Existence Justifies Reason: A Data Analysis on Chinese Classifiers Based on Eye Tracking and Transformers

Yu Wang **Emmanuele Chersoni** **Chu-Ren Huang**
Department of Chinese and Bilingual Studies,
The Hong Kong Polytechnic University
janet-yu.wang@connect.polyu.hk, emmanuele.chersoni@polyu.edu.hk,
churen.huang@polyu.edu.hk

Abstract

This study investigates the impact of classifiers on language comprehension using eye-tracking data and the transformer language model. Recent research suggests that classifiers can facilitate the understanding of subsequent nouns. However, quantitative studies exploring the role of classifiers in language comprehension are scarce. By analyzing eye-tracking data from 1.33 million gaze points, we examine the fixation time differences for nouns with and without classifiers. Our findings reveal that words with classifiers have significantly shorter average duration (P value < 0.05) compared to words without classifiers, with an average reduction in fixation time of 20.632%. Additionally, we utilize the transformer language model BERT to predict masked words based on language distributions and sentence context. Through word prediction experiments on a data set of 100,000 segmented and classifier-tagged sentences, we demonstrate that retaining classifiers significantly facilitate the prediction of the transformer language model. Notably, classifiers not only improve accuracy rates for subsequent nouns (2.56 times higher), but also for preceding verbs (1.25 times higher), which is a novel finding not reported in previous research. Moreover, measure words exhibit an unexpected and noteworthy capacity to contribute to prediction, while event classifiers and approximation classifiers offer greater advantages in predicting verb semantics compared to general individual classifiers. This observation suggests that the Chinese classifier system operates as a lexical-semantic system motivated by ontology.

1 Introduction

In languages that incorporate classifiers, the omission of classifiers from a sentence does not significantly impact comprehension. For instance:

1a. 桌上有一张 *(CL)* 纸。
zhuō shàng yǒu yī zhāng (CL) zhǐ.
There is a piece (CL) of paper on the table.

1b. 桌上有一纸。
zhuō shàng yǒu yī zhǐ.
There is a paper on the table.

Despite the absence of the classifier "张" (zhāng - piece), 1b remains intelligible, and the difference in meaning between 1a and 1b is minimal. *Does this mean that classifiers are not useful?*

The answer is no. Recent studies have revealed that classifiers can improve the prediction of subsequent nouns that share a classifier (Kwon et al., 2017; Chou et al., 2014; Srinivasan, 2010). Consider the following example:

2a. 男人喜欢一小瓶 *(CL) [MASK]* 陪自己的哥们。
nánrén xǐhuān yī xiǎo píng (CL) [MASK] péi zìjǐ de gēmen.
Men like a small bottle (CL) of [MASK] to accompany their buddies.

2b. 男人喜欢一小 *[MASK]* 陪自己的哥们。
nánrén xǐhuān yī xiǎo [MASK] péi zìjǐ de gēmen.
Men like a small [MASK] to accompany their buddies.

In 2a, the presence of the classifier "瓶" (píng - bottle) facilitates the inference that the word hidden with [MASK] likely denotes wine or beer. Conversely, in 2b, the absence of the classifier introduces a range of possibilities for the word hidden with [MASK], such as a plate of fries, a cup of coffee, a cigarette, or a bottle of wine, etc. In this case, the presence of the classifier simplifies the prediction process and reduces ambiguity.

1.1 Research Gap

In the past three decades, research on the function of classifiers has predominantly relied on corpora or surveys (Allan, 1977; Tai, 1994; Wu and Bodomo, 2009; Cheng and Sybesma, 2012; Chen et al., 2022; Wang and Walther, 2023). However, with the advancement of Event-Related Potentials (ERPs), several neurological measurement studies have emerged (e.g., Chou et al., 2014; Kwon et al., 2017). These studies consistently demonstrate that the mismatch between classifiers and predicted words influences the N400 effect in the human mind. The N400 effect demonstrated a graded response among the mismatching classifiers, with a smaller effect observed for classifiers that were semantically related to the predicted word compared to classifiers that were semantically unrelated. These findings provide further evidence for the importance of classifiers in processing subsequent nouns.

