

# Linguistics D conference

2023.11.22

November 21, 2023

1. Analysis on differences between Vietnamese and Japanese languages: Lam Vi Toan .....	2
2. Pleasure reading in a foreign language and its effect on communication ability: Jensen Nicholas John ..	4
3. Qualitative Analysis of Alphabetical Change In Indonesian Sanskrit Loanwords: Firdaus Zuhail .....	6
4. The role of motivation in language acquisition : Miyagi Shota .....	8
5. Japanese Loanwords: Garry Pranata Kusuma .....	10
6. A Comparative Study of the Distribution of reduplicative in Classical Chinese by examples: ZHANG RUI 12	
7. A Brief Research on Japanese Loanwords in Chinese: WU SHENGYE .....	14
8. WORDS FREQUENCY IN MY RESEARCH SUMMARY: Hu Yunfan .....	16
9. Differences and similarities between Chinese and Japanese onomatopoeias: Long Yanke .....	18
10. The Rapid Katakanaification of the Japanese Language: Van Alstine Nathan .....	20
11. Mood Detection in Messages using Textual Signals: VARDHAN Chetan .....	22
12. The Effects of Syllable Complexity on Verbal Information Transmission Speed: Lindahl Jacob .....	24
13. English Vowels at ELSI A Comparative Study of English L1 and Japanese L1 Speakers: Taylor Austin	26

# Analysis on differences between Vietnamese & Japanese languages

Vi Toan LAM

Life Science & Technology, Tokyo Institute of Technology

## 1. Introduction

- Vietnam & Japan are two countries with long history of linguistic development.
- Both are Sino languages [1].
- Chinese words (kanji) were introduced to Japan during 5<sup>th</sup> century AD through trade [2] and cultural exchange.
- In Vietnam, Han characters (Hán tự) were introduced in 3<sup>rd</sup> century BC after ancient empire of Vietnam (Âu Lạc) had been dominated by Zhao Tuo, a former Qin official.

## 2. Methods

- Analysis on Vietnamese and Japanese based on history of languages.
- Comparison of grammars between two languages.

## Why the two languages differ from each other

## 3. Results & discussion

Table 1. Difference in syllabic systems

Examples	Modern Vietnamese	Sino - Vietnamese	Japanese
	Monosyllabic		Polysyllabic
I meet the teacher	Tôi gặp thầy	Ngã ngộ kiến lão sư 我遇见老师	僕は先生に会います

Sino – Vietnamese was formed due to the domination of ancient China, it inherit not just the monosyllabism, but also the language system from old Chinese. Whereas kanji was adopted to Japan just through trading with Chinese immigrants from Korea [2].

Table 2. Difference in grammar systems

Examples	Modern Vietnamese	Sino - Vietnamese	Japanese
	Adjectives stay after nouns	Adjectives stay before nouns	
Red shoes	Giày đỏ (shoes + red)	Hồng hài 红鞋 (red + shoes)	赤い靴 (red + shoes)

Modern Vietnamese writing system was formed by missionaries from Portugal during 17<sup>th</sup> century [3]. Therefore, modern Vietnamese inherited not just the Latin alphabet, but also the grammars of Portuguese.

Table 3. Difference in temporal expressing systems

Examples	Modern Vietnamese	Sino - Vietnamese	Japanese
	Add extra indicators beside the verbs to differ the tenses		Add suffix to verbs or change the pronunciation to differ the tenses
Drink Drank Drinking	Uống Đã uống (past indicator) Đang uống (continuous)	Ẩm 飲 Ẩm liễu 飲了 Chính tại ẩm 正在飲	飲む 飲んだ 飲んでいる

Due to the monosyllabic tendency in Vietnamese and Chinese, both languages use extra indicators to distinguish the temporal expressions. Whereas in Japanese, writing in kanji might be still counted as extra indicator, but the pronunciation is changed sometimes.

## 4. Conclusion

- Comparison of grammars between Vietnamese & Japanese languages as well as analysis on historical aspects were carried out to shed light on the story behind the differences of two languages.
- Vietnamese, which inherited from both ancient Chinese and modern Latin languages, shows its non-identical properties in comparison with Japanese in writing, phonetic and grammar systems.

## References

[1] R.J. La Polla (2006), "Sino-Tibetan Languages", Encyclopedia of Language & Linguistics, 2<sup>nd</sup> edition, pp. 393-396.  
 [2] Mathieu (2017), "The History of Kanji ", It's Japan Time.  
 [3] Jacques, Roland (2004), "Les missionnaires portugais et les débuts de l'Eglise catholique au Viêt-nam", vol. 1, Reichstett, France.

# Pleasure reading in a foreign language and its effect on communication ability

Nicholas J. Jensen

Systems and Control Engineering, Tokyo Institute of Technology

## 1 Introduction

A common question in the context of acquiring a foreign language is “*what is the best method of learning a foreign language that leads proficient communication that language?*” There are many ideas in the linguistics literature and even more when it comes to personal situations. One theory by Stephen D. Krashen claims that the best way to acquire proficiency in a foreign language is to utilize *The Power of Reading* [2]. This idea is captured in Krashen’s 2004 book, which describes a specific kind of reading. Krashen says, “free voluntary reading (FVR) means reading because you want to: no book reports, no questions at the end of the chapter” [2].

The concept of FVR and its affect on communication in a foreign language is the focus of this research. Specifically, the research question in mind is “*does reading for fun in a foreign language lead to better communication than not reading for fun?*” This poster includes the research methods, followed by a brief presentation and discussion of the results before concluding.

## 2 Methods

The approach to address this research question was similar to the work done in [1]. Both studies are replications of Krashen’s theory, but are still valuable. A comment made by Cho deserves repetition that says, “Studies of journal policy have shown that replication is not highly respected in academia. Replication, however, is extremely important for both theory and practice: each replication is another test of a hypothesis, can increase or reduce confidence in its validity and can encourage or discourage applications of the hypothesis” [1]. In light of this comment, a survey was conducted in a different setting to continue testing Krashen’s theory.

Similar to [1], the survey was based on “self-reporting” to record the frequency of reading in a foreign language and the participants perceived ability to communicate in that language<sup>1</sup>. The survey was given to Tokyo Tech students and other people living in Tokyo. The questions asked the participant to rate the frequency they read in a language on a scale between one and five<sup>2</sup> and their ability to communicate on a scale between one and five<sup>3</sup>.

