Information Locality in the Processing of Classifier-Noun Dependencies in Mandarin Chinese

Hailin Hao (hailinha@usc.edu)
Department of Linguistics, University of Southern California
Los Angeles, CA 90089-1693 USA

Yang Yang (yangyanggw@gdufs.edu.cn)
Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies
Guangzhou, 510420, China

Michael Hahn (mhahn@lst.uni-saarland.de)
Department of Language Science and Technology, Saarland University
66123 Saarbruecken, Germany

Abstract

In this paper, we report three reading time (RT) experiments (one using self-paced reading and two using A-Maze) that tested the cognitive mechanisms underlying the processing of classifier-noun dependencies in Mandarin Chinese (MC). We leveraged prenominal relative clauses and the contrast between general and specific classifiers in MC, which offered a good testing ground for existing theories of sentence processing. Results from the A-Maze experiments showed both locality and expectation effects. More importantly, we observed an interaction between locality and expectation in the way of *Information Locality* (Futrell, 2019; Futrell, Gibson, & Levy, 2020): Expectation-driven facilitation was highly constrained by locality effects. To capture the results, we implemented a resource-rational Lossy-Context Surprisal model (Hahn et al., 2022) for MC, which successfully replicated the key patterns in the A-Maze experiments.

Keywords: sentence processing; locality; probabilistic expectations; computational modeling; Mandarin Chinese; classifiers

Introduction

A key goal of psycholinguistic research is to characterize the processing difficulty of a word (or any other linguistic unit) in context (e.g., Miller & Chomsky, 1963; Frazier & Fodor, 1978). *Expectation*-driven facilitation (Hale, 2001; Levy, 2008) and *locality*-driven retrieval difficulty (Gibson, 1998, 2000; Lewis & Vasishth, 2005) have been established as two crucial factors in determining incremental sentence processing difficulty. The Hale and Levy expectation-based account (also referred to as Surprisal Theory) holds that the processing difficulty of a word w is proportional to the negative logarithm of its probability given its preceding context c, as formulated in (1). The theory has received ample empirical support from both controlled experimentation (e.g., Jäger et al., 2015; Levy & Keller, 2013; Linzen & Jaeger, 2016; Vasishth & Drenhaus, 2011) and reading time (RT) corpora (e.g., Shain et al., 2022; Smith & Levy, 2013; Wilcox et al., 2023).

(1) processing difficulty $\propto -\log P(w|c)$

One key prediction of Surprisal theory is anti-locality effects, whereby increased distance between two co-dependents leads to faster processing time. For example, completing the subject-verb dependency between **administrator** and **drove** should be easier in (2b) than in (2a), as the expectations for the verb **drove** are shaper in (2b). Such effects have been observed in various languages like Hindi (Husain, Vasishth, & Srinivasan, 2014), German (Konieczny, 2000; Levy & Keller, 2013), Japanese (Nakatani & Gibson, 2010) and Chinese (Lin, 2011).

(2) a. The **administrator drove home.**
b. The **administrator** who lived in the suburb **drove** home.

By contrast, locality-based accounts predict increased cost of completing a dependency when two co-dependents are farther away from each other. Such accounts will thus predict that "met" is harder process in (2b) than in (2a). The explanation is that in order to integrate **drove**, one must retrieve the subject **administrator** from working memory, and as the linear distance between the two co-dependents increases, the memory representation of the first co-dependent is weakened as a result of decay (Gibson, 1998) or interference (Lewis & Vasishth, 2005), leading to more retrieval difficulty. Locality effects have been widely documented as well (e.g., Bartek et al., 2011; Demberg & Keller, 2008; Grodner and Gibson, 2005; Shain et al., 2016).

However, despite their empirical success, expectation and locality effects have mostly been investigated separately in previous work, and how they interact with each other remains under-studied. To build a complete theory of sentence processing, both factors, and their potential interaction, must be taken into consideration. Our current work aims to shed light on this matter using classifier-noun dependencies in Mandarin Chinese (MC), which allow us to test the predictions of expectation-based and locality-based theories, as well as their interaction. In the upcoming part of this section, we will begin by introducing Lossy-Context Surprisal (Futrell, Gibson, & Levy, 2020), a recently

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

proposed theory that offers a distinct prediction regarding the interaction between expectation and locality, and discuss prior empirical research. Following that, we will elucidate the rationale behind our current work and provide a review of earlier investigations into the processing of classifier-noun dependencies in MC. We will also outline the predictions of each theory.