Previous literature has provided valuable insights into the function of classifiers, but a **research gap** remains in understanding their effects on sentence comprehension.

1. There is a notable lack of **quantitative studies focusing on cognitive aspects** that investigate how the presence or absence of classifiers affects word prediction and comprehension in sentences.

2. There is a lack of research exploring the relationship between the function of classifiers and the **verbs** that govern them.

3a. 我刚刚 *[MASK]* 了一班 *(CL)* 飞机。
wǒ gānggāng [MASK] le yī bān (CL) fēijī.
I just [MASK] a scheduled departure (CL) of plane.

3b. 我刚刚 *[MASK]* 了一飞机。
wǒ gānggāng [MASK] le yī fēijī.
I just [MASK] a plane.

In 3a, the inclusion of the classifier "班" (bān - schedule) provides a strong indication that the word hidden with [MASK] likely represents the action "took". Conversely, in 3b, the omission of the classifier introduces a broader range of possibilities such as "saw a plane", "took a plane" or even "bought a plane", etc. This example illustrates that the presence of the classifier can also impact the comprehension of preceding verbs in the sentence.

3. There is no research on the variations in the influence on word prediction and understanding among **different types (Chen et al., 2022) of Chinese classifiers.** For example:

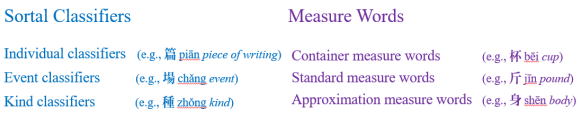

Figure 1: Examples of different types of Chinese classifiers

These classifiers possess distinct meanings, and it would be inappropriate to make a simplistic inference that they have identical semantic influences.

1.2 Research Questions

To address these research gaps, this article investigates the role of Chinese classifiers in sentence processing by employing eye tracking and the transformer language model.

The study employs eye tracking data to provide a quantitative measure of human sentence comprehension, focusing on the cognitive aspect of classifier research. Additionally, a transformer language model is utilized to investigate the machine aspect of classifier research, specifically analyzing big data on contextualized embeddings. The **research questions** are:

1. What is the impact of Chinese classifiers on noun processing in humans?

2. How does the presence or absence of classifiers affect word prediction performance in Chinese sentences?

3. What are the variations in the influence on word prediction among different types of Chinese classifiers?

2 Eye Tracking Experiment

The duration of eye movement reflects the time required for humans to comprehend sentences. And it can be seen as a measure of processing complexity and cognitive load. Our experiment is to compare the fixation time differences for nouns with and without classifiers. In this paper, duration refers to the first fixation duration (Hollenstein et al., 2021), calculated from subtracting onset time from offset time.

2.1 Data

The dataset we use is *The Database of Eye-Movement Measures on Words in Chinese Reading* (Zhang et al., 2022) which contains 1,718 participants, 8,015 Chinese sentences, and nearly 1.4 million fixations. It calculates nine eye-tracking metrics for 8,551 Chinese words. After preprocessing, we obtain a dataset where words are arranged in natural sequences and accompanied by their corresponding duration times. The data example is presented as Figure 2.

Experiment	Sentence_ID	ROI_Beginning	Word_Length	Word_Order	Words	Duration
exp1	1	0	2	1	月蓉	256
exp1	1	2	1	2	和	291
exp1	1	3	2	3	学诚	170
exp1	1	5	2	4	喜欢	142
exp1	1	7	1	5	吃	191
exp1	1	8	1	6	剁	392
exp1	1	9	1	7	椒	342
exp1	1	10	1	8	鱼	180
exp1	1	11	1	9	头	170
exp1	1	12	1	10	，	
exp1	1	13	2	11	昨天	172
exp1	1	15	2	12	聚餐	180
exp1	1	17	2	13	月蓉	142
exp1	1	19	1	14	点	191