## 3 Results

The survey collected results from 93 participants. Of those participants, 15 stated that they can’t speak or write in a foreign language while 78 could speak at least one foreign language. From the 78 participants that could communicate in a foreign language there were 152 language instances<sup>4</sup>. From the 152 language instances 73 reported reading for fun while 79 did not include reading for fun. A visual of reading for fun and not reading for fun is given in Figure 1. The survey results were processed further to compute the mean assessment levels reported by the participants. The results are shown in Table 1.

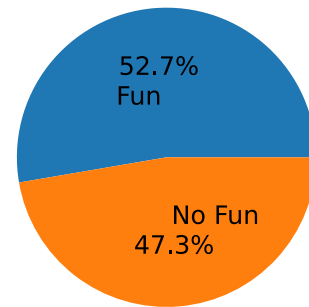


Figure 1: Language instances that included reading for fun and did not include reading for fun

Table 1: Survey results of reading in a foreign language. Mean  $\mu$  and standard deviation  $\sigma^2$  are provided

	Fun	No Fun
Reading freq. $\mu$ ( $\sigma^2$ )	3.38 (1.52)	2.65 (1.52)
Comm. ability $\mu$ ( $\sigma^2$ )	3.18 (1.33)	2.67 (1.28)
Correlation	0.68	0.61

## 4 Discussion

A few notes regarding Table 1 are: (1) the spread of each category is similar, meaning the distribution shapes are similar (2) reading for fun had a higher mean reading frequency and communication ability than reading not for fun (3) fun reading frequency is higher than fun communication ability and opposite for reading with no fun (4) the correlation between reading for fun and communication ability is higher than when not reading for fun and its correlation with communication ability. Based on (2-4) it can be said that reading for fun leads to more frequent reading and better communication in a foreign language. These results seem to affirm Krashen’s theory of *The Power of Reading* with FVR [2].

## 5 Conclusion

A survey was conducted to assess foreign language pleasure reading frequency and its effect on foreign language communication ability. The survey was “self-reported” and showed that reading for fun has higher correlation with communication ability than reading without fun. These results replicate Stephen Krashen’s work on the *The Power of Reading* and provide an answer to the question of *what is the best way to learn a foreign language?*

## References

- [1] Kyung-Sook Cho and Stephen Krashen. Pleasure reading in a foreign language and competence in speaking, listening reading, and writing. *TEFLIN Journal - A publication on the teaching and learning of English*, 30(2):231, aug 2019.
- [2] Stephen D. Krashen. *The Power of Reading: Insights from Research*. Libraries Unlimited, 2004.

<sup>1</sup>A link to the survey is here: <https://forms.gle/N17ruSuUeYXoHrPr7>

<sup>2</sup>One rating description was “Occasionally”. Five rating description was “Every day all day”

<sup>3</sup>One rating description was “I can only say or write a few things”. Five rating description was “I am fluent”

<sup>4</sup>A language instance refers to each additional language spoken by each participant. If one participant speaks two languages in addition to their native language then this is two language instances

# Qualitative Analysis of Alphabetical Change In Indonesian Sanskrit Loanwords

Zuhal Maulana Firdaus

Electrical and Electronic Engineering, Tokyo Institute of Technology

## 1. Introduction

Indonesia is a country in the South East Asia region, as a country rich in resources Indonesia is frequently visited by foreign merchant since pre-colonial time. One of the earliest contact with foreign civilization is with the Indian culture which dates back to ancient times[1]

Due to the nature of trade and contact with foreign merchant, many loanwords are diffused in the Indonesian language, one of it being the Sanskrit language from India in which this study aims to answer questions regarding:

- How does the alphabet changes for loanwords from the Sanskrit language to Indonesian language
- Does the definition of those word shifted to something other than the original definition in Sanskrit language?

## 2. Methods

This study uses a qualitative approach in which a research is focused on qualitative data containing word collected through interviews, documents, or observation[2].

The source of data used in this study is collected from the book "Unsur-unsur Bahasa Sansekerta Dalam Bahasa Indonesia" (Elements of Sanskrit Language in Indonesian Language)[3]

Words are chosen in the manner that there is an alphabetical change of the Sanskrit language in Indonesian language while the original Sanskrit meaning is the same with the Indonesian meaning.

Analysis is conducted by observing the alphabetical change and finding the pattern of the change

## 3. Result

The data is selected based on [3], six Sanskrit words are selected with a considerable alphabetical change in it's spelling but same or similar meaning, as displayed in Table 1.

Table 1. Indonesian Loanwords from Sanskrit Language With It's Original Meaning

No	Sanskrit Word	Meaning	Indonesian Word	Meaning
1	Artha	Wealth	Harta	Wealth
2	Cavi	Chili	Cabai	Chili
3	Naraka	Hell	Neraka	Hell
4	Vāhana	Vehicle	Wahana	Vehicle
5	Jīva	Soul	Jiwa	Soul
6	āsthāna	Castle	Istana	Castle

## 4. Discussion

Based on the selected corpora, it is evident that alphabetical change occurs on the Sanskrit word in it's conversion into the Indonesian loanwords. The Sanskrit word "Artha" is converted into "Harta", adding the "H" letter in its pronunciation. The word "Cavi" is localized into "Cabai". "Naraka"'s first vowel is converted from "a" to "e" while maintaining the rest of the vowel and consonant the same. The word "Vahana" is localized into "Wahana" and the word "Jiva" is localized into "Jiwa". The last word is "Asthana" which is loaned into "Istana".

From all the words in Table. 1 there are a few patterns that is discernable, some of Sanskrit word that starts with a vowel in the initial position has its localization changes as initial vowel change from "a" to "e" as depicted in the 6<sup>th</sup> word in Table 1 while the other pattern is that of the 1<sup>st</sup> word, "Artha" is shifted to "Harta" in which a consonant "H" is added in initial position. The second pattern is that vowel in the medial position is also subject to change as depicted in the 3<sup>rd</sup> word of Table 1, the vowel "a" is changed to "e" in medial position. One of the most common pattern in Sanskrit-Indonesian localization pattern is in the word that involves the letter "v" in the Sanskrit word. Words that has the "v" letter tends to be localized into the "w" letter which can be inferred from 4<sup>th</sup> and 5<sup>th</sup> word on Table 1, this change happened regardless of position with a few exception like that of the 2<sup>nd</sup> word.

