An Investigation into Chinese Classifiers: The Creation and General Acceptance of New Nominal Connections in Mandarin Chinese

by

Nathaniel Aaron

B.A. in Chinese Languages and Literature & B.S. in Emergency Medicine, University of Pittsburgh, 2023

Submitted to the Undergraduate Faculty of the

David C. Frederick Honors College in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DAVID C. FREDERICK HONORS COLLEGE

This thesis was presented

by

Nathaniel Aaron

It was defended on

April 7, 2023

and approved by

Alan Juffs, Professor, Department of Linguistics

Kun Qian, Associate Professor, Department of East Asian Languages & Literatures

Jie Zhang, Associate Professor, Department of Modern Languages, Literatures, and Linguistics

Thesis Advisor: Yi Xu, Associate Professor, Department of East Asian Languages & Literatures

An Investigation into Chinese Classifiers: The Creation and General Acceptance of New Nominal Connections in Mandarin Chinese

Nathaniel Aaron, BA BS

University of Pittsburgh, 2023

This study investigates the usage and selection of nominal classifiers in Mandarin Chinese. The first part of the study reviews the categories and historical evolvement of Chinese classifiers, with introductions to the default classifier 个 *ge* and complications in classifier usages. Dialectical differences and the evolvement of classifiers in the Modern era including the Internet Age are also discussed. The second part of the thesis reports an empirical study that aimed to reveal trends of classifier selections for various objects, with a particular focus on patterns of use for unconventional objects and new inventions. Forty-four Mandarin Chinese native speakers completed the survey. The findings indicate that participants tend to default to strong associations with 台 *tai* and *ge* for inventions; participants also demonstrated positive associations with salient features particularly shape and animal-based classifiers. These results shed light on the classification trends seen in the realm of technology and provide insight into how the classification system may evolve to accommodate new nouns in the future.

Keywords: Chinese; Classifier(s) (CL); classification system; salient features; inventions

Table of Contents

List of Tables

Preface

I am incredibly grateful for the guidance and support provided by thesis advisor, Dr. Yi Xu, throughout the drafting of my work. Her insightful comments and critiques have helped me shape and refine my research. Her mentorship and dedication to scholarship have inspired me to pursue my academic goals.

Dr. Xu's constructive feedback and unwavering support have been invaluable in overcoming challenges and navigating uncertainties in the writing process. Working with her has been an honor, and I owe her a debt of gratitude for her exceptional contributions to this work.

In addition, I would like to express my heartfelt appreciation to Dr. Alan Juffs, Dr. Kun Qian, Dr. Jie Zhang, and Mr. Hin Ming Chik for their invaluable guidance and support during this project. Their extensive knowledge, constructive feedback, and commitment to their students have been pivotal in shaping my work. I am truly indebted to them for the significant time and effort they have dedicated to my academic development.

Before concluding, I would like to acknowledge the University of Pittsburgh Department of East Asian Languages and Literature and the Frederick Honors College for their academic support in the creation of this paper. It is an honor to have had the opportunity to work alongside such inspiring mentors and colleagues. I am deeply grateful for their guidance and encouragement.

1.0 Introduction

A salient feature of Mandarin Chinese and related family dialects is the requirement of classification when referring to nouns in spoken and written forms. Nominal chunks must be linked with a Classifier, where such Classifiers will precede the nominal heads and act with little to no semantic function in the general usage. The term Classifier and Measure words are used synonymously but do differ linguistically. Classifiers refer to inherent or permanent properties of entities while Measure Words refer to the temporary or contingent properties of objects (Tai, 1994). Referring to counts, quantities, and groups of objects are Measure Words as they are not characteristics, but an interim state used to help relay a countable state, quanta, or arrangement of objects. Classifiers are permanent, intrinsically highlighting features of objects to emphasize meaning and categorization.

一条鱼 yi tiao yu	一群鱼 yi qun yu
One [CL for long, thin objects] Fish	One [group, school CL] Fish
条 : Classifier	群 : Measure Word

In this study, the term Classifier(s), CL, will be used as an umbrella term to contain Measure Words, MW, where appropriate. With such a grand requirement for the Chinese language to have a dependency on CL, the vast choices of CL are further subdivided for linguistic characterization and clarity.

This literature synthesis and subsequent investigation aim to gain insight into how certain inventions may adopt a CL-Noun connections and the associations formed in the mental schema

of native speakers. In particular, this paper will focus on how inventions have a unique opportunity to establish dominant CL-Noun connections for future use. The investigation will involve presenting new inventions to participants, through a survey, and asking them which CL they would use—to identify patterns and capabilities of CL usage.

Before the investigation, the literature synthesis below will provide an overview of the current knowledge on CL usage, habits, history, and generalizations in Mandarin Chinese and similar dialects. This contextualization will help shed light on potential findings and conclusions seen in the survey.

2.0 Literature Synthesis

2.1.1 Classification in History

The classification system most notably developed during the Han dynasty during dynastic China, but the structure of CL appeared very different than today's system. In the first and second centuries, the first evidence of specific types of CL appeared in ancient Mandarin Chinese and we see a dramatic increase in the utilization of a CL system (Sybesma, 2017). Although the Han dynasty was not the first instance of CL, it was the greatest change to the system too much of what it looks like today.

Back in the oracle bone carving era, CL looked very different. As the structure varied, one can hypothesize the motivations for using a system of nominal classification ought to be similar to today—to aid in object classification and separate noun chunks from other sentence particles. Peyraube (1998) discusses the many forms of CL structure:

[Noun + Numeral + Noun] [Numeral + Noun] [Noun + Numeral]

Three main structures existed concurrently in ancient Mandarin Chinese. These materialized with examples looking like:

羊一羊 *yang yi yang*	人十有六人 *ren shi you liu ren*	羌十人 *Qiang shi ren*
[Sheep one sheep]	[people ten and six people]	[Qiang 10 people]
One Sheep	sixteen people (men)	10 Qiang people

These variations on the Noun + Numeral chunks may be proto-classifiers as they attempted to modify the object being described to some extent. This way of classifying was a start to the creation of the system we know today, but it was still ages away from utilizing specified CL for specific nouns (Sybesma, 2017).

Not until the Han Dynasty did, we see a standardized system form, like in modern Chinese. Peyraube (1998) finds that over the next 1,000 years, the system would change the specialties of CL and how they are semantically used. Large transitions from the [Noun + Numeral + CL] form to the modern [Numeral + CL + Noun] were seen (Sybesma, 2017). A major change as the idea became accepted was that CL modified the noun. In the modern era, CL are typically in a [Numeral + CL + Noun] structure, exceptions do exist in grammatical phrases.

2.1.2 Classification Categorization

With such a large and diverse collection of CL, many linguists will categorize based on broad characteristics of form and function. Lobben et al. (2022), chooses 3 overarching categories of: Sortal, Container, and Default CL.

1. Sortals are CL that aids in context with their object to describe a salient feature of the carried object. Examples of: 件 *jian*, 张 *zhang*, 只 *zhi*, 条 *tiao*, 块 *kuai*.
2. Container CL being separated for quantity or vessel. Examples of: 杯 *bei*, 盘 *pan*, 瓶 *ping*.
3. Default CL is the general CL of Chinese, 个 *ge*, for which context and use will be explored in this paper.

While some linguists divide CL differently, Tai (1994) names 7 categories based on CL purpose, to which only four categories that apply to Chinese, "(1) material, (2) shape, (3) consistency, and (4) Size" (Tai, 1994). Highlighting a few examples from each category as follows:

(1) Material: 匹 *pi*, 条 *tiao*, 口 *kou*, 根 *gen*, 尾 *wei*, 棵 *ke*.

(2) Shape: 条 *tiao*, 只 *zhi*, 根 *gen*, 支 *zhi*, 张 *zhang*, 块 *kuai*, 把 *ba*, 封 *feng*.

(3) Consistency (Rigidity): 根 *gen*, 条 *tiao*, 团 *tuan*, 块 *kuai*, 支 *zhi*, 尾 *wei*, 领 *ling*.

(4) Size: 头 *tou*, 粒 *li*, 座 *zuo*.

Tai (1994) subdivides Consistency into objects that are flexible, hard, and non-discrete. The Shape Category gets divided into objects that are long, flat, and round, then further into dimensions with CL such as 张 *zhang* (1st dimension), 条 *tiao* (2nd dimension), 只 *zhi* (3rd dimension), and 根 *gen* (3rd dimension).

Regardless of the method of categorization, the critical features that help group CL remain constant. To preface, Lobben et al. (2022) approach to classification will be used as an umbrella categorization structure for the remainder of this paper.

2.1.3 Sortals

A CL will act as a bridge to help emphasize an attribute of the noun to the audience, sortal CL do this remarkably well in Chinese context. Sortals, one of the largest of the CL categories, focuses on matching salient features of an object to a CL with similar 'hinted at' features. While CL contains no concrete semantic meaning themselves, they tend to have implicit characteristics based on their classification habits and features of the nouns they carry, which this paper investigates.

The pairing between sortals and nouns endorses a tight bond, one capable of invoking a deep understanding of specific features of the intended noun it is paired with (Lobben et al., 2022). Some say Chinese has 100 commonly used sortal CL, the majority have a wide set of nouns capable of being carried, while others may be specific to only a few nouns. (Sybesma, 2017). The vast category of sortals has grown to service a wide variety of nouns, such as:

一根香蕉 *yi gen xiangjiao*	一张床 *yi zhang chuang*	一头羊 *yi tou yang*
One [long cylindrical CL] banana	One [flat CL] bed	One [livestock CL] sheep

This act of underscoring specific, definable traits of nominals in pairs in turn, has some profound benefits in language exchange—allowing better remembrance and focus by the audience (Lobben et al., 2022). This makes sense as it is easier to remember an object when speakers add a notable descriptor.