Figure 2: Example of the eye-tracking dataset with natural word sequences and corresponding duration

2.2 Methods

To assess the impact of classifiers, we will compare the average duration of words accompa-

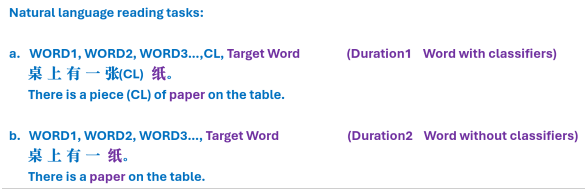

Figure 3: Example of the experiment task

nied by classifiers to the average duration of the same words without classifiers. These measurements will be conducted within the context of natural sentences (Figure 3). Due to the involvement of multiple participants in this dataset, variations in fixation time may occur among individuals. To enhance the reliability of the findings, we focused on words that were accompanied by classifiers at least five times and and calculated the average fixation time.

2.3 Results

From the dataset, we collected data on 29 nouns that appeared after classifiers more than 5 times. Among these nouns, there were a total of 242 fixations recorded on words that appeared after classifiers, while there were 1335 fixations on the same words without classifiers. We conducted statistical analyses to compare the average duration of the 29 nouns with classifiers and the average duration of the same nouns without classifiers. Due to the non-normal distribution of the data and the independent nature of the two groups (as they come from different sentences in natural sequences), we employed the Wilcoxon rank-sum test to determine the significance. With Wilcoxon rank-sum test (statistic: -2.43, P value: 0.015), the analysis revealed that the average duration of a word with classifiers is **significantly shorter (P value $<$ 0.05)** than the average duration of the same word without classifiers. On average, the fixation time for nouns with classifiers is **reduced by 20.632%**.

word	Frequency	Word_length	Word_position_with_classifiers	Word_position_without_classifiers	Word_predictability_with_classifiers	Word_predictability_without_classifiers	With_classifiers_avg_duration	Without_classifiers_avg_duration	Difference
菜	4.17E-05	1	0.628602736	0.662430553	3.77E-10	2.71E-07	199.3846154	269.12	69.73538
小时	0.000282	2	0.183333333	0.694455923	4.98E-10	2.45E-08	176.8	246.3571429	69.55714
药水	2.14E-06	2	0.547562317	0.483333333	5.65E-08	5.04E-07	183.1666667	249.6666667	66.5
图	0.00038	1	0.081333333	0.275891635	2.65E-10	1.48E-08	207.8	273.8	66
领导	0.000347	2	0.06462585	0.237263443	5.60E-09	1.22E-07	194.8571429	256.1780822	61.32094
小说	0.000151	2	0.210365605	0.282278354	2.94E-10	1.73E-08	185.4166667	240.6060606	55.18939
街	9.12E-05	1	0.097402597	0.549206349	3.91E-09	1.85E-08	175.4285714	228.5	53.07143
眼镜	2.24E-05	2	0.474275362	0.212229527	3.37E-08	5.93E-08	176.6	219.6428571	43.04286

Figure 4: Example of results

2.4 Exception Analysis

Despite the overall finding that the average duration of a word with classifiers is significantly shorter, there are certain exceptions where nouns exhibit longer fixation times when accompanied by classifiers.

时间	shíjiān	time
事情	shìqing	thing
电脑	diànnǎo	computer
画	huà	painting
电视剧	diànshìjù	drama
皮毛	pímáo	limited knowledge
业务员	yèwùyuán	salesman
画家	huàjiā	painter

We conducted a comprehensive analysis of all exceptions present in the dataset and found that they can be categorized into two distinct situations.

