Japanese Primer

In this section will do a **quick primer** on the ==**different features of the Japanese Language**==.

Why does Japanese makes for a good use-case for learning LLM/GenAI workloads?
- Japanese is a high context language.
- Japanese has multiple writing systems
- To get good performance with Japanese output with LLMs you need to use custom models

Japanese Writing System

にほん

Hiragana
Represents Japanese native words.

ニホン

Katakana
Represents foreign loanwords, technical terms, onomatopoeia, and for <u>emphasis</u>.

日本

Kanji
Borrowed Chinese characters
~3000 Kanji are used in common conversation

nihon

Romaji
representation of Japanese sounds using the Latin alphabet

JP

Emoji (Bonus)
representation of Japanese sounds using the Latin alphabet

⚠️ We are representing the same word in different writing systems.
The word we are using here Japan would often be represented as Kanji eg. 日本

Hiragana

n	w-	r-	y-	m-	h-	n-	t-	s-	k-	
ん	わ	ら	や	ま	は	な	た	さ	か	あ	a-
		り		み	ひ	に	ち	し	き	い	i-
		る	ゆ	む	ふ	ぬ	つ	す	く	う	u-
		れ		め	へ	ね	て	せ	け	え	e-
を		ろ	よ	も	ほ	の	と	そ	こ	お	o-

There are variants with markers: **ぼ** (bo)

There are diagraphs by smaller kana **きょ** (ki + yo) = Kyo

Katakana

n	w-	r-	y-	m-	h-	n-	t-	s-	k-	
ン	ワ	ラ	ヤ	マ	ハ	ナ	タ	サ	カ	ア	a-
		リ		ミ	ヒ	ニ	チ	シ	キ	イ	i-
		ル	ユ	ム	フ	ヌ	ツ	ス	ク	ウ	u-
		レ		メ	ヘ	ネ	テ	セ	ケ	エ	e-
	ヲ	ロ	ヨ	モ	ホ	ノ	ト	ソ	コ	オ	o-

Furigana

Furigana is when Hiragana Japanese appears beside Kanji to decipher its reading.

ゆう びん きょく
郵便局

Most Mangas will have Furigana so that all levels of readers can read the Kanji.

Kanji

Kanji can have multiple types of readings.
You cannot expect the same kanji to always make the same sound.
You simply need to remember its multiple readings.

あの 人　Ano **hito**　(Kun)

日本 人　nihon**jin**　(On)

三 人　san**nin**　(On)

You can see person can have **five different possible sounds**

jisho

Draw　部 Radicals　Voice

Kanji ▾ | 人 #kanji

人

2 strokes
Radical: man, human 人 (亻)
Parts: 人

person
Kun: ひと、 -り、 -と
On: ジン、 ニン

Words starting with 人　Words

Kanji has groups of readings:
1. On'yomi (音読み - "Sound Reading")
   • Based on the original Chinese pronunciation
2. Kun'yomi (訓読み - "Meaning Reading")
   • Based on the native Japanese word that matches the meaning of the kanji.

Words might also sound different due to devoicing, sound blending or irregularities.
For instance 人(hito) will sound more like shto

What is a Kana?

A **kana** character represents a specific <u>syllable</u> or mora (a unit of sound).
A Hiragana or Katakana character can represent a Kana but a Kanji cannot.

Let's look at the Japanese word **hajimete** which means "for the first time"

はじ
初 <u>め て</u>   When writing with a Kanji it only has two kana

<u>は</u> <u>じ</u> <u>め</u> <u>て</u>   When written completely in Hiragana it has four kanas

Both would have 4 moras (4 units of sound)   ha, ji, me, te

No Spaces Between Words

The Japanese language does not use spaces to separate words.

私は今日東京駅で電車に乗ります。

I will ride a train at Tokyo Station today.

Japanese readers do no need spaces because they tell where new words starts because they use a mix of writing systems.