Lossy-Context Surprisal and Information Locality

LCS (Futrell, Gibson, & Levy, 2020) was proposed to address the lack of theory unifying expectation and locality. It differs from the standard Surprisal Theory, which holds that processing difficulty is derived from expectations over a perfectly retained memory representation of context. By contrast, LCS acknowledges that processing is not only expectation-based, but also constrained by memory limitations (Gibson, 1998; Lewis & Vasishth, 2005). Therefore, processing difficulty is determined by expectations derived not from veridical context but from probabilistic inference over imperfect memory representations of the context, as formulated in (3). The model could explain both expectation and locality effects: Words are easy to process when they are easy to predict, as suggested by expectation-based models, and in the meantime if the relevant contextual information is not well-preserved in memory, it can lead to incorrect anticipation of upcoming words, resulting in processing difficulty, as predicted by locality-based accounts. Distinctively, it additionally predicts an interaction between expectation and locality, dubbed as the *information locality* effects: Expectation-driven facilitations are weakened under strong locality constraints. Imagine the word pair 'doctor' and 'diagnosed' in 'the doctor diagnosed the patient', where 'diagnosed' should enjoy an expectation-driven facilitation from 'doctor'. However, if the word pair is separated by extra linguistic material, such as a relative clause, the memory representation of 'doctor' may have undergone distortion by the time 'diagnosed' processed, leading to less sharp expectations.

(3) $$\text{processing difficulty} \propto -\log P(w|c') = -\log \sum_c P(w|c) P(c|c')$$

However, the predicted information locality effects have not been empirically borne out. By contrast, some studies have found the opposite patterns, whereby expectation-driven facilitations play an even bigger role under strong locality constraints. For example, Husain, Vasishth, & Srinivasan (2014) manipulated the predictability of the verb given a preceding noun, and the distance between the verb and the noun. The authors observed locality effects in low expectation conditions, but anti-locality effects in high expectation conditions. They explained the lack of locality effects in the high expectation conditions via the *strength-of-expectation* hypothesis: If a word is highly predictable given a preceding word, it could already be pre-activated and integrated when the preceding word is processed; this way, there will not be any retrieval cost later. However, the robustness of this hypothesis is also unclear. Schwab, Xiang, and Liu (2022) observed similar results using relative clauses in German, but Safavi, Husain, and Vasishth (2016) did not find such an interaction with noun-verb constructions in Persian (see Ming & Wang, in prep, for a similar conclusion based on data from *wh*-dependencies in MC).

Current Study

Considering the theoretical significance of understanding the interplay between expectation and locality and the mixed and limited empirical findings, further testing is necessary. In our current work, we used classifier-noun dependencies in MC as an empirical testing ground. In MC, a noun must be preceded by a classifier in certain contexts. While different nouns are compatible with different *specific* classifiers, there is a *general* classifier GE that almost all countable nouns can take. Although the two options can be used interchangeably (Ma, 2015), a specific classifier render it easier to predict the upcoming noun. This contrast allowed us to create varying levels of expectation. We in addition manipulated the distance between the two classifiers and nouns by inserting prenominal relative clauses that modify the noun. These manipulations allow us to test the main effects of expectation and locality, as well as their interactions.

Another notable gap in the existing literature is the absence of direct testing of the cognitive mechanisms involved in processing classifier-noun dependencies in MC (and other languages) using RT measures. The majority of previous work has examined how classifier information guides the predictions of upcoming noun using neurological measures (e.g., Chan, 2019; Chou et al., 2014; Hsu et al., 2014; Qian & Garnsey, 2016) and visual world eye-tracking (e.g., Grüter, Lau, & Ling, 2019; Lobben et al., 2023). The RT signature of classifier-noun dependencies therefore remains unclear. Some other work has used RT measures to investigate classifier-noun dependencies, but the research questions usually focus on how (temporarily mismatching) classifiers facilitate the processing of relative clauses (e.g., Tang, Nelson, & Tollan, 2024; Wu, Kaiser, & Vasishth, 2017). Therefore, further work is required for a thorough understanding of the processing mechanisms of classifier-noun dependencies.

To preview, we will present three reading experiments. Experiment 1 was conducted in the lab using self-paced reading (SPR). Experiment 2 was conducted online using A-Maze (Boyce, Futrell, & Levy, 2020). Experiment 3 was an in-lab replication of Experiment 2. Our results in general provided support for main effects of expectation and locality, as well as information locality effects.