Word	Sentence_ID	Sentence
电脑	5049	那 台 电脑 换 了 显示器 之后 才 更 好 用 一些
Computer	5049	That computer works better after changing the monitor.
电视剧	6936	在 这 部 电视剧 中 出演 正面 形象 是 他 本人 的 意思
TV drama	6936	In this TV drama playing a positive role means it is him in person.
画	2774	这 幅 画 将 游泳 运动员 在 竞赛 中 勇往直前 的 精神 表现 得 淋漓尽致 。
Painting	2774	This painting vividly portrays the spirit of swimmers striving forward in the competition.
时间	364	这 段 时间 南方 各 省 洪水 泛滥 ， 所以 政府 当前 需要 调动 一切 力量 积极 应对
Time	364	Due to the flooding in various provinces in the south during this period, the government currently needs to m
时间	4898	使用 一 段 时间 之后 手机 开始 变 得 比较 缓慢
Time	4898	After using it for a while, the phone starts to become slower.
事情	4983	报喜 不 报忧 早已 成 了 当今 中国 官场 中 平常 的 一 件 事情 。
Matter	4983	Reporting good news but not bad news has long become a common practice in today's Chinese officialdom.
业务员	3094	这 位 业务员 如此 催促 付款 让 人 觉得 十分 厌恶
Salesman	3094	This salesperson's constant urging for payment is extremely repulsive.

Figure 5: Example of exceptions for longer fixation time with classifiers

1. The nouns are “时间” (shíjiān - time) or “事情” (shìqing - thing/event) which are light nouns. They do not have too much

semantic content. People have to combine aforementioned contexts to understand them. It takes more fixation time.

2. These nouns are found in the structure of "这/那 (zhè/nà - this/that - demonstrative pronouns) + Classifiers + Noun" at the beginning of a sentence. People require additional time to determine the referent of this demonstrative structure. For instance, when encountering the phrase "这部电视剧" (zhè bù diànshìjù - this drama), individuals need time to retrieve the specific TV drama in the mind that this structure refers to (Figure 5).

The two situations can be accounted for by a single factor, namely co-reference resolution. By employing co-reference resolution, it is possible to provide a potential explanation for the observed exceptions. The processing of language by individuals is a complex process involving multiple factors. On the one hand, the inclusion of classifiers can reduce the fixation time. On the other hand, the task of resolving co-reference requires additional fixation time for individuals to comprehend. Consequently, this results in an overall increase in fixation time.

This phenomenon is intriguing. Given the constraints of the current study, investigating the intricate system of human language processing and exploring the multitude of factors involved, as well as their interactions, will be the subject of our future investigations.

2.5 Summary

The eye tracking experiment suggests that classifiers play a beneficial role in language processing for humans, aiding in the comprehension of language. On average, the fixation time for nouns with classifiers is reduced by 20.632%.

3 Transformer Language Model Experiment

To investigate the impact of classifiers on word prediction performance at a large scale, we employ the transformer (Vaswani et al., 2017) language model, specifically BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019), to predict masked words. BERT was selected due to its bidirectional nature and training methodology using masked language modeling, which aligns perfectly with our experiment's objective of predicting masked words. Our experiment aims at comparing the performance of BERT when the input includes classifiers versus when it does not.

It is worth noting that a previous research paper, (Järnfors et al., 2021), focused on how BERT performs in choosing classifiers based on contextual cues. However, our research differs from their primary focus: our research specifically focuses on assessing the impact of classifiers on word prediction performance.

3.1 Data

The dataset we use is the *Chinese Classifier Dataset* (Peinelt et al., 2017) which contains more than 100 million sentences illustrating the usage of Chinese classifiers, sourced from three language corpora: the Mandarin Lancaster Corpus, the UCLA Written Chinese Corpus, and the Leiden Weibo Corpus. The data has been cleaned and processed for context-based classifier prediction tasks.