Furthermore, it is typical that when a CL is correctly used in a sentence, the speaker and audience both gain context and understanding of the object being described. A simple example in English could perhaps be:

The apartment is two streets over.

What is interesting about this context is how the audience may begin to make leaps in knowledge. They may extrapolate to better understand the *streets* being described—are they windy, long, small, terraneous, etc. Whether consciously or not, the audience will produce an understanding of the object. The same situation comes up in Chinese. However, CL in Chinese helps facilitate mental-image production or nominal understanding within context.

一条路 *yi tiao lu*

One [long, thin CL] road

条 *tiao* (‘long, thin CL’) intrinsically carries the meaning for long, thin objects, so while the road could be of any size and condition, an audience’s mental image of a conventional roadway would be a sufficient starting point given the context.

2.1.4 Sortal Complexities

Nouns, however, do not always fit in these categories so uniformly. Many nouns carry conceptual meanings and interpretations, and thus some CLs seem to extend their ‘intrinsic’ meanings to carry those odd or nonconforming nouns. Such abstract occurrences create dissent in the CL pairing process and some sortals are selected for reasons unknown.

Tiao may be chosen in favor of more ‘semantically-close’ CLs. Snake is a good example:

一条蛇 *yi tiao she*	←	一只蛇 *yi zhi she*
One [long, thin CL] snake		One [animal CL] snake

In this case, although the animal CL would be a sound connection, the salient feature of snakes that is selected for is their long and thin bodies, to which *tiao* is preferred.

一条裤子 *yi tiao kuzi*	←	一件裤子 *yi jian kuzi*
One [long, thin CL] pants		One [clothing CL] pants

In the case of pants, *tiao* is also preferred. Although the clothing/apparel CL would be an intelligible match, *tiao* is commonly used instead.

The standard *tiao* CL for long, thin objects can be subdivided further as 根 *gen*, being typically used for long, rigid objects, and 支 *zhi* for long, cylindrically, rigid objects (Tai, 1994).

These subsets of *tiao*, however, do not always stand as the correct CL to use. Take a simple object like toothpaste; a review of a Chinese-to-English workbook (Li & Xiang, 2009) shows CL 管 *guan* ('tube CL') as the correct pairing. An English grammar workbook shows *tiao* as the correct pairing (Su, 2019). And a Mandarin Chinese linguistic book shows 支 *zhi* as the correct pairing (Zhang, 2006). This object will also be explored in the survey investigation, however, surely an interesting object to spark numerous possible CL connections. CL such as *tiao* has hundreds of nouns it can classify, and thus the extensions are equally as large.

一条牙膏 *yi tiao yagao*	一支牙膏 *y izhi yagao*	一管牙膏 *yi guan yagao*
One [long gen. CL]	One [long, cylindrical CL]	One [tube, pipe CL]
toothpaste	toothpaste	toothpaste

As discussed earlier, the extension of CL to cover abstract objects and concepts carries throughout many CL, and specifically *tiao* which endorses a connection between many invisible and conceptual objects (Tai, 1994).

一条新闻 yi tiao xinwen	一条法律 *yi tiao falü*
One [long gen. CL] news	One [long gen. CL] legal cause
一条意见 *yi tiao yijian*	一条命令 *yi tiao mingling*
One [long gen. CL] opinion	One [long gen. CL] order

A misstep in associations with CL can lead to erroneous interpretations similar to the following:

一根线 *yi gen xian*	一条线 *yi tiao xian*
One [long, cylindrical CL] Thread	one [long, thin gen. CL] line
One thread	One subway line

There are differences in how animals are connected to certain CLs. For example, a wolf and a dog are not categorized the same, and so carry different CL. Yet, evolutionary the wolf and dog are closely related (Tai, 1994). Note that CL usage may be subject to dialectical differences as well, see *Dialectical Findings*.

一只狗 *yi zhi gou*	一条狼 *yi tiao lang*
One [small animal CL] dog	one [long gen. CL] wolf

The same pattern is seen for foxes and cats. While evolutionarily and shaped similarly, foxes and cats host different CL similar to dogs and wolves.

一条狐狸 *yi tiao huli*	一只猫 *yi zhi mao*
One [long gen. CL] fox	one [small animal CL] cat

With objects such as 凳子 *dengzi* (‘bench’) and 毛巾 *maojin* (‘towel’), using *tiao* as the CL would be only in the case of extreme length, otherwise, they would take on their respective CL outside of *tiao* (Tai, 1994).

一张凳子 *yi zhang dengzi*	一块毛巾 *yi kuai mao jin*
One [flat CL] bench	one [piece CL] towel

Within the animal CL groupings, those intrinsic sortal CL meanings serve the nominal in differing ways such as:

一个羊 *yi ge yang*	一只羊 *yi ge yang*	一头羊 *yi tou yang*
One [gen. CL] Sheep	one [animal CL] sheep	one [head/animal CL] sheep
Neutral, refers to sheep	one, small sheep	one tame/domesticated sheep

The CL for sheep may vary among speakers, context, and surely dialectal differences explored later in this paper. But the complexity is of necessary note.

Further complexity in that a classifier's inherent meaning defines different features of the noun it carries (Sybesma, 2017). As in with:

一支花 *yi zhi hua*	一朵花 *yi duo hua*
One [flower CL] flower	One [flower CL] flower
Stem and flower reference	Petal reference

This pairing 支花 *zhihua*, refers more to the flower and stem entire unit of the flower, whereas the specific CL 朵花 *duohua*, semantically speaks to the flowering petal portion of the flower (Sybesma, 2017).

As illuminated, the CL system and nuances are quite vast. Further investigations into the CL system and categories will continue, but such variations are of note in the robustness and entangled drives of Chinese CL.

2.1.5 Container CL

In addition to sortals, CL may be sorted as Containers (Lobben et al., 2022). This category differs from sortals, in that containers are used to define units, highlight arrangements, and characterize objects in natural units for countability (Lobben et al., 2022). For example, CL for plate, cup, box, set, bottle, etc serve this role.

一盘 *yi pan*	一杯 *yi bei*	一盒 *yi he*	一套 *yi tao*	一瓶 *yi ping*
One [plate CL]	One [cup CL]	One [box CL]	One [set CL]	One [bottle CL]

Contrasting with the sortal category, container CL offers no meaning to the noun or audience (Lobben et al., 2022). Instead, they offer counts or units in context. No grammatical differences are required for either CL type.

Many linguists may see container CL as a form of Measure Word in Chinese, insofar as Lobben et al. (2022) denotes container CL as a defined CL category. Common examples include:

一盘米饭 *yi pan mifan*	一杯茶 *yi bei cha*	一壶茶 *yi hu cha*
One [plate CL] of rice	One [cup CL] of coffee	One [kettle CL] of tea

A unit of measurement is subjective in nature, and while the container CL has strict usage rules, it does offer some flexibility. When compared to sortals, the link between a noun and CL

pair is much looser in this relationship. More flexibility allows for speaker preference and personal decision to play a role to some extent. Whether described as a cup, bowl, or plate of rice for instance is within the speakers' discretion.

一碗米饭 *yi wan mifan*	一盘米饭 *yi pan mifan*	一小碗米饭 *yi xiao wan mifan*
One [bowl CL] of rice	One [plate CL] of rice	One [small bowl CL] of rice

The usages vary between incidences, leading to a finding that a "unit" of water may boast 7 different types of container CL ranging from bottle, glass, puddle, drop, can, etc (Lobben et al., 2022). This incidentally adds clarity and context to the noun being described, but how does having such a broad category of containers influence the use of CL itself? The next section may clarify such decisions.

A unique feature present in this category is the strict linguistic requirement of using a container or general CL when counting objects. As we will investigate, the ability or mistaken occurrences of omitting a sortal CL is not shared within the container and general CL categories, thus when counting objects, linguistic rules point to a container or general CL being required.

In a study by Erbaugh (2002) to test Mandarin Chinese and Cantonese speakers' ability to retell a story, they looked at semantic mistakes and found only a fraction of errors made in a container or general CL omissions across participants, with an error rate of 0.3% in Cantonese speakers (Erbaugh, 2002). The study goes on to draw a correlation between container and general CL omissions to poor language acquisition and development (Erbaugh, 2002). Sortal CL misuses, however, seem to carry a more acceptable nature, as the sortal CL class is complex and not representative of normal language development (Erbaugh, 2002) (Stokes and So, 1997).

2.1.6 Default CL

The last category by Lobben et al. (2022) is Default CL—for solely the CL *ge*. The generalized CL, *ge*, is of a special function in Chinese. *Ge* can be used in a variety of cases when a specific CL is unknown, for quick reference, for syntax, to 'fill' CL gaps, and many more. This CL hosts the widest range of nouns, with notable nouns being daily objects, abstract concepts, and uncountable objects, among others.

一个工厂 *yi ge gongchang* One [gen. CL] factory	一个耳朵 *yi ge er duo* One [gen. CL] ear	一个答案 *yi ge da'an* One [gen. CL] answer
一个考虑 *yi ge kaolü* One [gen. CL] think	一个腔调 *yi ge qiang diao* one [gen. CL] accent	一个台灯 *yi ge tai deng* one [gen. CL] desk lamp

As *ge* is the default CL among much research, it truly hosts a wide variety of nouns. As Lobben et al. (2022) choose to classify *ge* as a default CL, Tai (1994) correlates it to inanimacy or a default CL with only the value of "existent entities". And Loke (1994) agrees that the innate function of *ge* is to provide neutrality to nouns and classify them by conceptual over perceptual features. *Ge* having its own category showcases its true variety and vastness of function.