## 5. Conclusion

Based on Table 1 and discussion, it can be understood that alphabetical change occurs in the initial, medial, and final position of words which implies that before Sanskrit language came in contact with Indonesian civilization the way that native Indonesian speaks before that is already defined in which Indonesian language is already existing in some form before modern Indonesian language exist. The way that the original meaning between Sanskrit and Indonesian Sanskrit loanword is the same implies that there is no strong enough reason or influence in which Indonesian people need to differentiate the meaning of that word.

Reference:

[1] Trivedi, S. (2010). Early Indian Influence in Southeast Asia: Revitalizing Partnership between India and Indonesia. *India Quarterly*, 66(1), 51-67. <https://doi.org/10.1177/097492841006600104>

[2] Miles, M. B., & A. M. Huberman. (1994). *Qualitative Data Analysis: An Expanded Source Book* (1st ed.). London: SAGE Publications.

[3] Sharma, M. M. (1985). *Unsur-unsur Bahasa Sansekerta Dalam Bahasa Indonesia*. Denpasar: Wyasa Sangraha

# The role of the motivation in language acquisition

Shota Miyagi

Social and Human sciences, Tokyo Institute of Technology

## 1. Introduction

The motivation is a crucial factor in language acquisition. There have been numerous studies dealing with a group of language learners to see if there is any difference in the degree of language acquisition depending on their motivations. This study, on the other hand, focuses on the language learning history of a certain person who has acquired several languages.

## 2. Methods

To investigate the research question, I interviewed a person who has acquired several languages at different life stages. The interviewee is a Hungarian female person, who grew up in a Hungarian community (Sfantu Gheorghe) in Romania and later on went to a university in Hungary where she majored in Japanese language. She speaks Hungarian as first language, Romanian as second language, English as third language, and Japanese as fourth language.

## 3. Result

The outcomes are as follows: She feels most comfortable speaking in Hungarian, and then English, Japanese followed by Romanian in the end. Although she grew up in Romania, she was least motivated to speak Romanian. On the other hand, She is gradually feeling more and more comfortable in Japanese language since she lives in Japan and speaks Japanese everyday.

Language	The learning period	Motivation
Hungarian	30 years	Native language
Romanian	24 years	Education
English	23 years	Education
Japanese	11 years	Japanese culture

Table 1: Motivation and learning period

## 4. Discussion

In the field of linguistics, the following category is widely known for motivation: Integrative orientation and instrumental orientation. Integrative orientation is that the language learners seek to integrate themselves in the community where the target language is spoken. Instrumental orientation is that they are motivated to learn the target language because of their career. In her case, the integrative orientation can be seen with the Japanese language while the instrumental orientation can be seen with Romanian and English. According to the result, the instrumental orientation might be only effective for a certain language such as English.

## 5. Conclusion

In this research, I examined how influential the motivation can be for language acquisition using the interview method. For the target person, the instrumental orientation was seen only with English, while the integrative orientation was strong for Japanese.

<sup>5</sup> Reference: 守谷 (2002) 第二言語教育における動機づけの研究動向 - 第二言語としての日本語の動機づけ研究を焦点として, 言語文化と日本語教育

# Japanese Loanwords

Garry Pranata Kusuma

Electrical and Electronic Engineering, Tokyo Institute of Technology

## 1. Introduction

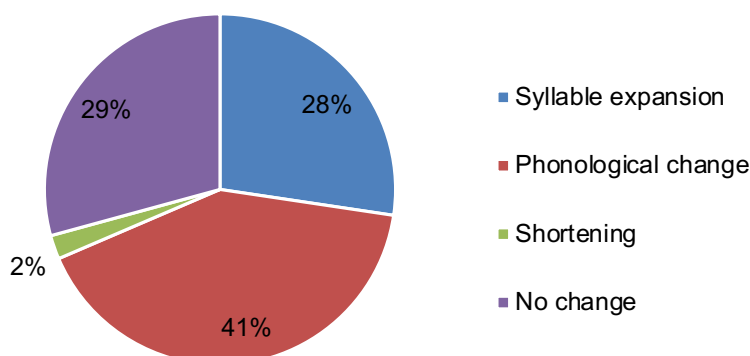
This study investigates how Japanese assimilates borrowed terms, or 'loanwords'—words adopted from one language and incorporated into another with varying degree of modification. The study aims to highlight the adjustments Japanese employs to accommodate these foreign sounds within its own phonetic constraints as well as the origin of each word.

## 2. Methods

A set of loanwords will be gathered from a recognized database. They will be categorized by focusing on the change each word undergoes. We will then count the the frequency to reveal how they are nativized. The origin of each word will also be analyzed and tallied up.

## 3. Result

Loanwords can experience: syllable expansion (e.g., ピーナツ), shortening (e.g., ビル), phonological changes (e.g., ペンギン), or little to no change (e.g., トマト). The following data is compiled from 318 Japanese loanwords (excluding names and countries).



6

Fig. 1. Changes experienced by selected loanwords

## 4. Discussion

According to Fig. 1, most LWs in the dataset undergo phonological change. This is because in Japanese, there are only 5 vowels whereas in English, there are 12 or more ways. This means that many English words will sound differently when written in katakana. Due to katakana's monosyllabic nature, some loanwords from English which have consonant clusters end up being expanded. Shortening in LWs probably happens to save time when speaking. The words in "no change" category in Fig. 1 are not ones which do not experience any change at all. It is referring to words which are phonetically identical to the original words and do not undergo any expansion. However, this is highly subjective since different accents might have different opinion on whether a LW undergoes "no change" or not.

Table 1 reveals that most of Japanese LWs in the dataset are derived from English. After WWII, Japan aspired to creating a country similar in strength to the US. Hence, perhaps Japanese began to use more English words in their own language unconsciously. Having English as an international language will also increase the number of English words adopted into Japanese LWs.

## 5. Conclusion

Japanese loanwords can undergo syllable expansion, shortening, phonological change, or minimal change. They come from various origins, the most common of which is English.

Reference: Olah, Ben (2007) English Loanwords in Japanese: Effects, Attitudes and Usage as a Means of Improving Spoken English Ability, 文京学院大学人間学部研究紀要, Vol. 9, No. 1, pp.178~180.