After preprocessing the data, we obtain the experimental dataset consisting of masked sentences with target word and classifiers.

id	tagged	Classifier
LWC_3407254	以防 今晚 上 失眠 , 下午 一 <CL> <h>人</h> 在 市中心 压 马路 , 从 崇安寺 到 南禅寺 , 然后 再 从 南禅寺 到 保利 广场	个
LWC_3401766	从 前 有 一 <CL> <h>神鲸</h> , 后来 它 病 了 。	条
LWC_3406272	补补补 , 五 <CL> <h>男士</h>	位
LWC_3405410	哈哈打 1 <CL> <h>钱</h> 赢 了 25 有 一 铺 4 炸	块
LWC_3405099	上传 了 17 <CL> <h>照片</h> 到 相册 猫咪希	张

Figure 6: Example of Chinese Classifier Dataset

Input_sentences	Target_word	Classifiers
以防 今晚 上 失眠 , 下午 一 个 [MASK] 在 市中心 压 马路 , 从 崇安寺 到 南禅寺 , 然后 再 从 南禅寺 到 保利 广场	人	个
从 前 有 一 条 [MASK] , 后来 它 病 了 。	神鲸	条
补补补 , 五 位 [MASK]	男士	位
哈哈打 1 块 [MASK] 赢 了 25 有 一 铺 4 炸	钱	块
上传 了 17 张 [MASK] 到 相册 猫咪希	照片	张

Figure 7: Example of preprocessed experimental dataset

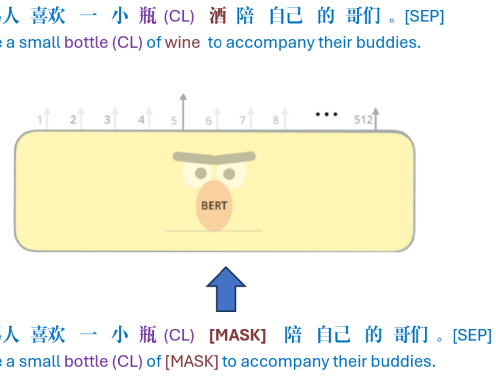

Figure 8: Example of the prediction experiment

3.2 Methods

We conducted two experiments using word-level BERT (wobert_chinese_base) to predict subsequent nouns and preceding verbs in two different environments: one with classifiers and one without classifiers (they are the same sentences with the only difference being a classifier). The input for both experiments is the masked sentence, and the output is the predicted word for the masked position.

To analyze the results, we compare accuracy, perplexity of the predictions, and semantic similarity between the words predicted by BERT when using the MASK token and the target word. The accuracy measures how often the predicted word matches the target word, while the perplexity measures how well the language model predicts the target word given the context, and the semantic similarity assesses the closeness of meaning between the two words.

3.3 Results

Through word prediction experiments on a dataset containing 100,000 segmented and classifier-tagged sentences, we demonstrate that among 100,000 sentences retaining the classifier the semantic similarity between predicted words and target words are **significantly (Wilcoxon signed rank test, P value $<$ 0.01) higher** compared to cases without the classifier. We chose the Wilcoxon signed rank test due to the non-normal distribution of the data and the paired nature of the observations. Additionally, we observed that the prediction accuracy and average perplexity also improved with the presence of the classifier.

3.3.1 Results for Predicting Nouns

Notably, the prediction accuracy for nouns following a classifier is **2.56 times** higher than for nouns without a classifier (49.89% vs. 19.48%). Furthermore, the average perplexity when using classifiers (5.43) is significantly lower than the perplexity without classifiers (7.01) (lower perplexity indicate more certain in its prediction). Additionally, when utilizing classifiers, among 100,000 sentences, we observe a significantly higher (Wilcoxon signed rank test, P value $<$ 0.01) semantic similarity between the words predicted by BERT using the MASK token and the target words, compared to scenarios without classifiers. On average, the semantic similarity scores were 0.9642 and 0.9396, respectively.

Nevertheless, it is intriguing to note that **measure words display an unexpected and noteworthy capacity to contribute to prediction**, despite being traditionally as-

