusa	Mami
Shinichiro Watanabe	Rin	Kachidoki	Yasogami	Ryuki	Suzuya
Raidou	Kongou	Kakashi	Tsukishima Kei	Gasai Yuno	Juuzou
Hitomi	Nozaki	ryu	Sakura Chiyo	tezuka	Milkychan
aoba seragaki	Misa	Nagisa Hazuki	kagami	Tanaka	Nabeshin
Suiseiseki	Usakumya	Putumayo	Nakamura Risa	mikasa	hare hare yukai
Makoto	Uta	Kyashii	Nanako	Baofu	Akihiko
Raiho	ryuko matoi	kamui senketsu	Haruhi	haruharukari	Tenchi Muyo
Yuki	subaru	yoshi	Tsubaki	haiku	Ookami-San
UtaPri	Kuroshitsuji	pokemon	Haganai	Magi	fukawa
Inazuma Eleven	Saint Seiya	Yu-Gi-Oh	Seiekn Densetsu	Dangan Ronpa	touhou
Yume Nikki	TouRabu	miyuki	gouhou	maoyuu	Madoka
uzumaki	toku	KanColle	Naruto	Totoro	SakuraCon
boku no pico	Obake no Q-taro	Haikyuu	Kamikaze Girls	Fanime	taigakun
Usagimon	katsucon	otakon	banchomako	youma	hentarmageddon

teko	mahou	Mugi-chan	mitsuwa	kinokuniya	Taiyou-con
Ryukishi	Animazement	Anime North	momocon	Oni-con	ichibancon
saboten	Con-nichiwa	kikori	takoyakicon	otakuthon	Kira kira con
kumoricon	hoshicon				

Appendix 4[44]: Lexicon of Japanese romaji words[45] found in the /CGL/ corpus

cosplay/cosplays/cosplayers/cosplayed/cossupurai/cossuprai	*anime/animu/animes*	*manga/mango*
kawaii/kawaiiest	*hentai*	*washi*
otaku	*sushi*	*kimono/ kimonos*
gi	*hakama*	*kabuki*
yukata/yukatas	*gaijin/gaijins*	*geisha*
yen	*bento*	*katakana*
hiragana	*kanji*	*nigiri*
ramen	*seppuku*	*sudoku*
katana	*haori*	*futon*
geta	*obi*	shimapan
waifu [wife]	wa-lolita	mori /mori kei
tenshi	seifuku/ fuku	himekaji
bara	burando/brando [brand]	sukusukus
yokai	kigurumi/ kigu	sukumizu
gohan	chibi	teitoku
husbando [husband]	nanchatte	nominoichi/ nomi
yaoi	gunpla	shibari
ikemen	yuri	dakimakura/ daki/ dakis
onigiri	ouji	shinigami
teko	gyaru/ gyarus/gyaru-o	kodona
omurice	ganguro	otome/ otome kei
kotatsu/kotatsus	miko	figma
neko/nekos	konmari	uke
shinai	-chan	kanzashi
kei	dansou/dansous	henjin
yamanba/manba	fukubukuro	kuro
bishoujo	parapara	zentai
kogal	bishounen/bishies	hime
furisode	guro [grotesque]	kissu [kiss]
meido/maido [maid]	gijinka	aidoru [idol]
shounen	akamoji-kei	koneko-chan
desu	purikura	ita/ itas/ ita-tier/ita-ing/not-ita
baka	ganbare	kyuu/kyun/kyuun
-kun	moe/moe moe/moeshit	sage/saged/autosage/autosaged
onii-chan	kowai	-senpai/ senpai
watashi	-tan	nee-chan
mahou	kamisama	nya/nya nya nya
shota	rorita/roritas/rori	dramu/duramu/dramus [drama]
kokoro	himitsu	friend [friend]
okubo	nekomimi	megane
ne	moshi-moshi	shiro
yo	sugoi	kemonomimi
beito	ka	

[44] Note: Loanwords italicised

[45] Except for words tagged JNC, JNS, JNE (see Appendix 7 for those words)

Appendix 5: Comparison of frequency of part of speech tags in the CGL and Troper Tales corpora

	CGL		Troper Tales	
Tag	**Raw Frequency**	**Percentage of Total Romaji Tokens**	**Raw Frequency**	**Percentage of Total Romaji Tokens**
JPHT	6	0.18	12	2.41
JPHS	102	3.00	27	5.43
JPV	330	9.71	33	7.16
JPN	1890	55.60	105	21.13
JPNA	446	13.12	13	2.62
JPPA	8	0.24	61	12.27
JPIE	2	0.06	126	25.35
JPSS	15	0.44	14	2.82
JPPR	2	0.06	31	6.24
JPAD	162	4.77	44	8.85
JNS	76	2.24	9	1.81
JNC	360	10.59	18	3.62
Total Romaji Tokens	**3399**	**100**	**497**	**100**
Corpus Word Count	**403972**	**0.84**[46]	**11251**	**4.42**[47]