Loanwords Source:

[https://www.bunka.go.jp/kokugo\\_nihongo/sisaku/joho/joho/kijun/naikaku/gairai/index.html](https://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/kijun/naikaku/gairai/index.html)

Table 1. The origins of selected loanwords

Origin	Number of loanwords
English	299
French	6
German	3
Italian	1
Korean	1
Dutch	2
Danish	1
Latin	2
Portuguese	1
Chinese	1
Greek	1

# A Comparative Study of the Distribution of reduplicative in Classical Chinese by examples

ZHANG RUI

EE, School of Engineering, Tokyo Institute of Technology

## 1. Introduction

Reduplicative is a word in the form of “字字”, e.g. 孜孜, 往往, 悻悻, 翩翩... that is widely used from classical to modern Chinese. This study will reveal the frequencies of reduplicative by examples, and talk about their differences and causes. I have chosen the 詩經, 楚辭, 全唐詩, 史記, 漢書 and 舊唐書 as the sample set. They are five ancient Chinese classics with different characteristics

Table 1: the characteristic of 5 books

	Types	Compiled time
詩經	verse	1100 ~600 BC
楚辭	verse	300 BC
全唐詩	verse	618~907 AD
史記	narrative	91 BC
漢書	Narrative	82 AD
舊唐書	narrative	941 AD

## 2. Methods

Write a program to find all repeated character words, excluding exceptions, and count their occurrences.

You could access my code through:  
<https://github.com/rinmyo/lingui>

## 3. Result

Table 2: the frequencies result of reduplicative

	Total characters	Number of Reduplicative	Frequencies
詩經	28425	704	2.48%
楚辭	26229	377	1.44%
全唐詩	3297733	20332	0.617%
史記	500702	629	0.126%
漢書	1297788	4226	0.326%
舊唐書	1900071	2344	0.123%

## 4. Discussion

All the results show that verse, such as poetry, tend to use reduplicative words more frequently than non-rhymed texts, like prose. This preference may be linked to the structural and aesthetic characteristics of rhymed texts. Verses emphasize rhythm and harmony in sound, and reduplicative words contribute uniquely to these aspects.

In classical Chinese poetry, this use of reduplicative words is a common and significant artistic technique. The main purpose of non-rhymed texts is to record, and therefore they usually do not pay special attention to pronunciation. The main purpose of non-rhymed texts is documentation, hence they usually do not give special consideration to sound and rhyme.

On the other hand, we find that pre-Qin(200 BC) rhymed texts, compared to the 全唐詩, show a higher frequency. This may because, during the pre-Qin era, Chinese had not yet experienced the phenomenon of differentiation between literary and colloquial languages. Therefore, the influence of oral habits, which varied more rapidly with region and time, was more pronounced on rhymed texts. But after the Qin Dynasty, literary and colloquial Chinese became distinct, and the linguistic system of classical Chinese was fixed, no longer influenced by changes in spoken language.

## 5. Conclusion

This study analyzes the frequency of reduplicative words in the Book of Songs, Chu Ci, Complete Tang Poems, Records of the Grand Historian, Book of Han, and Old Book of Tang, and explores the possible causes of the differences among them.

Reference:  
 漢リポ Kanseki Repository:  
<https://www.kanripo.org>



# A Brief Research on Japanese Loanwords in Chinese

WU Shengye

Mathematical and Computing Science, Tokyo Institute of Technology

## I. INTRODUCTION

Many scholars believe that Chinese has had a significant influence on the development of Japanese. Despite the general consensus in academia that Chinese and Japanese originate from two distinct grammatical systems, both languages utilize Chinese characters for written communication. Furthermore, a multitude of Chinese characters and words share remarkably similar pronunciations in both languages. However, what is intriguing is the prevalence of Japanese loanwords in modern Chinese. Therefore, we aim to investigate the phenomenon of Japanese loanwords to explore certain patterns in the evolution of languages.

## II. METHODS

First of all, we need to define Japanese loanwords and which words can be loanwords. A loanword (also a loan word, loan-word, or borrowing) is a word at least partly assimilated from one language (the donor language) into another language (the recipient language, also called the target language).[1][2] For instance, "... 化 (-ise)"; "... 主义 (-ism)" are typical Japanese loanwords in Chinese. In our experiment, the Japanese corpus we employed was mainly referenced from [3].

To consider about our research topic, we choose three popular Chinese Website and randomly download a texture for each of them. We use Python to do segmentation on these textures and compare them with a list of Japanese loanwords. We calculate about the frequency of Japanese loanwords in these textures and the result is as follows:

TABLE I  
THE FREQUENCY OF THE JAPANESE LOANWORDS

Chinese Corpora	Size of the text	Loanwords frequency(%)
China Daily	1737	7.1
WeChat Official Accounts	638	8.3
Wiki	263	11.5

We can find out that the Chinese corpus from Wiki has the most frequent Japanese loanwords.

## III. DISCUSSION

According to [4], the percentage of Japanese loanwords in modern Mandarin Chinese corpora ranges from 5% to 20%. A range of 10% to 20% is also a relatively high percentage, which might lead some scholars to believe that the number of loanwords is significantly higher than the actual count. Due to the absence of a standardized Japanese loanword corpus at present, statistical data may be subject

to various influences. For instance, it can be challenging to determine whether the meaning of a particular word already existed in Chinese or was derived from Japanese.

Nevertheless, in our brief experiment, we still observed a significant number of Japanese loanwords in modern Mandarin Chinese texts, especially in scientific texts such as Wikipedia. This is largely attributed to the fact that during the development of modern Chinese, a substantial portion of scientific terminology originated from Japanese. Consequently, in texts with a higher level of specialization, the proportion of Japanese loanwords tends to be notable.

## IV. CONCLUSION

According to [5], One crucial aspect for a language to endure and pass down through generations is its adaptability. It needs to evolve and create new vocabulary staying current with the times.

Considering the influence of Japanese on the development of Chinese is meaningful. Historically, the perception was often that Japanese was influenced by Chinese, particularly through the adoption of Chinese characters to form its own writing system. However, what is interesting is that in the process of modern Chinese development, due to the shared use of Chinese characters in Japanese, it became more readily accepted by the Chinese people (partly due to historical reasons). The Japanese localized Western advanced concepts and subsequently introduced them to China in the form of Chinese characters.