Masked_sentence_with_classifiers	Target_word	Predicted_word	Semantic_similarity
为什么 买 件 [MASK] 这么 难？	衣服	衣服	1
Why is it so difficult to buy piece of [MASK]?	clothes	clothes	
虽然 还 是 惯例 的 三 首 [MASK] , 但是 图片 是 RIF 还 蛮 欣慰 的 喂	歌	歌	1
Although it is still the usual three *shou* (CL for songs or poems) [MASK], the pic	songs	songs	
只要 你 一 句 [MASK] 绝对 好使	话	话	1
As long as you one *ju* (CL for words) [MASK], it will definitely work.	sentence	sentence	
Masked_sentence_without_classifiers	**Target_word**	**Predicted_word**	**Semantic_similarity**
为什么 买 [MASK] 这么 难？	衣服	彩	0.9257905
Why is it so difficult to buy [MASK]?	clothes	color	
虽然 还 是 惯例 的 三 [MASK] , 但是 图片 是 RIF 还 蛮 欣慰 的 喂	歌	刀	0.952944
Although it is still the usual three [MASK], the pictures are quite satisfying for RI	songs	knives	
只要 你 一 [MASK] 绝对 好使	话	看	0.9614661
As long as you [MASK], it will definitely work.	sentence	see	

Figure 9: Example of noun prediction results

sociated with counting or measuring rather than classifying, as suggested by scholarly discourse. This is a phenomenon that has not yet been well discussed within scholarly literature.

Classifiers	Count	Average_similarity_of_nouns_with_classifiers	Average_similarity_of_nouns_without_classifiers	Accurate_times_with_classifiers	Accurate_times_without_classifiers	Ratio
本	52	0.673326993	0.650225342	29	6	4.83333
张	250	0.658654584	0.622359962	137	31	4.41935
辆	25	0.114775417	0.097644558	11	3	3.66667
块	81	0.61164624	0.592084689	43	12	3.58333
篇	96	0.702991492	0.693028392	21	6	3.5
对	28	0.328785524	0.314990641	7	2	3.5
根	27	0.455741	0.432371189	10	3	3.33333
杯	42	0.450428754	0.431866974	13	4	3.25
把	39	0.764373253	0.749302692	16	5	3.2
句	175	0.789781876	0.770397667	117	38	3.07895
片	61	0.367503204	0.330527925	9	3	3
滴	70	0.506533647	0.498603992	20	7	2.85714
部	123	0.472460536	0.449089391	54	20	2.7
只	112	0.507577223	0.495307459	34	13	2.61538

Figure 10: Example of classifier performance in noun prediction

For instance, the container measure word “杯” (bēi - glass) exhibits a substantial influence on noun prediction. In the presence of “杯”, the accuracy of predictions amounts to 13, whereas in its absence, the accuracy diminishes to merely 4, resulting in a noteworthy ratio of 3.25. This dataset encompasses various other measure words that similarly exert a profound impact on enhancing noun prediction, such as “片” (piàn - piece) and “滴” (dī - drop), among others.

3.3.2 Results for Predicting Verbs

Regarding the prediction of verbs preceding classifiers, our analysis reveals that among 100,000 sentences, retaining the classifier results in a **significant increase (Wilcoxon signed rank test, P value < 0.01)** in the semantic similarity between predicted words and target words. And on average, the semantic similarity scores were 0.961 and 0.957, respectively. Furthermore, the prediction accuracy is also higher when classifiers are retained, reaching **1.25 times** that of cases without the classifier (54.36% vs. 43.42%). Additionally, the average perplexity when classifiers are employed (4.27) is notably lower compared to cases without classifiers (5.34).

Moreover, it is worth noting that event classifiers and approximation measure words offer greater advantages in verb semantic prediction as compared to general individual classifiers.

Classifiers	Count	Average_similarity_of_nouns_with_classifiers	Average_similarity_of_nouns_without_classifi	Accurate_times_with_classifiers	Accurate_times_without_classifiers	Ratio
类	23	0.9785815	0.973797815	12	7	1.71429
根	27	0.981827506	0.970862137	15	9	1.66667
片	61	0.821293035	0.781302772	36	25	1.44
堆	42	0.936900002	0.93249495	24	17	1.41176
代	25	0.897755793	0.891291617	14	10	1.4
对	28	0.983012111	0.974674852	18	13	1.38462
只	112	0.852935576	0.831411248	57	42	1.35714
本	52	0.741218687	0.738914916	19	14	1.35714
场	127	0.883301424	0.876822687	65	48	1.35417
号	94	0.873416343	0.868755335	47	35	1.34286
部	123	0.894776404	0.921760185	55	41	1.34146
滴	70	0.862182606	0.884515384	32	24	1.33333
碗	31	0.915016388	0.908512439	16	12	1.33333
套	30	0.903569024	0.965584084	12	9	1.33333