As earlier discussed, the importance of a generalized CL among languages is invaluable. In Chinese, the use of *ge* is the common generalized CL for seemingly most nouns. The linguistics currently investigating *ge's* usage is compelling. Many conclusions persist that when a speaker has the choice of using a specific CL versus general CL, speakers tend to opt for generalization.

This finding was seen in both child and adult, sophisticated, and common vernaculars in spoken Chinese (Lobben et al., 2022). Surely native, fluent Chinese speakers gain a deeper

understanding of CL purpose and obstacles throughout their lifetime, but interestingly, even Chinese language learners and preschool speakers can understand the unwritten rule of using a general or sortal CL when unsure what the noun should be held by (Erbaugh, 2002). The pervasive use of the generalized *ge* CL, streamlines communication as speakers can typically default back to *ge* when uncertainty in classification arises.

This generalized CL is also pervasive in both high-class and colloquial speech, leading to its general acceptance and usability within Chinese (Frankowsky & Ke, 2016). As many Chinese language learners regard *ge* as a 'catch-all' CL for any instance they may be unsure of the correct CL to utilize, which holds similar to the native speaker mindset in modern times, specifically the way children speak (Sybesma, 2017).

2.1.7 Ge Investigations

Ge can refer to both physical and abstract nouns—placing it in a very versatile role (Sybesma, 2017). As mentioned earlier, sortal CL may be extended beyond their generally accepted 'intrinsic' meaning to make unconventional associations. Recall 一条蛇 *yi tiao she* selected over 一只蛇 *yi zhi she*. While native speakers generally can select *yi tiao she* over *yi zhi she*, they may still struggle with whether to extend the meaning of *tiao*. This question was highlighted in Frankowsky & Ke (2016) investigation. Where Frankowsky & Ke (2016) looks at when *ge* may be selected for over such 'extensions of preexisting sortal CL'.

The anthropocentric center of a word is the intended meaning of a word as it relates to human beings. When relating humans to animals, those animals that have diverged in evolution more recently in history, such as monkeys, chimpanzees, and orangutans are more

anthropocentrically close to humans. And by the same logic, an insect, worm, or parasite would be farther from human beings.

Frankowsky & Ke (2016) discovered that the connection between a noun and *ge* is often either: very far from the noun's anthropocentric center or very. As Frankowsky & Ke (2016) sees it, the two examples of monkey and oyster are anthropocentrically close and distant from humans, respectively. Both were found to carry *ge* as the most used CL when investigated.

一个猴子 *yi ge houzi*	一个牡蛎 *yi ge muli*
One [gen. CL] monkey	One [gen. CL] oyster

Ge tends to take on the two extremes—preference given to nouns being either closely or distantly related to humans (Frankowsky & Ke, 2016). Looking at those anthropocentrically mediocre or not profoundly related to humans or non-humans. Speakers were found to not opt for *ge*, such as 天鹅 *tian'e* ('swan') and 羚羊 *lingyang* ('antelope') (Frankowsky & Ke, 2016). Instead, speakers extrapolating the meaning of a sortal CL like 只 *zhi* ('animal CL') would be more common.

Hypotheses swirl, but it is plausible that this phenomenon could be related to how the animal kingdom is divided by the 'big bucket' CL such as 只 *zhi*, *tiao*, 头 *tou* that can carry a huge number of animals each. That is, when an animal does not look like the in-group, it takes on *ge*.

Seemingly *ge* has both become a miscellaneous bucket for general CL, yet it also holds particular CL rules whether cognitively known to speakers or not. When speakers make errors in choice or preference in CL, it is regarded as a lesser speaking ability, evident in the process of language learning both in first language (L1) and second language (L2) contexts (Erbaugh, 2002). The variation in the CL system remains vast, and to claim certain CL validity is debatable.

Many argue whether the *ge* usage is arbitrary, generalized, or an intellectual choice by the speaker. In Frankowsky & Ke (2016) study, researchers sifted through the Peking University Center for Chinese Linguistics CCL Corpus Database of written and spoken Chinese in addition to Google searches for modern examples to investigate the usage of *ge* in high and low-frequency nouns. In the small subset of CL studied, 76 animal examples and the CL pairing of *ge* were determined to be identically selected in both high and low-frequency groups (Frankowsky & Ke, 2016). This research helps form a better picture of the *ge* context in Chinese—from this study, speakers opt to use *ge* in many situations, and the familiarity or comfortability of the noun (animals in this case) does not seem to be directly related to how speakers would choose to CL with *ge*.

2.1.8 Dialectal Findings

Throughout Asia and the globe, there are many dialects of Chinese spoken, each with unique rules and practices regarding CL usage. There is such variation that the noun 'tree' could carry 8 differing CL across popular Chinese Dialects (Tai, 1994):

一棵树 *yi ke shu*、一根树 *yi gen shu*、一株树 *yi zhu shu*、一头树 *yi tou shu*、

一颇树 *yi po shu* 、一兜树 *yi dou shu*、一枞树 *yi cong shu*、一条树 *yi tiao shu*

Each variation in CL above has adapted to refer to different salient features of a tree, such as the longness, trunk, stem, collection of leaves, head-like, etc giving rise to such a wide variation across speakers.

In this literature synthesis, I will briefly compare some features of Mandarin Chinese, Cantonese, and other popular dialects through various examples.

2.1.9 Similarities

Commonalities in dialectal CL usage are apparent. A striking one falls back on *tiao*. Tai (1994) found that across a majority of Chinese dialects, the CL for roads, rivers, and streets, tend to stay as *tiao*. Throughout the regional differences, semantic variations, and certain preferences among dialects, *tiao* has remained an unchanged CL for some objects (Tai, 1994). So even though there are seemingly hard guidelines in some CL usages such as, in most Dialects of Cantonese, Min, and Kejia the CL *gen* is not used, instead *tiao* and *zhi* are filled in, some nouns will carry the same CL throughout many dialects (Tai, 1994).

Similar among most dialects is the object being of hard or mushy nature CL, 块 *kaui* (‘small, hard piece CL’) and 团 *tuan* (‘round, mushy CL’) respectively (Tai, 1994). Cantonese may use 舊 *gau6* as well.

一块铁 *yi kuai tie*	一团棉花 *yi tuan chuan hua*
One [hard, piece CL] steel	one [mushy CL] cotton

There are far more similarities of note, but to highlight the complexity and unique changes various dialects have with their classification preferences—the following section will be more robust.

2.1.10 Differences

The most glaring difference to bear in mind between Cantonese and Mandarin Chinese is the categorical differences in CL. Mandarin Chinese and Cantonese may both contain upwards of

200 CL, but sortal CL makes up 38% of Mandarin Chinese (Wang and Wu, 1989) and 65% of Cantonese (Killingley, 1982). The number of CL may be roughly equal, but the majority of Cantonese are sortals in nature, and thus the biggest difference in CL usage is the mixed set of objects common CL may take.

Cantonese speakers often have a better chance of using a sortal CL when compared to Mandarin Chinese speakers (Erbaugh, 2002). As a whole, Cantonese has a wider array of sensory and transient CL used for abstract concepts like smells, sounds, flashes of light, etc that are not classed out in Mandarin Chinese to such an extent (Erbaugh, 2002). The apparent trend is that Cantonese offers more variety in what associations can be made in CL choice, something not as robust in Mandarin Chinese. Instead, Mandarin Chinese offers larger grouping CL where a small few, e.g., *zhi, tiao, zhang* can carry a vast number of nouns, not seen in Cantonese in the same light.

An interesting experiment conducted by Erbaugh (2002) showcased the variety of CL used by both Mandarin Chinese and Cantonese speakers. When prompted to retell the events of a story where the CL has intentionally removed the CL, 19 Mandarin Chinese speakers used a total of 43 sortal CL, while the 30 Cantonese speakers unsurprisingly used more, and reached a peak of 241 sortal CL used to describe the same objects as the Mandarin Chinese speakers, averaging one per every six nouns. Certainly, Cantonese CL usage varies, but of note is the amount of CL used by Mandarin Chinese speakers to retell a story with the same objects (Erbaugh, 2002). Moreover, the five most frequent Cantonese CL used accounted for 81% of all tokens, meanwhile, the Mandarin five most frequently used CL accounted for only 62% of all tokens (Erbaugh, 2002). There are a few conclusions we can draw from this study, such as Cantonese has a greater amount of sortals,

yet the most common CLs can cover a wider variety of objects when compared to Mandarin CL. But this study also looked at default CL as well.

The study concluded that as speakers learn Chinese, they are less likely to overuse the general CL compared to Cantonese native learners, potentially because the alternative sortal CLs are a language developmental milestone (Erbaugh, 2002). Furthermore, when a generalized CL is used for human-related nouns, it is used far more in Cantonese than in Mandarin Chinese, accounting for 90% and 64% of usages respectively, in a study conducted by Sybesma (2017) for native speakers. The Cantonese default CL equivalent is 个 *go3*, and the second is 啲 *di1* for plurality (Li & Leung, 2007). It holds that many of the Mandarin Chinese branch dialects treasure at least one generalized CL, but some dialects use two such as Southern Min adding 枚 *mo* (Wang, 2008), and Shanghainese with *ge* and *zhi* (Erbaugh & Yang, 2006) for practically (Sybesma, 2017).