This is indeed a process where two languages mutually contribute to each other's development. Similar phenomena are often observed in other linguistic systems (not limited to the Chinese character system). Such interactions contribute to the vitality of languages, enabling them to endure over time and continue to have numerous users.

## V. REFERENCE

- [1] Dictionary. Merriam-Webster. Retrieved 2 October 2022.
- [2] Jespersen, Otto (1964). *Language*. New York: Norton Library. p. 208. ISBN 978-0-393-00229-4. Linguistic 'borrowing' is really nothing but imitation.
- [3] Hou,R. *Organizing and Examining Loanwords in Japanese*
- [4] Mike Opper, *Statistics on Japanese Loanwords in the Modern Chinese Corpus*, IACL-19,Jun 16, 2011
- [5] Eckert, P., Rickford, J. R. (2001). *Style and Sociolinguistic Variation*. Cambridge University Press.



# WORDS FREQUENCY IN MY RESEARCH

## SUMMARY

HU YUNFAN

GEDES, Tokyo Institute of Technology

### 1.Introduction

During this Linguistic course, I learned what is Zipf's law.

I wonder my own research summary could fit Zipf's law or not.

Also, I am curious about the frequency of each word in the summary, whether the most frequent noun match the content of my research or not.

### 2.Method

- Use a program to count the frequency of each word.
- Turn the data into pictures and tables.

### 3.Result and Discussion

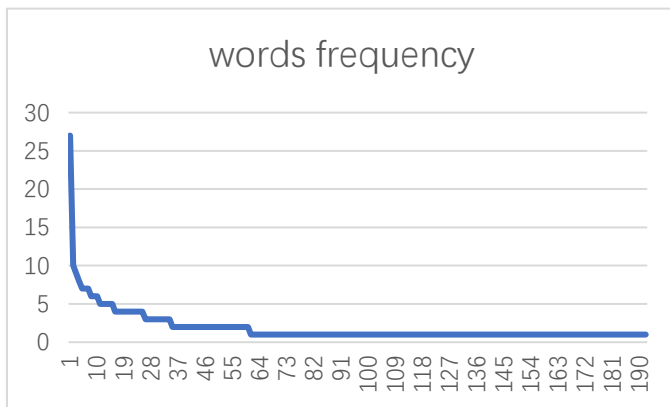


Figure 1

Zipf's law is expressed by the following equation.

$$f_r \approx \frac{c}{r} \quad (r = 1, 2, 3, \dots, n)$$

Where, the frequency, rank, and constant are  $f$ ,  $r$ ,  $c$  respectively.

From the **Figure 1**, it can be seen that the frequency of the first word and the second word comply with Zipf's law. Then it comes with several words with

same frequency of occurrence.

Table 1

the	27
of	10
a	9
in	8
and, extraction, to	7

Table 2

extraction	7
liquid	6
biodiesel, model	5
production, results, catalyst, feed, acids, program, study	4

**Table 1** shows the top 5 words frequency. **Table 2** shows the top 4 noun frequency.

Simply from the words frequency you cannot determine the context of this research, the most important key word 'extraction' shows 7 times in the summary.

But if we saw the frequency rank in noun, we might be able to infer the general content of the study.

### 4.Conclusion

- L shaped distribution looks fine, but Zipf's law might not such universal. Maybe it's because the text of objective is not enough.
- Count noun words number can be helpful in understanding the content of research article.

# Differences and similarities between Chinese and Japanese onomatopoeias

Long Yanke

Department of Systems and Control Engineering, Tokyo Institute of Technology

## 1 . Introduction

Japanese is often said to be a language full of onomatopoeia. Chinese also has a lot of onomatopoeia, but it is difficult for native Chinese speakers to learn onomatopoeia in Japanese.

This report will focus on the similarities and differences between onomatopoeia in Chinese and Japanese, finding the possible reason of this problem.

## 2 . Methods

Collect the commonly used onomatopoeias for database. Divide them with different forms, and calculate the number of words with each form.

## 3 . Result

There are 17 typical types of onomatopoeia in Japanese. In the table1, it is divided into four categories, and a portion of words that do not fit into either form is unclassified. "ABAB" style bisyllabic repeated tetrasyllabic words are the most common ones. (e.g., にこにこ, ふわふわ). Table2 for Chinese onomatopoeia, on the other hand, is divided into four categories, of which the most numerous is the bisyllabic words.

Table 1: Number of onomatopoeic words with different syllables in Japanese (with 995 words)

Bisyllable	Trisyllable	Bisyllabic repeated tetrasyllable	Nonrepetitive tetrasyllable	Unclassified
50	160	373	140	10 272

## 4 . Discussion

Considering that Japanese are epigraphic characters and Chinese are ideographic characters, Japanese onomatopoeia can be combined in more ways than Chinese onomatopoeia. Chinese words needs to take into account how well the meaning of the character fits the semantics. As a result, like showed in the table1, Japanese will have many more forms of onomatopoeia, as well as many onomatopoeias that cannot be categorized.

## 5 . Conclusion

In terms of lexical morphology, Japanese onomatopoeia can be recognized by "repetition" and the use of つ, ー, and ん. Chinese onomatopoeia can be partly recognizable by repetition, but has no distinctive features. The composition of Chinese and Japanese onomatopoeia differs due to the different language families.

Reference:

- [1] Yang, Lingling. Comparison of Chinese and Japanese onomatopoeia onomatopoeia.2007.Shandong Normal University,MA thesis.
- [2] Fan Minlei. A comparative study of onomatopoeia and onomatopoeia in Chinese and Japanese languages--Taking "どんどん" as an example[J]. Jiangsu foreign language teaching research,2014(02):50-55.
- [3] KAKUOKA K. A study of the morphology of the Japanese onomatopoeic lexicon[J]. Annual Report of Ryukoku University International Center, 2006, (5): 611-635.

Table 2: Number of onomatopoeic words with different syllables in Chinese (with 576 words)

Monosyllable	Bisyllable	Trisyllable	Tetrasyllable
96	309	108	63

# The Rapid “Katakana-ification” of the Japanese Language

Nathan Van Alstine | [vanalstn@elsi.jp](mailto:vanalstn@elsi.jp)

M2 Student - ELSI, Department of Earth and Planetary Sciences, Tokyo Institute of Technology

## Introduction

How and why do languages evolve and die? Native speakers of a language acquire words and meaning naturally of their language as they grow from a child. The “generation” we learn language in is simply a contemporary version of the language we speak. Changes in cultures, international relations, improvement of technology, etc. introduces new words into the lexicon seemingly out of nowhere and as a result, newer and more frequent words unknowingly push other words into less frequent use. At some point, with large enough changes over time, a language will sound like a very different language than it did in the past. This is the case with Old English vs modern English – and is a path modern Japanese is on.