Predictions

Expectation-based accounts should predict a speedup at the noun when a specific classifier is used. It may additionally predict an anti-locality effect, whereby increased distance between the two co-dependents leads to a speedup. A preliminary analysis using the Chinese part of the Universal Dependencies (UD; Nivre et al., 2016) supports this

prediction. Out of 2,245 classifier-noun dependencies, the noun directly appears after the classifier for 781 times (34.79%). For 219 times, there is a relative clause following the classifier, and among these, the noun directly appears after the classifier for 114 times (52.05%). Therefore, the probability of a noun coming up is higher when there is an intervening relative clause. By contrast, locality-based account should predict locality effects, whereby increased distance leads to higher processing difficulty. Regarding the interaction effects, LCS predicts that the facilitation from specific classifiers should be weakened as distance increases. The strength-of-prediction hypothesis, by contrast, predicts the opposite pattern of interaction, whereby locality effects are attenuated in specific classifier conditions.

Experiment 1: In-Lab SPR

Methods

Participants Ninety-six students who self-identified as native speakers of MC from Guangdong University of Foreign Studies took part in the study. Their participation was financially compensated.

Material & Design We crossed Classifier (Specific vs. General) and Distance (Local vs. 1RC vs. 2RC) in a 2X3 design, leading to 6 conditions. An example set of stimuli is in (4). In all examples, GE is the general classifier, while ZHANG is a specific classifier that matches with 'zhuozi' (desk). In (4a, b), the classifier and the noun are adjacent to each other. In (4c, d), they are separated by one relative clause. We added a passivizer 'bei' between the classifier and the relative-clause-internal noun 'Anna' to avoid a temporarily misleading parse. In (4e, f), one additional relative clause is inserted, which modifies 'Anna'. For all conditions, the critical region (CR) is the noun 'zhuozi'. We created 36 target items in total, along with 54 filler items.

(4) a. General Classifier; Local
Mali tingshuo na-liang-GE zhuozi duzhu-le lu.
Mary hear that-two- GE desk block-PERF road.

b. Specific Classifier; Local
Mali tingshuo na-liang-ZHANG zhuozi duzhu-le
Mary hear that-two-ZHANG desk block-PERF
lu.
road.
'Mary heard that those two desks blocked the road.'

c. General Classifier; 1RC
Mali tingshuo na-liang-GE bei Anna nuozou de
Mary hear that-two-GE PASS Anna move REL
zhuozi du-le lu.
desk block-PERF road.

d. Specific Classifier; 1RC
Mali tingshuo na-liang-ZHANG bei Anna nuozou
Mary hear that-two-ZHANG PASS Anna move
de zhuozi du-le lu.
REL desk block-PERF road.
'Mary heard that those two desks that LiuNa moved blocked the road.'

e. General Classifier; 2RC
Mali tingshuo na-liang-GE bei culude ganzou-le
Mary hear that-two-GE PASS rudely chase-PERF
Katie de Anna nuozou de zhuozi du-le lu.
Katie REL Anna move Rel desk block-PER road.

f. Specific Classifier; 2RC
Mali tingshuo na-liang-ZHANG bei culude
Mary hear that-two-ZHANG PASS rudely
ganzou-le Katie de Anna nuozou de zhuozi
chase-PERF Katie REL Anna move Rel desk
du-le lu.
block-PER road.
'Mary heard that those two desks that Anna that rudely chased away Katie moved blocked the road.'

Procedure The study was conducted in a sound-proof lab in-person. We used the non-cumulative self-paced moving window method (Just, Carpenter, & Woolley, 1982). Stimuli were presented using PCIbex (Zehr & Schwarz, 2018). The Latin square design ensured that each participant saw each item in only one condition. The target items and fillers were pseudo-randomized for each participant. Out of 2/3 of the stimuli, participants had to answer a binary-choice comprehension question, to make sure that they paid attention to the experiment.

Analysis

Exclusion All participants achieved more than 80% accuracy on the comprehension questions. We excluded RTs that are either below 50ms or above 5000ms.

Model Structure For statistical analysis, we fitted Bayesian linear mixed effects regression models on log-transformed RTs at the critical word region (i.e., the noun) using the brms package, version 2.12 (Bürkner, 2017) in R. Relatively uninformative priors were chosen, which allowed for a plausible yet wide range of RTs and effect sizes in either direction. The factor Classifier was contrast coded, where Specific is coded as -1 and General as 1. For Distance, we applied successive differences coding (Venables & Ripley, 2002), as in Table 1. This coding schema allows us to compare one level with the means of the other two levels. When a reliable interaction was observed, we would conduct a follow-up nested analysis.