The common CL, *zhi* can refer to a wide variety of objects in Mandarin Chinese but in Cantonese *zhi* takes on different groups of objects (Erbaugh, 2002). And interestingly enough, the Hakka/Kejia can use *zhi* for people—strictly forbidden in Mandarin Chinese (Erbaugh, 2002) (Tai, 1994).

Zhi variety of object families

Mandarin	birds, cats, tigers, ears, hands, shoes, socks, etc
Cantonese	horses, oxen, teeth, eggs, loose earrings, pans, battleships
Hakka/Kejia	People

Specific CL may override the general animal CL of *zhi* (Tai, 1994). But as some dialects emphasize specific CL for animals, the inverse is also heard. A few dialects such as Xinhua,

Kaiping, Enping, Guangdong, and the Hainan Islands will consider *ge* as their animal CL (Tai, 1994).

Most Cantonese dialects and some Kejia refer to 被 *bei* (‘quilt’) and 刀 *dao* (‘knife’) with CL *zhang* for objects with the salient feature of being "flat" (Tai, 1994).

Cantonese and some Kejia Findings:

一张刀 *yi zhang dao* One [flat CL] knife (Not often used anymore)	一张被 *yi zhang bei* one [flat CL] quilt

Mandarin Chinese Differences:

一把刀 *yi ba dao* One [hand-held obj CL] knife	一条被 *yi tiao bei* one [long gen. CL] quilt

As we dive further into CL differences, in Mandarin Chinese the CL for animal ranges from *zhi, pi, tiao, tou, kou*. But in Dialects originating from Sichuan, and Shaanxi they will use *gen* to classify long-shaped animals, not seen equivalently in Mandarin Chinese (Tai, 1994). Similarly, in Southern Min Dialects, the CL 尾 *wei* is used for long-tailed animals like fish, dragons, etc (Tai, 1994).

Mandarin Chinese *tiao* for snake and dragon

一条蛇 *yi tiao she* One [long, thin CL] snake	一条龙 *yi tiao long* One [long, thin CL] dragon

Sichuanese, Guanzhong Varieties: *gen*

一根蛇 *yi gen she*	一根龙 *yi gen long*
One [long, thin CL] snake	One [long, thin CL] dragon

Southern Min: *wei*

一尾蛇 *yi wei she*	一尾龙 *yi wei long*
One [long tail CL] snake	one [long-tailed CL] dragon

In Southern Min, CL *tou* is for large animals. Cantonese and Kejia dialects commonly extend the usage of *tou* to plants as well (Tai, 1994). Furthermore, Cantonese speakers used to use *kou* and 眼 *yan* ('hollow CL') in more colloquial speech, also also found in more classical usages of Mandarin Chinese (Tai, 1994). Nowadays, differences in Hong Kong and Guangzhou Cantonese may further marginalize these findings.

Classical Mandarin Chinese Usage and Older Cantonese Usage

一眼针 *yi yan zhen*	一口猪 *yi kou zhu*
One [hollow CL] needle	one [large CL] pig

More modern Cantonese Occurrences:

一条针 [*yi tiao zhen*] and 一只猪 [*yi zhi zhu*]

The variation in CL usage from person to person is an amalgamation of all their past experiences, semantic preferences, cognition, and nominal connections they make during comprehension; no wonder such differences are woven throughout CL frequency.

Beyond differences in CL in dialects, there is also a semantic difference in how those CL are interpreted by the audience. A Cantonese bare noun carries no implicit connection with the speaker or actor. The phrase 'Dogs like meat' has no connection to the speaker, but in Mandarin Chinese that could in addition to a bare expression, be interpreted as '[My, our, your, his, or her] Dogs likes meat'. This is also seen with CL carrying demonstrative usages. Object locators like 这/那 *zhe/na* ('this/that'), in Mandarin Chinese are necessarily needed, but in Cantonese, a CL may hold such a demonstrative function as well e.g., the phrase 啲学生 *di xuesheng* ('those group of students').

Additionally in Cantonese, bare nouns can only be interpreted as singular, whereas Mandarin Chinese endorses an ambiguity in translation (Sybesma, 2017). Surely to avoid misinterpretations a pluralization of the noun can be done through CL as well. Both dialects support a plural CL such as 些 *xie* in Mandarin Chinese and the Cantonese addition of 啲 *di* can be used to refer to a group or multiple objects but is strictly not allowed in either dialect for counting (Sybesma, 2017). As mentioned earlier, it has been shown that misuse or mistakes in CL usage are rarely seen in counting, and that trend is woven into pluralization in many dialects as well.

2.1.11 Modern Changes

As the modern age continues to grow in complexity and new words sprout up from the internet, pop culture, etc many speakers fear the use of CL is becoming less emphasized (Sybesma, 2017) and being replace by a widening extension of *ge*. The researcher notes *ge* to referring to 牛 *niu* ('cow'), has become popularized, leaving its original *tiao*/ *tou* sortal CL in far less usage.

一个牛 yi ge niu

one [gen. CL] cow

We are unable to postulate whether the CL system will or will not faulter, however, it is evident that CL use and preferences are varied among speakers, a trait liked by some. But findings that speakers utilize older or more mundane CL for new objects bring back life to the CL system and strengthen its existence for speakers. In Hong Kong, the use of code-switching or 'English influence' has brought about a new CL: one pack of paper could be said as *'yat pack ji'* (Erbaugh, 2002). Code-switching has been readily seen in Pidgin Chinese and such dialects are more susceptible to number and quantity code switches (Erbaugh, 2002).

2.1.12 Internet Age

The adaptation of the CL system is inevitable as the world exponentially becomes interconnected, but numerous examples help showcase that the system, while varied among dialects, is still vibrant. In the age of the internet, new technology and products are constantly being released and used by the general public, while this is certainly an advantage for humanity, it also poses a linguistic opportunity to look into how these objects come to be carried by certain CL.

On the internet, language boundaries and variations run with such freedom. Zhang (2016) looks at internet language as less constrained by language norms, more abundant vocabulary, and usage specific to online chatting. Nonetheless, it seems the Internet language adheres to its development path by utilizing emoticons, numerical representations of words, and incorporating English and pinyin into text (Zhang, 2016). Unique expressions of language are at the forefront of online Internet language, and its development has interesting characteristics.

There are many ways internet language has adapted and proliferated throughout the modern age as some of the examples above showcase its varying nature. However, as Su (2022) puts it, Internet language is a complex of multiple language styles, combining old and new vocabulary and structures, breaking through the conventional ways to express oneself through language. The notion that internet language combines old and new styles is what is so fascinating. A text message is seemingly simple. Well, how did the internet's mystical concept like a text message come to take on the CL *tiao*? And why did the abstract idea of a text message catch on in spoken Chinese as something that best fits within the salient long and thin CL groups?

一条短信 *yi tiao duanxin*

one [long thin CL] text

But this concept is not new, it happens every time a new invention, object, or abstract concept is discussed. Associating a CL with an object is the internal struggle native speakers must mediate when speaking about an object that is 'not yet quantified' in mainstream language. When text messages were introduced to the public, one can hypothesize there was disagreement as to which CL should become commonplace. Whenever that was resolved is unknown, but the foundation of this process holds. As more technology, innovations, inventions, concepts, and abstracts become daily objects of living or commonly discussed, generally accepted CL arise.

When questioning this practice more questions pop up, naturally. Dictionaries denote CDs having two CL either *zhang* or 片 *pian* (Yellow Bridge) (MDBG). A medication tablet is also carried by *pian* (MDBG). Satellites use the CL 棵 *ke* (MDBG). Rockets use the CL 枚 *mei* (MDBG) and countless more inventions sprung up in recent years that have found unconventional CL to some.

一张光盘 *yi zhang guangpan* One [flat CL] CD	一棵卫星 *yi ke weixing* one [small sphere CL] satellite	一片药片 *yi pian yaopian* one [flat slice CL] medication tablet	一枚火箭 *yi mei huo jian* one [coin, medal CL] rocket

Combined with the knowledge of the CL system genesis and how dialects may differ in CL trends, the system of classification is alive and active in Chinese. A system so complex and intricate that even when faced with new or unclassified objects, native speakers can assimilate these objects into CL groups without restructuring existing nominal bonds. When taking a step back and admiring the entire system of Chinese classification, it is both profound and puzzling. While the answer to all the questions above will not be found in this study, the next section will aim to investigate how Chinese speakers perform and critique certain CL when faced with new inventions and objects that have not come to a formalized CL.

3.0 Methodology

3.1.1 Research Introduction

The purpose of this study is to better understand how Chinese speakers make CL associations when faced with objects, inventions, unconventional animals, etc that have not been classified in common knowledge or frequently used in spoken Chinese. Trends in CL selection and comfortability by native speakers could lead to clues of unspoken or undefined CL rules within the Chinese classification schema.

One motivation for this study was the implicit principles that appear relevant to new inventions such as 一条短信 *yi tiao duanxin* (‘one text message’) which might strike non-native speakers and learners as idiosyncratic, as learners are taught CL tend to have meanings related to objects. As non-native Chinese learners dive deeper into the language, more of these ‘interesting’ CL-nominal pairings pop up. Begging the question, are there common CL trends in specific circumstances? Do native speakers usually select a similar CL? And do regional differences play a role in CL selection?

3.1.2 Participants

Participants were selected by the following inclusion criteria being 18 years or older and a native/ fluent Chinese speaker. The survey solicited participants’ demographic characteristics of age and if they speak any additional Chinese dialects.