Katakana is used in Japanese mainly to transcript foreign and loan words (外来語) into Japanese, roughly originating around 800AD to transliterate Chinese writings into Japanese. With the massive improvements in technology for international interaction in the last few centuries, particularly WW2 and beyond, the Japanese language has seen an influx of katakana loanwords more than ever. From 1956 to 1994 alone, foreign loanwords have seen a near three times increase in usage. This begs the questions – At what point will the influx and usage of different words alter the Japanese language enough that it is incomprehensible to those in the past, and how much language-change is required to cause this?

## Methods

To look at this, I began by finding different popular national Japanese dictionaries, their statistics posted since 1956, word databases, and databases of loan word frequency from popular magazines. With this data, the proportion of loan words was counted vs. total count of words. One major Japanese dictionary examined was the Koujien (広辞苑) for multiple editions published across decades (Igarashi, 2007). Two databases of words were also examined from the Academic Repository of the National Institute for Japanese Language and Linguistics (国立国語研究所学術情報リポジトリ) and CiNii (国立情報学研究所).

## Discussion

Increasing frequency of katakana and loanword usage show a clear upward trend, especially post 1980s. The small increase between 1956 to 1991 may be a case of Japan and Japanese media maintaining relatively closed off influence from the West or other cultures. The dramatic increase post 1990s, is likely a result of the proliferation and greatly increasing use of the internet - as not only was Japanese media much more easily accessible to the rest of the world, but Japan also now had much greater access to the rest of the worlds media, and thus influence. Peculiarly, loan word would include all foreign words that enter the Japanese language however, “loan word (外来語)” in Japanese generally refers to words taken from Western origin, which shows a recognition that a large influence of change in the language is due to Western influence.

TOTAL PROPORTION OF LOAN WORDS VS. JAPANESE WORDS OVER TIME

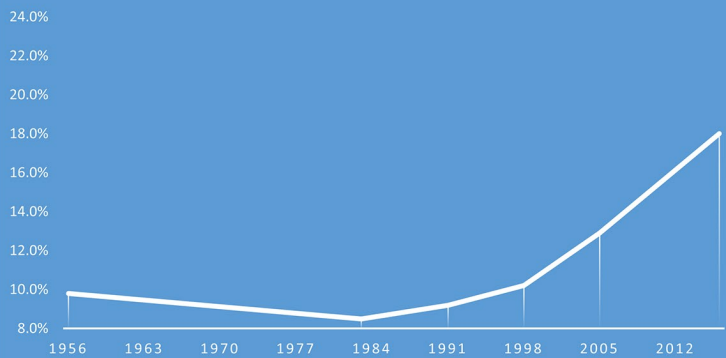


Figure 1. Increase of proportion of loan words to total words in the Japanese language since 1956 to modern times. Sources from Koujien and popular magazine statistics.

Referring back to the latter half of this projects original question, how much language change is required to completely change the Japanese language – this also begs the question: Why is the Japanese language so receptive to taking in loan words versus creating a new word from it's own lexicon? I'd like to introduce common Japanese words and their loan word counterpart (Table 1) and common katakana loan words and potential Japanese alternatives (Table 2). Though some of the examples in Table 1 may have nuanced differences, their general use is interchangeable. Table 2 creates “new words” from pre-existing Kanji that would suffice the meaning of the original nouns purpose and would be simple to decipher having known the Kanji prior. Despite this, the Japanese language has chosen to adopt the original loan word instead of adapting it to the language.

Common Katakana	Japanese Alternative	Original Japanese	Common Katakana
アナウンサー	発表者 (はっぴょうしゃ)	扉	ドア
エスカレーター	自動階段 (じどうかいだん)	会計	チェック
ハンカチ	小布 (こぬの)	建物	ビル
ホテル	泊館 (はくかん)	上着	コート
		画面	スクリーン

Tables 1 and 2 (left and right respectively). Table 1 demonstrates common katakana words and Japanese alternatives that are not used. Table 2 shows original Japanese words and their common katakana counterparts.

## Results

Data from popular magazines and Koujien shows little change post WW2 until around the mid-80s where influx of loan words beings to increase greatly over the decades. From 1983 to 2016, a change from 8.5% to 18% is shown. The data for 1983 to 1998 is based entirely on Koujien statistics, which show a steady gradual increase over 15 years. Data points 1956 and 2005 to 2016 are sourced from popular magazines word usage data (Ebeid, 2016), which shows a higher percentage of loanwords in 1956 versus Koujien. Post 2005 data shows a sharp increase over the decade to 2016 of 12.9% to 18.0% loanword usage. Compared to changes shown in Koujien, this is roughly a 4x quicker ratio of increase over time.

## Conclusion

As the rapid influx of loan words continues to enter the Japanese language, would we ever expect a complete “English/Western-ification” of the language? Is there a quantifiable point that this would result in Japanese being a dead language or the consideration of modern Japanese to be “Old Japanese”? In looking at dictionary data, it is very telling when loan words are now “officially” accepted into the Japanese language. In looking at magazine data, it shows a potential path, as many modern and nuanced topics covered in magazines will use the most recent lingo which has been accepted by sub-communities, but not yet nationally recognized as vocabulary. In observing and predicting this change, is the fate of Japanese and its inclusion of loan words a natural path of language change or a predictable language death – and more importantly should this change be allowed to happen naturally, or should there be an intervention or effort to preserve the language as much as possible as is?

## References

Igarashi, Y. (2007). The Changing Role of Katakana in the Japanese Writing System: Processing and Pedagogical Dimensions for Native Speakers and Foreign Learners. [Dissertation] University of Victoria  
Ebeid, E. A. (2016, May 3). 日本語を脅かす外来語. nippon.com. <https://www.nippon.com/ja/column/g00195/#>  
Databases Provided by:  
<https://repository.ninjal.ac.jp/>  
<https://ci.nii.ac.jp/books/>



## Introduction

While estimating the mood in a sentence is easier in face-to-face communication, since the body language and facial expression give many hints, understanding mood in a written message is much harder due to lack of such cues. As such, many ways are used in modern messaging to communicate the mood. While emojis and emoticons are the most popular example of such ways, an emerging way is the use of abbreviations and other textual signals at the end of the message. This study aims to check the influence of these signals in determining the mood of a sentence.