[46] Percentage of total *Romaji* tokens in corpus

[47] Percentage of total *Romaji* tokens in corpus

Appendix 6: The Top 20 Most Frequently Appearing Japanese Words in the CGL Corpus

Ranking	Word[48]	Translation	Frequency	Percentage[49]
1	***cosplay/cossupurai****	Lit. costume play	1597	69.50
2	***anime***	animation	265	11.53
3	***kawaii***	cute	74	3.22
4	***kimono***	Traditional item of clothing	50	2.18
5	ita	From 'itai' meaning 'ouch': something that is painful to look at	50	2.18
6	otome	Maiden (fashion style)	35	1.52
7	***manga***	comics	31	1.35
8	gyaru*	Lit. gal (fashion style	30	1.31
9	***otaku***	Person obsessed with popular culture	24	1.04
10	yaoi	Media genre featuring male homosexual relationships	21	0.91
11	waifu*	Lit. wife	20	0.87
12	ouji	Lit. prince (fashion style)	17	0.74
13	burando*	Lit. brand	19	0.83
14	neko	cat	11	0.48
15	husbando*	Lit. husband	10	0.44
16	fuku/seifuku	Sailor style school uniform	10	0.44
17	chibi	Cute 'deformed' art style	10	0.44
18	***sushi***		10	0.44
19	daki/dakimakura	Body pillow with animated character pictured on case	7	0.30
20	meido/maido*	Lit. maid (referring to clothing style)	7	0.30

[48] Words in bold italics are loanwords according to the Oxford English Dictionary, words with * are *wasei eigo* words

[49] Percentage of total tokens

Appendix 7[50]: All Japanese Romaji multiple word segments found in the CGL corpus

kawaii desuuuu, sugoi!!!	moe moe kyuu	Ninhonjin desu ka?
moshi-moshi beito desu	*kawaii* desu ne	*kawaii* desu yo neeee
kawaii meido		

[50] Note: Loanwords italicised

Appendix 8: Main attitudes towards language use and the people who use it found in the Troper Tales text

General trends	Example in text
Negative attitudes	
Reference to a young or past age	This troper suffered this fresh off his trip to Japan when he was still a [[OldShame]] **failtacular teenager** who thought he was cool randomly injecting Japanese swears into conversations. Now he leaves poorly pronounced Japanese to his Otaku friends, who now look as silly as he once did. **They'll grow out of it**…
Reference to the person or linguistic behaviour as 'annoying' or 'unacceptable'	This troper knew a whole group of girls that did this. At first it was, like, cool, you guys like anime, too but it quickly got **annoying**, especially since all of them though it was so clever that they could do this.
Reference to degree of fluency in Japanese	If you speak a word of Japanese in everyday conversation (unless completely appropriate in context) [[BerserkButton]] **you'd better be fluent**. (**Note that I'm not fluent**, but I don't insert random words in sentences ever, either.)
Reference to the terms 'weeaboo' or 'otaku'	This troper has the tendency to shout "YATTA!" when something good happens to her, and is also prone to saying "itadakimasu" before she eats, but other that those two phrases, that is the extent of her use of Gratuitous Japanese, as she does not want to be seen as a **weaboo**.
Correct pronunciation is considered important	This troper used to speak like this. Now she knows a group of girls who do this all the time, and it is one of the most annoying things she's ever heard. The super-squeaky voices, the gratuitous honorifics, **the terrible pronunciation**…She is sincerely sorry she ever acted that way.
User feels need to 'defend' use of language	Top troper also **defends his use** of "Yare yare daze" as being situationally appropriate, muttered under his breath most of the time, and really fun to say.
Positive or neutral attitudes	
Japanese knowledge is high status	This troper has a decent knowledge of Japanese… • Sorry, but this is NOT 'decent knowledge'…So shut up and stop bragging…
Showing off unique or interesting language behaviours	Though, for some strange reason, his speech style seems to be…stoic-ish.
Acceptable if no direct translation	Tsundere is actually beginning to enter this troper's vocabulary. Since stealing from other languages is pretty much the entire basis of English, I figure it's fair. • There really **isn't an English equivalent** for the word anyway. • Yeah**, it's not quite so gratuitous in that particular case**
Acceptable in certain contexts	I only use sempai/kohai **with other J-leaners/Japanese people**
Examples of own language use given without justification or disclaimer	This Troper finds himself using [[Gintama]] "Zura ja nai, Katsurada!" as a swear, does this also count as a [[Gosh Darn It to Heck]]?