3.1.3 Instruments

A single survey collection method was used in this study with an additional optional follow-up chat, conducted by the PI, for brief interest in question response reasoning. Participants were contacted via email and sent a link to complete a survey online via Qualtrics individually. Participants were asked to reach out should any questions or complications arise during the test-taking period.

Forty-four native Chinese speakers completed the task. The survey utilized a variety of question types. After initial demographic questions, a total of 27 questions were used (multiple choice, free-text, ranking, and rating on a Likert scale). The use of different question types and content-presentation methods was in hopes to engage the participant and encourage discrimination of certain CL to aid in analyzing patterns and trends among participants. The survey was subdivided into three sections.

In Section I of the survey, participants were asked to answer 12 multiple-choice questions. The first five questions were relatively easy and participants' responses, which we expected to be relatively consistent, help to ensure the reliability of the survey. After the first five questions, the difficulty of objects to CL began to increase e.g 'My sister's apartment only had three tubes of toothpaste'. A list of commonly used CL or debated CL were listed as options. Participants could also write in responses. The basis for these questions was to gather a baseline and ensure participants' production accuracy.

Next, in Section II participants were asked to use the free-text response fields to write a CL they would use for the given objects, with no prompting of CL choice in this section. Seven new, innovative objects were presented along with a short description so the image could be understood

e.g., a robotic dog, an auto-stabilizing lipstick handheld applicator, virtual reality goggles, a book-like foldable laptop, etc.

Section III asked participants to rank preferences of CL choices for various objects and animals. This section was four questions, aiming to elicit which CL was the most comfortable or natural-sounding from a given list, the top-ranked being the most comfortable/ natural usage. A list of pre-determined CL were given based on common and uncommon usages.

Section IV asked participants to rate if certain CL sounds natural and comfortable to native speakers on a Likert scale of 1-5 stars, 1 being the most odd-sounding and 5 being very natural or common-sounding.

See Appendix A for the complete survey.

3.1.4 Data Collection and Analysis

For control sake, the same set of questions and images were shown to every participant in the survey. Images remained focused on animal and invention classifications. Moreover, to increase the survey validity, scholarly literature was consulted to create the survey questions and answer choices provided to participants (Li & Yan, 2009) (Su, 2019) (Zhang, 2006).

Participants' responses to the first five multiple-choice items in Section I showed more than 90% accuracy. As these questions were used to provide a baseline of participants' responses, high accuracy was expected, and it indicated that our survey items' reliability in reflecting native speakers' linguistic competencies in categorizing different nouns.

Data were examined both collectively and as two subgroups—with Taiwanese speakers compared to other dialect speakers. The results serve to provide insight and guidance in some of the features when CL are chosen by native speakers in new or unfamiliar contexts.

4.0 Results

4.1.1 Section 1: Multiple Choice Questions

Basic demographic information was collected before CL survey questions. Of the 44 participants who completed the survey, the age range breakdown was 31 participants between 18-28 years old, 10 between 28-38 years old, 2 between 38-48 years old, and 1 participant over 48 years old.

As for dialectal differences, 10 participants indicated no other dialects known. 22 participants within the Min dialect families (all indicating Taiwanese). Four participants within the Wu dialects, 2 Hakka, 2 Mandarin families, 1 Yue family, 1 Xiang family, and 1 within the Gan dialect families[1].

Questions one through five were used as warm-up questions for comfortability with the survey platform and were not informed in the following analysis. Specific items in Q1-5 are generally agreed-upon items within Mandarin Chinese with objectively correct answers (e.g., cat = *zhi*, fish = *tiao*, bed = *zhang*).

The following question in Section I introduced harder classifications. The results from Q7-13 show the pairs can be grouped into three subgroups from little to heavy variation of participant agreement.

[1] Of the 44 participants, only 43 responded to the dialect demographic question above.

Table 1 Multiple Choice Questions 7-13

Question	Top Choice	% Agreement	Second Choice	% Agreement	Third Choice	% Agreement	Variation
Q11 Seahorse	只	84%	匹	12%			Little
Q12 Catfish	条	91%	尾	9%			Little
Q7 Wolf	匹	67%	只	23%			Moderate
Q8 Toothpaste	条	56%	支	26%			Moderate
Q13 Crocodiles	只	61%	条	23%			Moderate
Q9 Window Curtains	张	26%	匹/套	19%	面	14%	Large
Q10 Rocking Horse	个	42%	只	30%	匹	21%	Large

From Q7-Q13 we see a variety of CL pattern distributions. Two classifier-noun pairs exhibited little Variation (over 70% agreement among participants): 'seahorse' in primary association with *zhi* and secondary association with *pi*; 'catfish' in primary association with *tiao* and secondary association with *wei*. These two different sea creatures show that participants gravitate to different salient features to classify. *Pi* is likely from the use of 马 *ma* ('horse') in context, although seahorses are not horse related.

In the moderate variation distributions (between 50-70% participant agreement), we see a wider variety of CL agreements. In 'wolves', primary association is with *pi*, secondary association is with *zhi*. *Pi* historically is used with horses, camels, donkeys, cloths, etc and the connection to wolves is unknown but hypotheses are that wolves could be mounted or ridden in ancient times

(MDBG). ‘Toothpaste’ distribution had slightly over half of the participant’s primary association with *tiao*, secondary with 支 *zhi*. Crocodile’s primary association was 只 *zhi*, secondary with *tiao*.

The moderate distributions show an emphasis on shape and animal-based CL or both as top associations. *Tiao*/支 *zhi* as two different conceptualized ways of categorizing toothpaste based on shape, and *tiao* is a shape-related animal classifier. The results confirm how classifier sorting conceptualizations overlap. Recall from above, the separation within the *tiao* group for 支 *zhi* dennotions being long, rigid, cylindrical objects (Tai, 1994). And how from three separate Chinese language books, toothpaste had different CL 支 *zhi, tiao,* and *guan* (Li & Yan, 2009)(Su, 2019)(Zhang, 2006).

Analyzing the Non-Taiwanese dialect participant’s responses shows ‘toothpaste’ at a top association with 支 *zhi* (43%) instead of *tiao*. *Tiao* had a 19% response from all non-Taiwanese dialect speakers, showcasing some regional CL preferences.

In the large variation distributions (less than 50% agreement), we see a vast variety in how participants classified the items including ‘curtains’ and ‘wooden horses.’ One common trait between the two nouns is that those are human-manufactured products created around pre-modern periods. In this aspect, they are in contrast with the animal-denoting nouns which typically exhibited little or moderate classifier variations. These objects are also dissimilar to ‘toothpaste’ because ‘toothpaste’ has a salient tube-like shape. In contrast, fabric curtains are associated with *zhang*, reflecting its flat, thin features similar to ‘a sheet of paper’, 面 *mian* (‘faced surface CL’) in association with the conceptualization that it has a ‘face’ side and ‘backside’, much like a wall or mirror, and 套 *tao* (‘set CL’) typically used for a set of furniture pieces such as ‘sofa’ or ‘furniture’. ‘Curtain’ is associated with *pi*, as that is the canonical classifier for ‘cloth’ or ‘fabrics’ rolled up similar to a scroll (ToneOz). In other words, ‘curtains’ can be conceptualized differently

depending on their shape and feature (flat, thin, and potentially has a 'face' side), their function (furniture), and the material that it is made of (fabric).

Rocking Horses are also very interesting as they had a majority of participants select *ge* the default CL, potentially as an option due to uncertainty. Recall from the literature synthesis that *ge* tends to do this with connections. The second and third most popular CL for Rocking horse are both animal featured CL, 只 *zhi* and *pi*, both in association with the animal horse. In other words, participants observed the salient shape that the toy resembles, while most of them recognized the function of the object as a 'toy,' making *zhi* and *pi* only their secondary choices. The wide variations observed in association with these objects confirm that many objects, especially manufactured products in premodern to modern times, may exhibit many features (e.g., shape, function, material) that correspond to different classifiers, leading to varied choices in classifier usage.

When analyzing participants' responses in terms of their dialectical background, i.e., speakers with Taiwanese backgrounds versus others, no apparent differences were seen between the two groups of participants. However, they had increased usage of *zhi* in 'Rocking Horse', with *zhi* as the top choice at 41% and *ge* at 27%, so effectively reversing the top two choices.

4.1.2 Section II: Free Response Questions

Section II of the survey Free-text response questions are as follows:

Table 2 Free Response Questions 14-20

Response	Q14 Elliptical	Q15 Lipstick Handheld Applicator	Q16 Robotic Dog	Q17 Exoskeleton/ Moving Cart	Q18 Computer Book	Q19 VR Glasses	Q20 Air Purifier
台	20	9	6	11	28	2	29
个	3	18	7	13	8	7	9
副	-	1	-	3	-	27	-
只	-	3	26	-	-	1	-
支	-	7	-	-	-	1	-

In these questions, participants were shown new technological inventions and input in a CL. The results show that 台 *tai* ('machine CL') and *ge* remain popular options among all items. Either *tai/ge* remained in the top results of each item, helping to verify that *ge* is the default CL and that the salient feature of *tai* for machines is continually selected. This helps affirm that as new inventions continue to be produced, *tai/ge* are CL that will continue to be used. This conclusion is the most prominent within this data set.

Additionally, the 'Robotic Dog' primary association was 只 *zhi*. For the 'handheld auto stabilizing lipstick applicator' the tertiary association after *tai/ge* was 支 *zhi*, likely attributing the machine's intended purpose to lipstick itself. The VR headset showed a similar association, while *tai/ge* had a presence, a vast majority of participants took the intended purpose of the headset's glasses-like features to associate with 副 fu ('glasses CL').