## Methods

6 example sentences were chosen from [1], half of which represented a positive situation other a negative one. lol, /sigh and /cry were attached randomly at the end of each of the 6 sentences. These 3 texts signal a positive, negative and an ambiguous mood respectively. This was done 10 times, to create a 10 sets of 6 sentences each. Each of these sets were given to 10 people, and they were asked to estimate the mood behind all the sentences, and to answer between "Positive", "Negative", "Neutral", "Ambiguous". The data was then studied and plotted using the Julia Language.

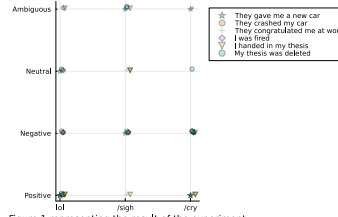
## Results

Here is a table showing number of textual signals and estimated moods in 60 sentences for both positive and negative situational sentences.

Table 1

	positive situational	negative situational
lol	9	14
/sigh	10	7
/cry	11	9
Positive	15	5
Negative	4	18
Neutral	5	4
Ambiguous	6	3

Here is a graph depicting things more accurately. Note that darker the color, the more number of people voting estimating the same result.



## Discussion

As shown from the data, the textual signals have some impact on the determination of mood in the sentence.

For example, there were 15 instances where the subjects couldn't determine that the sentence has a positive mood behind it, for a positive situational sentence. 6 of those couldn't determine the mood at all (ambiguous).

The same was reflected for negative situational sentences as well, where 5 thought that the mood was positive.

Other than that, similarly situational sentences paired with similar mood depicting texts gave the appropriate results in all cases.

## Conclusion

This experiment showed the impact of add textual signals at the end of sentences when determining the mood.

While there is considerable influence of textual signals, the majority still determined the mood close to the original situation (positive or negative) of the sentence.

References: Nerea Aldunate et al (2018), Frontiers in Psychology, Volume 9

# The Effects of Syllable Complexity on Verbal Information Transmission Speed

Jacob Lindahl

Mathematics & Computing Science, Tokyo Institute of Technology

## 1. Introduction

How do diversity of phonetic inventory and complexity of syllable structure affect the speed at which information can be communicated?

## 2. Methods

1. Survey of phonetic inventory cardinalities.
2. Sample reading durations of standard translated text.

## 3. Results

Data show a small negative correlation between all thin phonetic complexity metrics and reading duration. Removing outlier (French) reduces correlation.

## 4. Discussion

The strongest correlation is vowels:time with  $r = -0.5042$ ,  $R^2 = 0.254$ . The correlation is *weak*. More and better data are required to confirm the relationship.

## 5. Conclusion

Data are inconclusive. We cannot say for certain that syllable complexity affects the rate of communication.

### References

- Chew, Peter A. A Computational Phonology of Russian. Jesus College, University of Oxford, 2003.
- Kumar, Aman. "Aspects of Hindi Syllable Structure." Aligarh Muslim University, 2005.
- Labov, William, Sharon Ash, and Charles Boberg. The Atlas of North American English: Phonetics, Phonology, and Sound Change : A Multimedia Reference Tool. Walter de Gruyter, 2006.
- "Les Syllabes | Français Interactif." Accessed November 18, 2023. <https://www.laits.utexas.edu/fi/html/pho/05.html>.

Table 1: Reading Time / Phonemic Cardinality

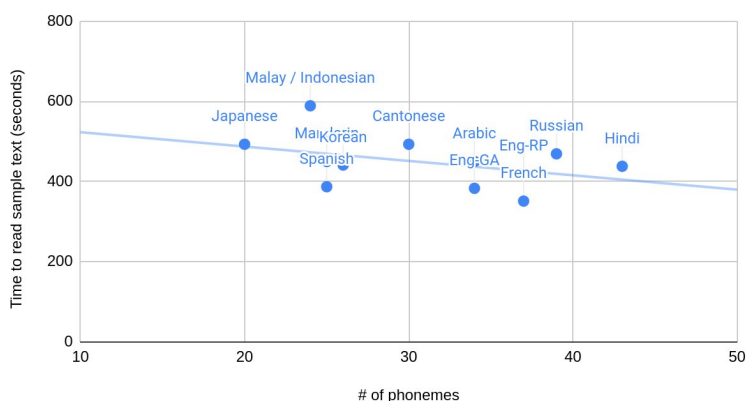
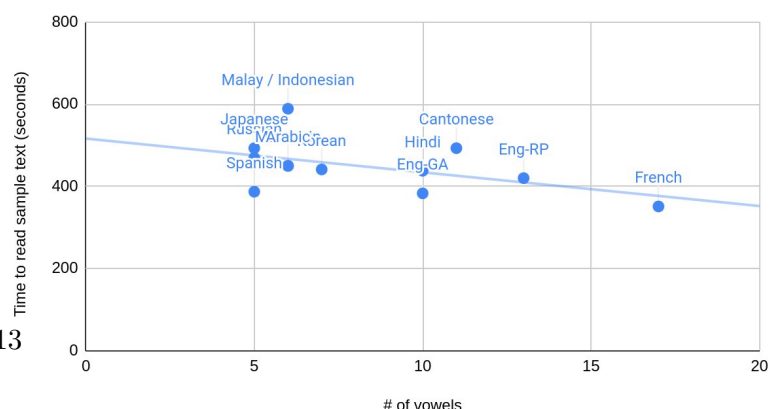