Table 1: Coding for the factor Distance.

	Distance 2-1	Distance 3-2
Local	0.33	0.67
1RC	0.33	-0.33
2RC	-0.67	-0.33

Results

Results of Experiment 1 are plotted in Figure 1. At CR, we observed locality effects. The 2RC conditions were read more slowly than the mean of 1RC and local conditions (β=0.112, CrI=[0.071, 0.154]). The local conditions were also read faster than the mean of 1RC and 2RC conditions (β=-0.097, CrI=[-0.146, -0.048]). Surprising, we did not observe any reliable effects of Classifier (β=-0.008, CrI=[-0.022, 0.005]). There was in addition an interaction between Distance and Classifier, whereby the effects of the Classifier became stronger in the 2RC conditions than in the 1RC and local conditions (β=0.035, CrI=[-0.022, 0.005]). Follow-up analysis suggested that Classifier only had a reliable effect on RTs in the 2RC conditions (β=-0.033, CrI=[-0.060, -0.005]), but not in the local and 2RC conditions (β=0.004, CrI=[-0.012, 0.020]).

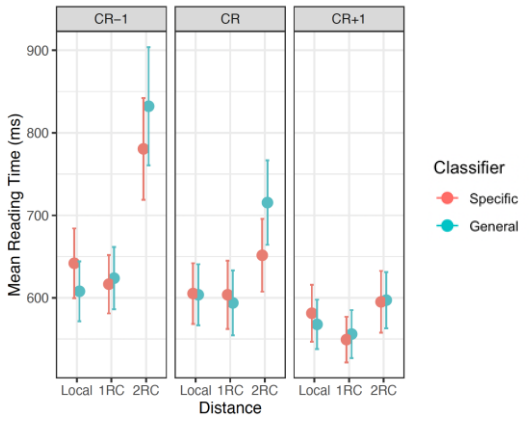

Figure 1: Results of Experiment 1 at CR, and one region before and after CR (error bars represent +-1SE).

Interim Discussion

We are cautious in interpreting the results, due to the lack of any expectation-driven facilitation from the use of specific classifiers in the local and 1RC conditions. This is very surprising considering the pervasiveness of predictive processing and stands in contrast to previous work on classifier-noun dependencies using ERPs or visual world paradigms. We consider two potential explanations here. The first is that processing the agreement between a specific classifier and its dependent noun requires a conscious checking, meaning that comprehenders need to spend extra resources to make sure that the noun is compatible with the classifier, as pointed out by Zhang & Zhou (2019), who found that a noun following a specific classifier is processed even more slowly. However, this proposal is incompatible with the facilitations we discovered in the 2RC conditions. Moreover, it is hard to reconcile this account with other work on predictive processing, which usually found facilitation. A second explanation is that the noun regions suffer from spillover effects from specific classifiers. Spillover effects are well documented in the sentence processing literature, whereby the effect of certain manipulation does not show up locally but only after the critical region (e.g., Wagers, Lau & Phillips, 2009; Smith & Levy, 2013). In our case, the specific classifiers might pose more processing difficulty than the general classifier as they are less frequent and morphological more complex, and this difficulty might have caused spillover effects in the noun region in the local and 1RC conditions, masking any expectation-driven facilitation. It also explains why facilitation is only observed in the 2RC conditions.

To investigate the second possibility, we conducted a replication of Experiment 1 using A-Maze (Boyce, Futrell, & Levy, 2020), an alternative to SPR, which has been shown to be more robust to noise, and especially, spillover effects (Forster, Guerrera, & Elliot, 2009; Witzel, Witzel, & Forster 2012; Boyce, Futrell, & Levy 2020).

Experiment 2: Online A-Maze

Methods

Participants Eighty self-identified native speakers of MC were recruited online. Their participation was financially compensated.

Material & Design We adopted the same design and material from Experiment 1.

Procedure The study was conducted online using A-Maze (Boyce, Futrell, & Levy, 2020), hosted by PCIbex (Zehr & Schwarz, 2018). As in other Maze tasks, participants' word-by-word RTs were recorded as they choose between the correct and an incorrect continuation, as illustrated in Figure 2. When the wrong continuation was selected, participants were prompted by an error message telling them to try again (with a penalty of 1000ms). In A-Maze, wrong continuations are automatically selected by a neural network language model. Following Levison et al. (2023), we used multilingual BERT (Devlin et al., 2019) to generate incorrect continuations that have very low contextual probability but match the correct ones in terms of frequency and then hand-corrected them.