When an invention has a prominent, specific appearance or presumed purpose, a greater association is present. The 'Robotic Dog' for animal (只 *zhi*), 'lipstick invention' for lipstick (支 *zhi*), and VR headset for glasses (*fu*).

The contrast to the above trend is seen in 'Air purifier'. As an air purifier can take various shapes and sizes, *tai/ge* were the greatest associations. Perhaps a mundane machine has no greater contextualization to other CL.

4.1.3 Section III: Ranking Questions

The results from the ranking questions are as follows. The first choice indicates the most comfortable/ natural-sounding CL to participants.

Table 3 Ranking Questions 22-26

Question		Most Common First Choice	Second Most Common First Choice	Third Most Common First Choice
Q22	Robotic Dog	个 (36%)	台 (33%)	只 (31%)
Q24	Hoverboard	台 (87%)	个 (14%)	-
Q25	Ocelot	只 (90%)	头 (5%)	-
Q26	Badger	只 (92%)	头 (5%)	-

Section II saw inventions take on *tai/ge* as large contributors to the common CL practices. 'Hoverboards' saw this association above. 'Ocelot' and 'Badger' saw 只 *zhi* as the first association with secondary association to *tou*. Linked to the literature synthesis above, *tou* is common in Southern Min as their animal CL, but in Mandarin *tou* innately relates to livestock, farm animals,

domesticated, and relatively tame animals (Tai, 1994), an interesting association made by some participants.

'Robotic Dog' in this section shows *zhi* as the first association, while in Section III, *zhi* occupied the secondary association behind *tai/ge*. The difference could be due to task variation. It would appear that in ranking tasks or when participants had sufficient time to consider all options, they may pay more attention to the function rather than the shape/appearance of the object.

Analyzing the responses of participants with dialectal differences was not significant in this section. Participants with a Taiwanese-speaking background showed greater cohesion in their collective choices i.e., a greater percentage would form a consensus on CL choices. The top choices above remained true within the entire cohort.

4.1.4 Section IV: Likert Scale Question

The final section, a review of the Likert scale questions for the most comfortable/ natural-sounding CL are as follows:

Table 4 Likert Scale Questions 21, 23, 55, 57

Question	Prompted CL for evaluation	Most Popular Choice	Percentage of Participants
Q21 Auto-stabilizing Lipstick Holder	把	Not natural/1	48%
Q23 Underground Car Subway Cart	列	Not natural/1	45%
Q55 Blob Fish	团	Not Natural/1	50%
Q57 Shrew Mole (Animal)	条	Neutral/3	27.50%

Section IV results show a majority of participants were not comfortable using 把 *ba* (‘handheld CL’) with the ‘handheld lipstick applicator’. Even though the applicator has a handle, the applicator may have other more salient features that render the usage of ba an unlikely choice. Recall that in Section II, participants primarily chose *tai* or *ge* with this noun, following its ‘machine-like’ function and some participants chose ‘支 *zhi’*, in association with the ‘lipstick’ itself. It appears that these characteristics of the applicator override the feature of the ‘handle’ in this case.

The machine/invention trend follows that *tai/ge* would be the most comfortable CL, with 支 *zhi* (the salient feature of its purpose) a second comfortable choice. So *ba*, yes while being ‘a handheld’ salient feature CL, various investigations about the use of *ba* as a CL, show that ‘the handheld-ness’ of an object is to be associated with *ba* when there is a function of that handle, so a doorknob, handle, knife, gun, etc. The use of the grip and immediate action discharged in the hand are innately connected with *ba* use (Li Fang, 2023) (Liao and Li, 2023), making the above invention perhaps not as closely related as originally thought.

The ‘Underground Car Subway Cart’ was pictured as one car cart in a train-like tunnel. Nearly half participants showed no association with 列 *lie* (‘segment, cart CL’), even when pictured in such a way as to correspond to other *lie* linked objects. The Ministry of Education (2021) illustrates *lie* as a CL that denotes some kind of row, column, sets, and arrangements that tend to be counted. While the underground subway car cart could be counted in such a way, participants showed little association with such a scheme, likely due to greater salient features to consider such as 辆 *liang* (‘vehicle CL’), *tai*, etc.

团 *tuan* associated with ‘blob fish’ was rejected by half the participants, likely due to a several factors. The fish’s animate and loose appearance does not elicit the same contextualization

that *tuan* connected nouns do (inanimate, round, moldable). In a solely squishy and bouncy sense, blobfish shape could be extrapolated to loosely fit into a *tuan* rough definition, however, this leap was not primarily endorsed by participants. Zhang (2019) speaks to the general evolution of *tuan* CL use, and how abstractions such as gasses, smells, groupings, clouds, and smoke have become the most common uses. *Tuan* while related to roundness and being a small mass, the salient features of blobfish do not elicit strong associations among participants.

Recall that 'Catfish' seen in Section I, can readily be associated with *tiao*, as they resemble fish archetypes. Extending the 'blob fish' to other CL groups on a subset of physical appearances—such as round, and pliability, does not induce an association among participants.

Lastly, 'mole' shows that *tiao* was a neutral associated among participants. From earlier results, shape-based CL seems to be preferential in animal associations. Categorical associations with *tiao* and 只 *zhi* to animals have remained strong throughout. The neutral association suggests at least a degree of acceptance or little opposition to the shape-based animal CL *tiao*.

Results from the above sections show that the connection between nouns and CL is not arbitrary. Specific salient features of objects are at the forefront of associations. Features that render certain CL suitable and others invalid work to establish trends in native speakers' CL use schema. Facing new inventions and unconventional objects, the Mandarin Chinese classification framework ebbs and flows to accommodate nouns.

5.0 Conclusions and Limitations

Classifiers are hard. There is a consensus on that. When conducting this study, I frequently ran into instances where native Chinese speakers would have to sit and ponder about what CL to use in a given scenario. However, the seemingly instinctual nature that native speakers have to select CL is something so foreign to language learners, no doubt a core curiosity sparking this study.

This study sought to pick out trends, if any, in CL usage and selection for new and exciting objects. Even with a limited study group, and some odd questions, there was indeed trending to be seen and pondered. In a realm of hundreds of CL to choose from, participants' collective preferences help illustrate how native speakers tend to choose CL—not by arbitration but by instantaneous innate categorical analysis of salient features of the object presented.

Interestingly enough, when asked if participants would change their CL after hearing a different one used in context, interviewees consistently said they would not. Object analysis and CL selection are categorically rigid to the schema within one's mind, and when we look at how participants select certain objects, habits and trends become visible.

The results above showcase several fascinating trends within the classification system. First, we notice that shape and animal-based CL remain a sturdy pillar of the CL associations. Among native speakers, we found that so much of the participant’s first and second associations gravitate towards the larger classification groups of shape and animal sortals (i.e., *tiao*, *zhi*). In several follow-up interviews with participants, when asked what features of an object they consider when determining a CL, an astoundingly common response was “shape and size”. Even with the largely varied distributions of objects like curtains, rocking horse, and toothpaste, the disputes

originated from which shape or appearance-based CL to use. Shape, animal-based, appearance, and size seem to be the instinctual CL categories that begin CL discrimination. Note, this trend appears throughout the survey results, but most concisely in Section 1.

Secondly, when faced with a new invention, participants not only verified that *ge* is the default CL, but in addition, *tai* remains categorically correct in each of the inventions presented. There was a particular attraction noted with *tai/ge* as primary and secondary associations. A third association lent itself to connect to the function or 'striking appearance' of the invention. For the lipstick gadget, the third association, 支 *zhi*, is related to its function—lipstick. Robotic Dog took on 只 *zhi* for its unmistakable dog appearance, VR headset with *fu* for glasses, and so on. When we extrapolate further, we can conclude that in the future when new inventions are introduced, the continued use of *tai/ge* will be accepted by the public. Regardless of invention size and weight, *tai* remained an instinctual choice for inventions. Additionally, functionality and object appearance should be considered for various supportive CL associations.

Lastly, a CL does not connect to an arbitrary feature, the association is specific and adaptive per object. While the lipstick holder has a handle, it would not commonly be used with CL *ba*. The same is seen with mugs, doors, tennis rackets, etc all containing handles, but classified in ways different than *ba*.

Blobfish being squishy, round, and with a bouncy appearance could draw a line to *tuan*, but the blobfish's animation and fish-like archetype excludes a *tuan* association with many participants. Concluding that speakers could latch on to a variety of salient features of any object, but rarely does random trait selection endorse enough of a response to elicit a CL association.

The cohesion and general trending of CL usage show how important, reliable, and self-aware participants are of CL usage in modern Chinese. While predicting the future of CL

habitations is not within the scope of a single study, the trends seen with inventions and unconventional objects shed light on how specific features attract CL and what popular associations are enhanced in new circumstances.

Despite the significant findings of the above research, there are a few limitations that warrant consideration. Firstly, while half of the survey participants spoke Taiwanese dialects and we analyzed the data based on the whole participant group and dialect subgroups, we could not report statistically significant comparisons due to the limited number of participants we had. In future studies, it would be beneficial to include representative groups from different dialects to complement the results of this study. The study also utilizes a single survey that was limited to 27 questions. It would be useful for future research to utilize a wider set of objects, inventions, and provide multiple settings to discriminate if CL use is context dependent.