Figure 11: Example of classifier performance in verb prediction

For instance, among the top 10 classifiers that exhibit a significant influence on verb prediction, 7 of them are event classifiers or approximation measure words. These particular classifiers play a crucial role in accurately predicting verbs.

3.4 Exception Analysis

Exceptions occur when the meaning of classifiers is highly generalized and abstract, lacking sufficient semantic information. Out of the 10,000 sentences analyzed, it was found that in 775 sentences where classifiers were present, the prediction performance for nouns was worse compared to sentences without classifiers. Similarly, in 613 sentences with classifiers, the prediction performance for verbs was worse compared to sentences without classifiers. Notably, the three main

Noun Exceptions 775/10000	
Classifiers	Worse performance
个	461
种	51
次	39
场	12
张	11
件	11
位	11
段	9
片	7
点	7

Figure 12: Example of exceptions for worse performance in noun prediction with classifiers

Verb Exceptions 613/10000	
Classifiers	Worse performance
个	277
种	46
次	40
句	14
张	14
部	11
篇	10
首	10
件	10
场	8

Figure 13: Example of exceptions for worse performance in verb prediction with classifiers

classifiers associated with these exceptions are "个" (gè), "种" (zhǒng), and "次" (cì). These classifiers are typical general classifiers that lack specific semantic information.

个 gè — a classifier used to count individuals or objects

种 zhǒng — a classifier used to count types, kinds, or species

次 cì — a classifier used to count occurrences of an action or event

3.5 Summary

After conducting large-scale investigations, the results indicate that the inclusion of classifiers can facilitate the prediction of transformer language models not only on subsequent nouns but also on preceding verbs.

4 Discussions

This section will address two key questions that require further discussion.

1. *Why can Chinese classifiers affect word prediction performance not only for nouns but also verbs?*

2. *Why measure words also display an unexpected and noteworthy capacity to contribute to prediction?*

One possible explanation is that the Chinese classifier system operates as a **lexical-semantic system that is ontologically motivated (Chen et al., 2022)**. It takes into consideration not only the perceptual attributes of objects but also factors such as functionality, event structure, and other factors (formal, constitutive, telic, participant, event participating, descriptive, and agentive (Huang, 2013)).

For instance, the classifier "件" (jiàn - a piece) exemplifies the formal aspect of the noun "衣服" (yīfu - clothes), while the classifier "班" (bān - schedule) represents the event participating aspect of the noun "飞机" (fēijī - plane). This is why we are able to predict phrases like "买件 [MASK]" (buy a piece of [MASK]) and "[MASK] 一班飞机" ([MASK] a flight) with the help of classifiers.

The underlying ontology of classifiers assists us in language processing.

5 Conclusion

This paper investigates the role of Chinese classifiers in sentence processing through the utilization of eye tracking and the transformer language model. The study reveals three primary findings:

1. Classifiers have a beneficial impact on language processing for humans. On average, the fixation time for nouns with classifiers is reduced by 20.632%.

2. Classifiers can facilitate the prediction of transformer language models, not only

for subsequent nouns (with a 2.56 times higher accuracy rate) but also for preceding verbs (with a 1.25 times higher accuracy rate).

3. Different types of classifiers exhibit varying abilities to improve prediction. For example, measure words exhibit an unexpected and noteworthy capacity to contribute to prediction; event classifiers demonstrate better performance than typical individual classifiers when predicting verbs, etc.

6 Limitations

Despite the promising results obtained in our data analysis, there are still several areas that require further improvement.

1. This paper utilizes eye tracking and transformer language models, but the integration between them is still superficial.