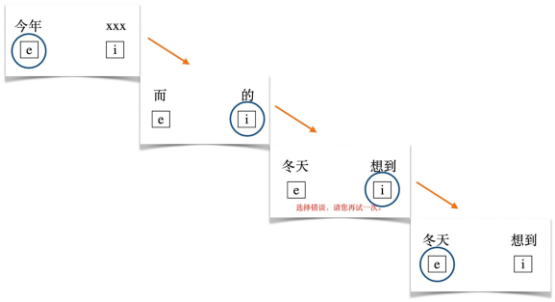

Figure 2: An illustration of the Maze task in Chinese.

Analysis

Exclusion We excluded nine participants whose overall A-Maze accuracy rate fell under 80%. We excluded RTs that are either below 50ms or above 5000ms. Additionally, we excluded trials where participants made a mistake at the critical word or before the critical word.

Model Structure The model structure and coding scheme was the same as in Experiment 1, except that the priors of the Bayesian model were adjusted slightly to account for the fact that reading takes longer in A-Maze than in SPR.

Results

Results of Experiment 2 are plotted in Figure 3. The model estimates showed a negative effect of Classifier (β=-0.086, CrI=[-0.110, -0.061]), suggesting that specific classifiers facilitated the processing of the subsequent noun, consistent with expectation-based accounts. We also observed locality effects. The 2RC conditions were read more slowly than the mean of 1RC and local conditions (β=0.111, CrI=[0.071, 0.152]). The local conditions were also read faster than the mean of 1RC and 2RC conditions (β=-0.086, CrI=[-0.110, -0.061]). Finally, there was also a reliable interaction between Distance and Classifier, whereby that the facilitatory effects of a specific classifier became weaker in the 2RC conditions in comparison to in the Local and 1RC conditions (β=0.058, CrI=[0.018, 0.098]), consistent with information locality effects.

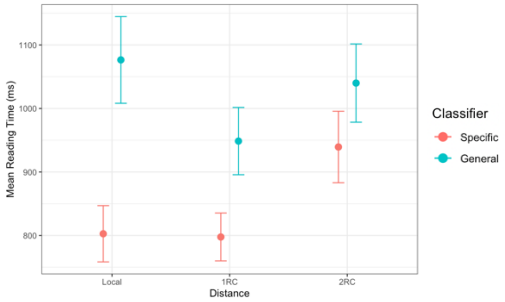

Figure 3: Results of Experiment 2 at CR (error bars represent +-1SE).

Nevertheless, follow-up analyses show negative effect of Classifier in the 2RC condition as well (β=-0.052, CrI=[-0.085, -0.020]), but it is of smaller magnitude compared to the effect Classifier in the local and 1RC conditions (β=-0.104, CrI=[-0.131, -0.077]). Moreover, in the local and 1RC conditions, there was also a negative effect of Distance, whereby the 1RC conditions was read faster than the local conditions. The main effect of Distance was modulated by Classifier, whereby its main effect was mostly driven by the general classifier conditions (this is confirmed in another follow-up analysis). That is to say, there was an anti-locality effect in the general classifier conditions between the local and the 1RC conditions. We argue that this interaction is consistent with information locality as well, as the difference between the general and specific conditions became smaller in the 1RC condition. We will discuss how the anti-locality effects in the general classifier conditions can be subsumed under information locality.

Interim Discussion

The use of A-Maze in Experiment 2 successfully avoided spillover from specific classifiers, and instead we observed facilitation from them, as predicted by expectation-based theories. We also found robust locality effects, whose magnitude was similar as in Experiment 1. More importantly, we found evidence for information locality effects: Expectation-driven facilitation from specific classifiers became weaker in 2RC conditions compared to the local and 1RC conditions.