While investigating this topic, as noted in the introduction, I primarily count measure words under the umbrella group of classifiers, following Lobben et al. (2022), as in many contexts and to the language learner, they are often treated the same. As this study is primarily concerned with language usage with objects and new inventions instead of theoretical distinctions, I felt this was an appropriate generalization. Nevertheless, it should be acknowledged that many linguists consider measure words and classifiers to be different (e.g., Her & Hsieh, 2010; also see Cheng & Sybesma, 2012 for relevant theoretical discussions), and future experimental studies can be designed to tease apart the potential usage differences between the two.

Despite these limitations, our study provides valuable insights into the relationships between CLs, inventions, and salient features. Moving forward, there are several promising directions for future research. Tracking CL usage in a longitudinal study could help uncover how CL usage adaptations and patterns manifest over time and shed light on the dynamics of CL

utilization in different contexts or stages of the invention process. Longitudinal data could provide a more comprehensive understanding of the changes in CL usage and their implications for invention outcomes.

Lastly, future research could explore potential mediating and confounding variables that may impact CL selection in various scenarios. Variables such as cognitive processes, social influences, or contextual factors could play a role in the relationship between CLs, inventions, and salient features. Identifying and examining these variables could provide a deeper understanding of the underlying mechanisms that play a role in CL selection and usage.

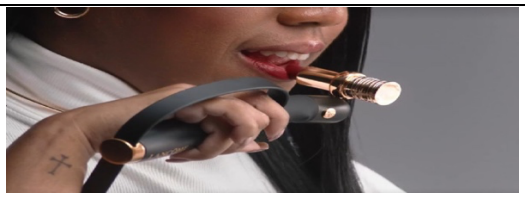
Appendix A Survey

Participants completed the following survey on Qualtrics, an online survey website.

Table 5 Survey

1. I have four fish. 我有四___鱼	Answer Choices: 个 条 只 头 其他
2. There are nine bottles of milk. 有九___牛奶	Answer Choices: 瓶 杯 只 个 其他
3. I will adopt seven cats. 我会收养七___猫	Answer Choices: 个 只 条 头 其他
4. Each room has two beds. 每个房间都有两____床	Answer Choices: 张 套 条 其他
5. At home we have five birds. 在家里，我们有五___鸟。	Answer Choices: 条 头 只 其他
6. End of Control Questions	
7. My dog chased four wolves. 我的狗追着四___狼。	Answer Choices: 个 匹 头 条 只 其他
8. My sister's apartment only had three tubes of toothpaste. 我的姐姐家，只有三 ____ 牙膏。	Answer Choices:条根支管其他
9. If you are going to the store, buy nine window curtains. 你去市场的话，应该买九____窗帘。	Answer Choices: 匹 面 张 套 其他
10. On Christmas I got a wooden horse. 圣诞节时，我收到了一 ___木马。	Answer Choices: 匹 只 架 台 其他
11. Where did the four seahorses go? 四___海马去哪儿？	Answer Choices: 条 只 匹 其他
12. I bought six catfish at the fish market. 在鱼市，我买了六___鲇。	Answer Choices: 条 个 尾 其他
13. The swimming pool had five crocodiles. 这游泳池有五 ___鳄鱼。	Answer Choices: 只 条 头 其他

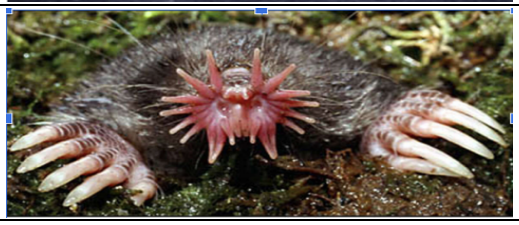
End of Multiple Choice Questions, Begin Free Response Questions

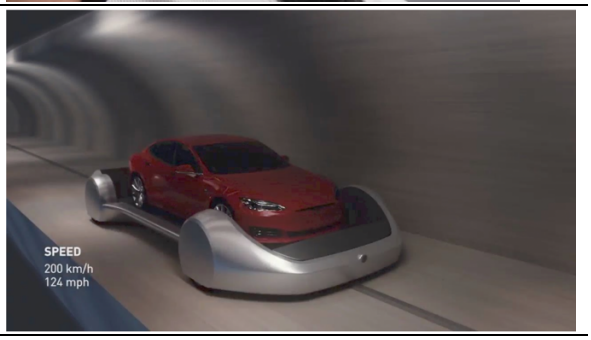
14. 图片：自动自行车 这种东西你会怎么讲？ 请买一 (量词)	 Free Response
15. 图片：自动稳定口红涂抹器这种东西你会怎么讲？请买一 (哪个量词)	 Free Response
16. 图片：机器狗 这种东西你会怎么讲？请买一(哪个量词)	 Free Response
17. 图片：搬箱子的机器外骨骼 东西你会怎么讲？请买一 （哪个量词）	 Free Response
18. 图片：像书一样的电脑 这种东西你会怎么讲？请买一(哪个量词)	 Free Response
19. 图片：虚拟现实眼镜 这种东西你会怎么讲？请买一 (哪个量词)	 Free Response
20. 图片：空气净化器 这种东西你会怎么讲？请买一 (哪个量词)	 Free Response

End of Free Response Questions, Begin Ranking Questions

21. 图片：跟猫很像、跟豹子很像 Rank the following CL in order of most comfortable/natural-sounding. “我拍了一_____Ocelot (奥斯洛特) #1: Most comfortable/natural. #3: Neutral. #5: Least comfortable/natural	 Ranking options: 一条 一个 一只 一头 一口
22. Rank the following CL in order of most comfortable/natural sounding. 我的狗有一_____Robotic Dog 的朋友。 #1: Most comfortable/natural. #3: Neutral. #5: Least comfortable/natural	 Ranking Options: 一只 一台 一条 一个 一匹
26. Rank the following CL in order of most comfortable/natural sounding. “看！这个獾家有三_____獾 ″ #1: Most comfortable/natural. #3: Neutral. #5: Least comfortable/natural	 Ranking Option: 三头 三个 三只 三条 三口
24. Rank the following CL in order of most comfortable/natural sounding. 我弟弟有五 "hoverboards” #1:Most comfortable/natural. #3: Neutral. #5: Least comfortable/natural	 Ranking Options: 五台 五条 五轮 五块 五张

End Ranking, Begin Likert Scale Questions

55. would the following be acceptable if a friend texted this to you: “我钓鱼时，捕鱼过一团鱼，看看 ″ * =not natural ** =Neutral ***** =Very normal.	
57. would the following be acceptable if a friend texted this to you: “在我花园下有两条鼹鼠 ″ *=not natural ** =Neutral ***** =Very normal.	

21. would the following be acceptable if a friend texted this to you: “我看了一把 “自动稳定口红涂抹器” *=not natural ** =Neutral ***** =Very normal.	
23. 图片：在地铁里这是跟雪橇很像会帮车去。would the following be acceptable if a friend texted this to you: "我城市有特别的交通，看一列"cart"会帮你车去啊。 *=not natural ** =Neutral ***** =Very normal.	

End of Survey

Appendix B English Glossary

Classifier

个 (ge)- the default, general CL for a variety of nouns and situations

匹 (pǐ) - a CL for horses or certain types of cloth

条 (tiáo) - a CL for long, narrow objects like rivers, pants, or roads

口 (kǒu) - a CL for family members

根 (gēn) - a CL for long, thin objects like trees or pencils

尾 (wěi) - a CL for certain animals like fish or snakes

领 (lǐng) - a CL for shirts or coats

棵 (kē) - a CL for trees, plants, or some vegetables

只 (zhī) - a CL for certain animals like birds and small mammals

支 (zhī) - a CL for long, thin objects like pens or chopsticks

张 (zhāng) - a CL for flat objects like paper or photographs

块 (kuài) - a CL for currency or some solid objects like cakes or bricks

把 (bǎ) - a CL for things that can be held in one hand like chairs or umbrellas

封 (fēng) - a CL for letters or some types of sealed containers

团 (tuán) - a CL for some round, solid objects like balls or clouds

领 (lîng) - CL for clothes, mats, screens, etc

眼 (yǎn) - a CL for the number of eyes on a person or animal

头 (tóu) - a CL for some animals or for counting people or livestock

粒 (lì) - a CL for small, round objects like beads or pills

座 (zuò) - a CL for large objects like mountains or buildings

杯 (bēi) - a CL for cups or some other vessels

盘 (pán) - a CL for dishes or plates of food

瓶 (píng) - a CL for bottles or other containers with necks

朵 (duǒ) - a CL for flowers or clouds

管 (guân) - a CL for tube shaped objects

盒 (hé) - a CL for boxes or containers of food

壶 (hú) - a CL for teapots or other similar vessels

颇 (pō) - a CL for some objects that are heavy or cumbersome to carry

兜 (dōu) - a CL for pockets or bags

株 (zhū) - a CL for some types of plants or trees

枞 (zōng) - a CL for some types of trees or evergreens

辆 (liàng) - a CL for vehicles like cars or bicycles

声 (shēng) - a CL for sounds or voices

套 (tào) - a CL for sets, collections, etc.