2. Language processing in humans is a complex system involving multiple factors such as classifiers and co-reference resolution. It is important to qualitatively investigate the impact of different factors on language comprehension and their interplay.

3. While this paper explains classifiers theoretically, it lacks exploration in practical applications, such as designing a system for learning classifiers and using classifiers to detect cognitive impairments and related disorders, etc.

7 Ethics Statement

We affirm our commitment to contributing positively to society, prioritizing the avoidance of harm, and maintaining honesty and trustworthiness in our work. We do not anticipate any significant risks associated with our research. All experiments conducted in this study were based on publicly available datasets.

References

Keith Allan. 1977. Classifiers. *Language*, 53(2):285–311. Publisher: Linguistic Society of America.

I-Hsuan Chen, Kathleen Ahrens, and Chu-Ren Huang. 2022. The Chinese Classifier System as a Lexical-semantic System. In Chu-Ren Huang, Yen-Hwei Lin, and I-Hsuan Chen, editors, *The Cambridge Handbook of Chinese Linguistics*, 1 edition, pages 550–577. Cambridge University Press.

Lisa Lai-Shen Cheng and Rint Sybesma. 2012. Classifiers and DP. *Linguistic inquiry*, 43(4):634–650. Publisher: MIT press.

Chia-Ju Chou, Hsu-Wen Huang, Chia-Lin Lee, and Chia-Ying Lee. 2014. Effects of semantic constraint and cloze probability on Chinese classifier-noun agreement. *Journal of Neurolinguistics*, 31:42–54. Publisher: Elsevier.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv:1810.04805 [cs].

Nora Hollenstein, Emmanuele Chersoni, Cassandra L. Jacobs, Yohei Oseki, Laurent Prévot, and Enrico Santus. 2021. CMCL 2021 Shared Task on Eye-Tracking Prediction. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 72–78. Association for Computational Linguistics.

Chu-Ren Huang. 2013. On the functional consistency of 'de'. In *Contemporary Studies on Chinese Grammar at the Frontier of Science*, pages 129–135, Beijing. Commercial Press.

Jani Järnfors, Guanyi Chen, Kees van Deemter, and Rint Sybesma. 2021. Using BERT for choosing classifiers in Mandarin. In *Proceedings of the 14th International Conference on Natural Language Generation*, pages 172–176, Aberdeen, Scotland, UK. Association for Computational Linguistics.

Nayoung Kwon, Patrick Sturt, and Pan Liu. 2017. Predicting semantic features in Chinese: Evidence from ERPs. *Cognition*, 166:433–446. Publisher: Elsevier.

Nicole Peinelt, Maria Liakata, and Shu-Kai Hsieh. 2017. ClassifierGuesser: A context-based classifier prediction system for chinese language learners. In *Proceedings of the IJCNLP 2017, System Demonstrations*, pages 41–44.

Mahesh Srinivasan. 2010. Do classifiers predict differences in cognitive processing? A study of nominal classification in Mandarin Chinese. *Language and Cognition*, 2(2):177–190. Publisher: Cambridge University Press.

James HY Tai. 1994. Chinese classifier systems and human categorization. *In honor of William S.-Y. Wang: Interdisciplinary studies on language and language change*, pages 479–494. Publisher: Pyramid Press Taipei.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, \Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yamei Wang and Géraldine Walther. 2023. Measure words are measurably different from sortal classifiers. In *Proceedings of the Seventh International Conference on Dependency Linguistics (Depling, GURT/SyntaxFest 2023)*, pages 81–90.

Yicheng Wu and Adams Bodomo. 2009. Classifiers\ne determiners. *Linguistic Inquiry*, 40(3):487–503. Publisher: MIT press.

Guangyao Zhang, Panpan Yao, Guojie Ma, Jingwen Wang, Junyi Zhou, Linjieqiong Huang, Pingping Xu, Lijing Chen, Songlin Chen, and Junjuan Gu. 2022. The database of eye-movement measures on words in Chinese reading. *Scientific Data*, 9(1):411. Publisher: Nature Publishing Group UK London.