私は今日東京駅で電車に乗ります。

Particles

Japanese uses **particles** to show relationships between words

Examples of particles:
- が (ga) - subject marker
- を (wo/o) - direct object marker
- に (ni) - indirect object, direction, location marker
- へ (e) - direction marker
- で (de) - location of action, means marker
- から (kara) - from, since
- まで (made) - until, to
- の (no) - possessive marker

私**は** 友達**と** 公園**で** 犬**の** 写真**を** 撮ってみたい**ね**。

私は	友達と	公園で	犬の	写真を	撮ってみたい	ね
I (Myself)	Friend	Park	of Dog(s)	Photo	want to take photo	seeks **agreement or confirmation** from the listener
topic	**with**	**at**	**possession**	**object**		

Sometimes native-Japanese speakers omit known particles.

More Similar Sound Words

Japanese language has many **homophones** (words pronounced the same.)

This could be due to two reasons:
- Japanese simple phonetic structure
- Borrowed pronunciations due to history of the borrowed Chinese Characters

Lets use the following example **こう/kou**

校 kō - school
高 kō - high/tall
公 kō - public
光 kō - light
港 kō - harbor/port
工 kō - construction/engineering
孝 kō - filial piety
考 kō - thought/consideration
効 kō - effect/effectiveness
広 kō - wide/broad
行 kō - going/journey
洪 kō - flood
講 kō - lecture
興 kō - entertainment/excitement

高校 kōkō - high school
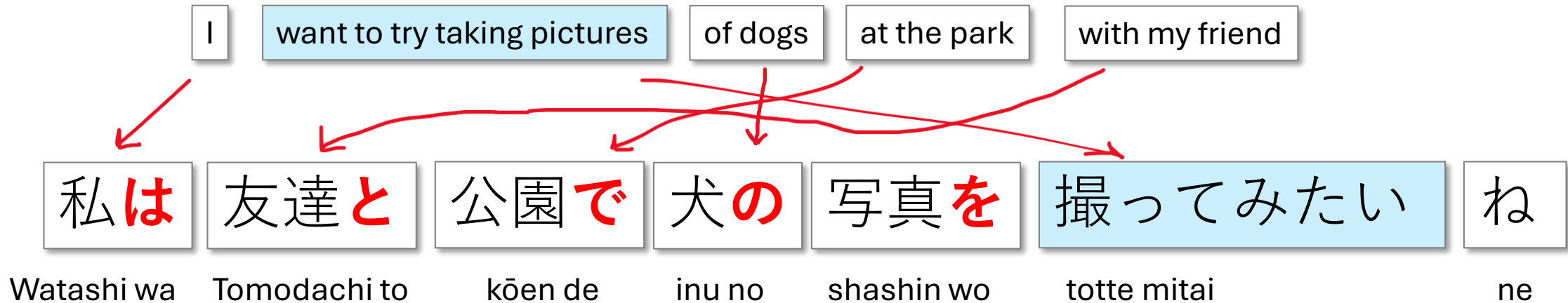公候 kōkō - public prince/duke
孝行 kōkō - filial piety
航行 kōkō - navigation
好好 kōkō - very good

Sentence Construction

Japanese sentence structure is **Subject-Object-Verb (SOV)**

English sentence structure is: **Subject-Verb-Object (SVO)**

| I | want to try taking pictures | of dogs | at the park | with my friend |

| 私**は** | 友達**と** | 公園**で** | 犬**の** | 写真**を** | 撮ってみたい | ね |

Watashi wa   Tomodachi to   kōen de   inu no   shashin wo   totte mitai   ne

Levels of Speech

In Japanese there are different levels of speech, which vary based on social relationship.

Humble Language
To your friend

昼ご飯をいただきます

hiru gohan wo itadakimasu

Honorific Language
About your boss

部長は昼ご飯を召し上がります

buchō wa hiru gohan wo meshiagarimasu

Polite Form
To your coworker

昼ご飯を食べます

hiru gohan wo tabemasu

Casual/Plain Form
To your friend

昼ご飯を食べる

hiru gohan wo taberu