啲 (di1)- Cantonese CL, similar to 些 xie in Mandarin for pluralization

枚 (mo)- Generalized CL in Southern Min

片 (piàn) - CL for slices, thin sheets

Objects

鱼 (yú) - fish

群 (qún) - a group or herd of animals, people, or things

样 (yàng) - style, kind, or type

人 (rén) - person

羌 (qiāng) - a minority ethnic group in China

香蕉 (xiāng jiāo) - banana

床 (chuáng) - bed

路 (lù) - road or path

蛇 (shé) - snake

裤子 (kù zi) - pants

牙膏 (yá gāo) - toothpaste

新闻 (xīn wén) - news

意见 (yì jiàn) - opinion or suggestion

法律 (fǎ lǜ) - law

命令 (mìng lìng) - order or command

线 (xiàn) - thread or line

狗 (gǒu) - dog

狼 (láng) - wolf

狐狸 (hú li) - fox

猫 (māo) - cat

花 (huā) - flower

凳子 (dèng zi) - stool or bench

毛巾 (máo jīn) - towel

米饭 (mǐ fàn) - cooked rice

茶 (chá) - tea

工厂 (gōng chǎng) - factory

考虑 (kǎo lǜ) - to consider or think about

腔调 (qiāng diào) - tone of voice or accent

耳朵 (ěr duo) - ear

灯 (dēng) - lamp or light

答案 (dá àn) - answer or solution

猴子 (hóu zi) - monkey

牡蛎 (mǔ lì) - oyster

树 (shù) - tree

龙 (lóng) - dragon

刀 (dāo) - knife

被 (bèi) - quilt or blanket

铁 (tiě) - iron

棉花 (mián huā) - cotton

针 (zhēn) - needle

猪 (zhū) - pig

牛 (niú) - cow or ox.

这 (zhe) - this

那 (nà) - that

学生 (xuéshēng) - student

生 (shēng) - life

嗯 (èn) - grunting sound

短信 (duânxìn) - text message

光盘 (guāngpán) - CD Disk

卫星 (wèixīng) - satelite

药片 (yàopiàn) - Medication tablet

火箭 (Huǒjiàn) = Rocket.

References

Ahrens, K. (1994). Classifier production in normals and aphasics. *Journal of Chinese Linguistics*, *22*(2), 202-247.

Cheng, L. L. S., & Sybesma, R. (2005). Classifiers in four varieties of Chinese. *Handbook of comparative syntax*, 259-292. DOI:10.1093/oxfordhb/9780195136517.013.0007.

Cheng, L. L.-S., & Sybesma, R. (2012). Classifiers and DP. *Linguistic Inquiry*, *43*(4), 634–650. http://www.jstor.org/stable/23358025.

Erbaugh, M. S. (2002). Classifiers are for specification: Complementary functions for sortal and general classifiers in Cantonese and Mandarin. *Cahiers de linguistique Asie orientale*, *31*(1), 33-69. DOI:10.3406/clao.2002.1602.

Erbaugh, M. S., & Yang, B. (2006). Two General Classifiers in the Shanghai Wu Dialect: A Comparison with Mandarin and Cantonese, *Cahiers de Linguistique Asie Orientale*, *35*(2), 169-207. doi: https://doi.org/10.1163/19606028-90000151.

Frankowsky, M. & Ke, D. (2016). Humanness and Classifiers in Mandarin Chinese: A Corpus Based Study of Anthropocentric Classification. Language and Cognitive Science. 2. 55-67. DOI: 10.17923/LCS201601003.

Gao, W. [高婉瑜] (2014). An Examination of the Netspeak Pronunciation Memes and Transmission-Using the Works of Jhai Nyu Siao Hong [網路語言的語音模因及其傳播-以宅女小紅的作品為例]. *Tamkang Chinese Journal* [淡江中文學報], (30), 277-315. DOI: 10.6187/tkujcl.201406.30-10.

Her, O. S., & Hsieh, C. T. (2010). On the semantic distinction between classifiers and measure words in Chinese. *Language and linguistics*, *11*(3), 527-551.

Kit-Ken, L., & Lu, J. [陸潔娟]. (1997). The grammaticalisation and regrammaticalisation of chinese numeral classifier morphemes [汉语量词词素之语化和再诘化]. *Journal of Chinese Linguistics*, *25*(1), 1–20. http://www.jstor.org/stable/23753978.

Kuo, J. Y. C., Wu, J. S., & Chung, S. C. (2011). Computer Assisted Learning of Chinese Shape Classifiers. 華語文教學研究, 8(2), 99-122. https://doi.org/10.6393/JCLT.201108.0099

LEE, P. (2020). On the semantics of classifier reduplication in Cantonese. *Journal of Linguistics*, 56(4), 701-743. DOI: 10.1017/S0022226720000110.

Li, F. [李芳] (2023). Analysis of the source of the individual quantifier "ba" [个体量词“把”的来源辨析]. *Journal of Henan University (Social Science Edition)* [河南大学学报(社会科学版)], 126-130+155-156. DOI: 10.15991/j.cnki.411028.2023.01.006.

Li, S., & Yan, X. [李硕 、闫翔] (2009). Lesson 15. In Qing Song Ying Yu Yu Fa [轻松英语语法] (pp. 253). *Ke Xue Ji Shu Wen Xian Chu Ban She* [科学技术文献出版社].

Liao Qiaoyun & Li Ruixiu. [廖巧云、李蕊秀](2023). Research on the evolution of the meaning of the quantifier "Ba" in the perspective of recognition [体认视域下量词“把”的词义演变研究]. *Foreign Languages* [外国语文]（01), 20-30. DOI:CNKI:SUN:SCWY.0.2023-01-003.

Liu, S., Mei, H., Williams, A., & Cotterell, R. (2019). On the idiosyncrasies of the Mandarin Chinese classifier system. *arXiv preprint arXiv:1902.10193*.

Lobben, M., Bochynska, A., Eifring, H., & Laeng, B. (2022). Tracking semantic relatedness: Numeral classifiers guide gaze to visual world objects. *PsyArXiv Preprints*. DOI:10.31234/osf.io/9jenr

MDBG English to Chinese dictionary. (n.d.). MDBG Chinese dictionary. Retrieved March 20, 2023.

Shun-chiu, Y., & 游顺钊. (1988). A cognitive approach to the genesis of nominal classifiers as observed in archaic chinese [从认知角度探讨上古汉语各量词的起源]. *Journal of Chinese Linguistics*, *16*(2), 264–277. http://www.jstor.org/stable/23757860.

Stokes, S.F., & So, L.K. (1997). Classifier use by language-disordered and age-matched Cantonese-speaking children. *Asia Pacific journal of speech, language, and hearing, 2*, 83-101.

Su, G. [苏光] (2022). Reflection on the application of network language in the era of new media. [新媒体时代网络语言应用的思考]. *Internet Weekly* [互联网周刊], (24), 92-94. DOI:CNKI:SUN:HLZK.0.2022-24-031.

Su, Y. (2019). Unit 47. In X. Bian (Ed.), Shi Yong Ying Yu Yu Fa Jing Bian (pp. 356-357). Nan Jing: Jiangsu Feng Huang Ke Xue Ji Shu Chu Ban She.

Sybesma, Rint. (2017). Classifiers, Nominal. *Encyclopedia of Chinese language and linguistics*. Vol I, 620-627.

Tai, J. H. (1994). Chinese classifier systems and human categorization. In honor of William S.-Y. Wang: Interdisciplinary studies on language and language change, 479-494.

Tai, J. H.-Y. (1999). Chinese grammar and bio-cognitive bases of human language. *Journal of Chinese Linguistics Monograph Series*, *13*, 130–146. http://www.jstor.org/stable/23826112.

TSE, S., LI, H., & LEUNG, S. (2007). The acquisition of Cantonese classifiers by preschool children in Hong Kong. *Journal of Child Language*, 34(3), 495-517. DOI:10.1017/S0305000906007975.

Wang, Y. [王瑶］(2021). Research on aspect classifiers based on the HSK vocabulary outline and teaching suggestions [基于 HSK 词汇大纲的体类量词研究及教学建议] *Chinese Character Culture* [汉字文化］(20), 10-11. DOI:10.14014/j.cnki.cn11-2597/g2.2021.20.004.

Wang, L. (2008). Historical and dialectal variants of Chinese general classifiers---on the criteria of general classifiers. In *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)* (Vol. 1).

Wiebusch, T. [魏婷兰] (1995). Quantification and qualification: Two competing functions of numeral classifiers in the light of the radical system of the Chinese script [从汉语的部首看汉语量词标量和分类两种功能的并存关系]. *Journal of Chinese Linguistics*, *23*(2), 1-41.

Zhang, Y. [张艳] (2019). A Diachronic Investigation of the Quantifier "Tuan". [量词 “团” 的历时考察]. *Modern Language and Literature* [*现代语文*] (05), 24-29. DOI:CNKI:SUN:YWLY.0.2019-05-005.

Zhang, C. [张春华](2016). A Brief Analysis Of The Lexical Structure Types And Characteristics Of Internet Language. [浅析网路语言的语汇结构类型和特点]. *Northern Literature* [*北方文学*], (20), 74-75. DOI:CNKI:SUN:BFWX.0.2016-20-046.

Zhang, K. (2006). Zao Zhuang Fang Yan Yu Fa Yan Jiu. Ren Min Jiao Yu Chu Ban She. (Original work published 2006) (pp. 55).

匹. Wiktionary. (n.d.). Retrieved March 24, 2023, from https://en.wiktionary.org/wiki/%E5%8C%B9#Readings.

教育部《國語辭典簡編本》. <列> 辭典檢視 - 教育部《國語辭典簡編本》2021. (n.d.). Retrieved April 11, 2023, from https://dict.concised.moe.edu.tw/dictView.jsp?ID=14299&q=1.

澳声通 ToneOZ. (n.d.). 窗帘 - 教育字典 快速核校注音 曉聲通. 拼音、注音 IVS 字型，选音编辑器 - 澳声通 ToneOZ. Retrieved March 24, 2023, from https://toneoz.com/ime/?dq=%E7%AA%97%E5%B8%98&lang=zh-CN&fnt=1.