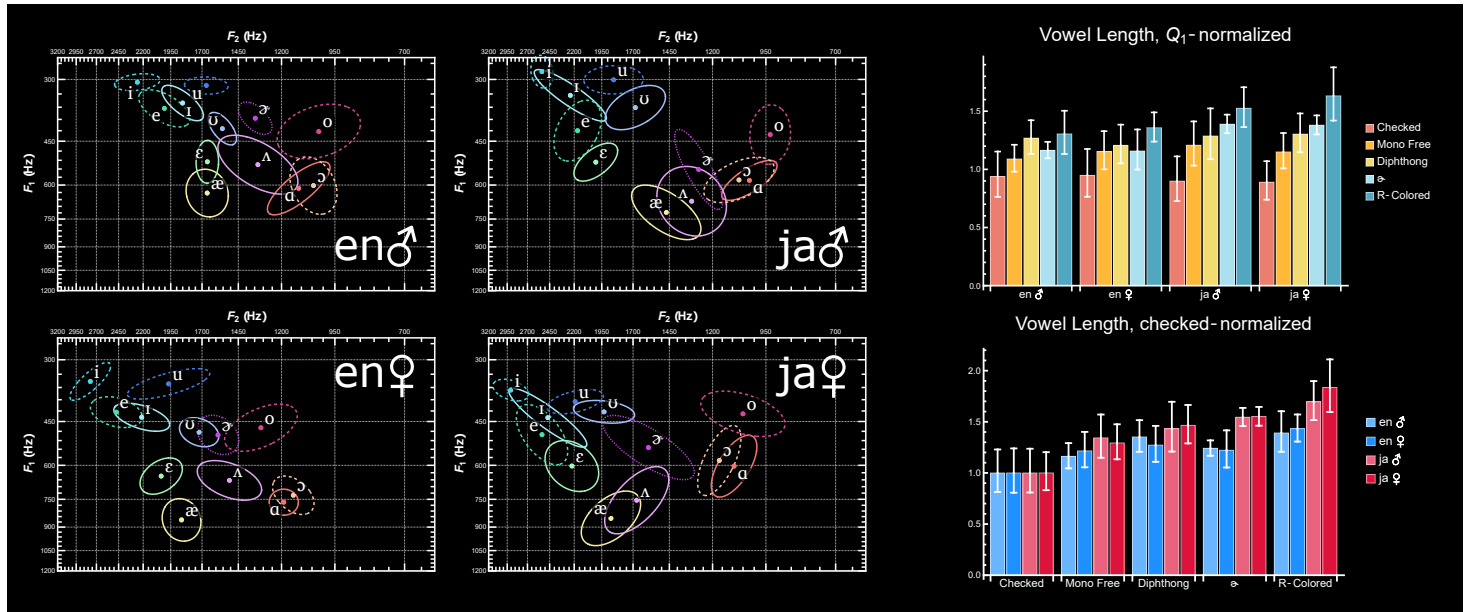
Table 2: Reading Time / Vowel Cardinality



# English Vowels at ELSI

## A Comparative Study of English L1 and Japanese L1 Speakers

Department of Earth and Planetary Sciences, Tokyo Institute of Technology  
**Austin Taylor**



## Background

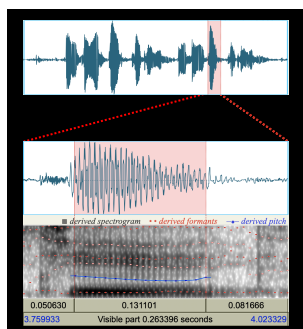
- Japanese has a limited vowel inventory, but distinguishes different vowel lengths
- English has many more vowels of different quality, but vowel length is phonemic
- **Japanese L1 speakers should have English vowels with overlapping formants, compensated by vowel length**

## Statistical Testing

t-tests at 95% significance were taken for vowel length  
 L1 Japanese speakers show significant differences in length for several categories; L1 English speakers do not

	Checked	Free Mono	Diphthong	schwa	R-Colored		Checked	Free Mono	Diphthong	schwa	R-Colored
en♂	1.23x 10 <sup>-1</sup>	7.47x 10 <sup>-3</sup>	9.47x 10 <sup>-2</sup>	2.95x 10 <sup>-1</sup>		ja♂	6.77x 10 <sup>-3</sup>	5.88x 10 <sup>-1</sup>	1.46x 10 <sup>-3</sup>	3.63x 10 <sup>-6</sup>	
		1.40x 10 <sup>-1</sup>	5.87x 10 <sup>-1</sup>	7.44x 10 <sup>-2</sup>				6.01x 10 <sup>-1</sup>	2.63x 10 <sup>-1</sup>	2.28x 10 <sup>-2</sup>	
			5.10x 10 <sup>-1</sup>	8.00x 10 <sup>-1</sup>					5.88x 10 <sup>-1</sup>	1.64x 10 <sup>-1</sup>	4.43x 10 <sup>-1</sup>
				3.76x 10 <sup>-1</sup>							
en♀	6.32x 10 <sup>-2</sup>	4.44x 10 <sup>-2</sup>	2.83x 10 <sup>-1</sup>	6.52x 10 <sup>-1</sup>		ja♀	1.05x 10 <sup>-2</sup>	1.19x 10 <sup>-1</sup>	1.42x 10 <sup>-3</sup>	3.04x 10 <sup>-7</sup>	
		6.87x 10 <sup>-1</sup>	9.80x 10 <sup>-1</sup>	8.24x 10 <sup>-2</sup>				2.53x 10 <sup>-1</sup>	1.46x 10 <sup>-1</sup>	1.14x 10 <sup>-3</sup>	
			8.26x 10 <sup>-1</sup>	2.77x 10 <sup>-1</sup>					6.66x 10 <sup>-1</sup>	5.46x 10 <sup>-2</sup>	
				3.81x 10 <sup>-1</sup>						1.94x 10 <sup>-1</sup>	

## Methodology



4 groups of 3 speakers  
 19 English vowels recorded  
 Vowels are analyzed with Praat, capturing length and taking formant data across the length  
 Data is normalized for speaker's lengths and statistics are taken from the logs of these values

## Discussion

As predicted by L1 interference, the L1 Japanese speakers at ELSI generally have overlapping formants compensated by a length contrast between checked and free monophthongs  
 L1 English speakers at ELSI are about half and half on the CAUGHT-COT merger  
 The speech samples are limited to prompted phrases and L1 Japanese have varying levels of English speaking ability. Because of the small sample, take these results with a grain of salt.

## Acknowledgements

Yazawa et al., *Spectral and temporal implementation of Japanese speakers' English vowel categories: A corpus-based study*, 2023  
 Kato et al., *Analysis of L2 English Vowel Production by Native Japanese Children in Domestic Elementary School*, 2019.  
 Oh et al., *A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting*, 2011  
 Paul Boersma and David Weenik, Universiteit van Amsterdam for producing and offering Praat on GPL version 3.



東京工業大学 地球生命研究所