Analysis on Local & 1RC Conditions When analysis was conducted on the local and 1RC conditions only, there was an interaction between Distance and Classifier, whereby increasing distance in the general classifier conditions led to speedups, but not in the specific classifier conditions. We argue that this might be due to the fact the intervening material in the 1RC conditions can provide extra information to predict the upcoming noun. This extra piece of information is especially helpful in the general classifier conditions, considering that virtually any noun can appear after it. The possible set of nouns that can come up would be drastically reduced when after the relative clause information is processed. By contrast, in the specific classifier conditions, the information that the relative clause provides does not provide extra information in predicting the noun. This still aligns with the information locality hypothesis, which states that words that highly predict each other should be kept close (e.g., a specific classifier and its associated noun). Another way to put it is that word pairs with low mutual information (e.g., a general classifier and its associated noun) suffer less from locality constraints, as the extra information may provide more useful cues for the prediction of the noun.

Experiment 3: In-Lab A-Maze

In Experiment 3, we did a replication of Experiment 2 in the lab. A successful replication can strengthen our confidence in the results. Also, we hope to make sure that the difference between the results of the first two experiments was not due to the in-person vs. online setup.

Methods

Participants Ninety-six students who self-identified as native speakers of MC from Guangdong University of Foreign Studies took part in the study. Their participation was financially compensated.

Material & Design We adopted the same design and material from Experiments 1 and 2.

Procedure The same procedure was used as in Experiment 2, except that this experiment was conducted in the lab.

Analysis

Exclusion All participants achieved higher than 80% accuracy for A-Maze choices. The same RT exclusion criteria as in Experiment 2 were adopted.

Model Structure The model structure and coding scheme was the same as in Experiment 2.

Results

Results of Experiment 3 are plotted in Figure 4. We successfully replicated the major patterns of Experiment 2. First, there was evidence that specific classifiers facilitated the processing of the subsequent noun (β=-0.095, CrI=[-0.111, -0.078]). We also observed locality effects. The 2RC conditions were read more slowly than the mean of 1RC and local conditions (β=0.120, CrI=[0.085, 0.157]). The local conditions were also read faster than the mean of 1RC and 2RC conditions (β=-0.102, CrI=[-0.156, -0.049]). Finally, there was also a reliable interaction between Distance and Classifier, whereby that the facilitatory effects of a specific classifier became weaker in the 2RC conditions in comparison to in the local and 1RC conditions (β=0.078, CrI=[0.042, 0.116]), consistent with information locality effects. Follow-up analysis showed the same patterns as well. Moreover, the magnitude of the effects across the two studies was very similar.

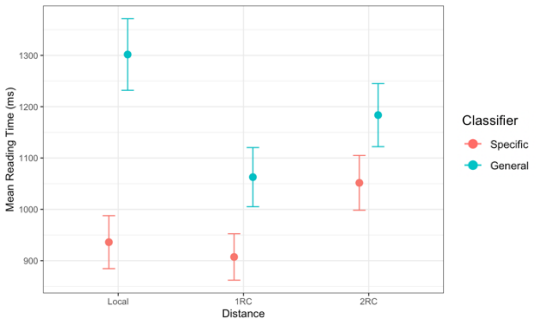

Figure 4: Results of Experiment 3 at CR (error bars represent +-1SE).

Interim Discussion

Experiment 3 replicated the results of Experiment 1 in the lab. It excluded the possibility that the differences among the first two experiments resulted different experimental setups and strengthened our confidence in interpreting the results.

Computational Modeling

Finally, we present a resource-rational LCS (RR-LCS) model (Hahn et al., 2022) for MC. Recall that in LCS, a word's processing difficulty is determined via probabilistic inferences over imperfect contexts. However, the model allows for great flexibility with regard to which components of the preceding context might be prone to memory loss (i.e., it is up to the modeler to choose a noise model). To remedy this, Hahn et al. (2022) proposed a resource-rational (Lieder & Griffiths, 2019) version of LCS (RR-LCS), which offered a principled way for approximating lossy context representations, based on the hypothesis that memory representations are optimized to minimize expected downstream processing effort given limited cognitive resources. The model calculates a retention probability for each past word, which was optimized to minimize the average model surprisal over large-scale text data, while constraining the overall fraction of deleted words among a context window of 20 words. We created an MC version of the model using large-scale MC corpus data from a variety of genres and sources (20GB of text). We then estimated $P(w|c)$ using a Chinese GPT-2 model (Zhao et al., 2019). Results of the RR-LCS model surprisal are plotted in Figure 5. Applying this model to our stimulus set, it successfully captured (i) expectation effects, whereby specific classifier conditions have lower surprisal (i.e., less processing efforts); (ii) locality effects, whereby longer-distance conditions have higher surprisal (i.e., more processing efforts); and (3) information locality effects, whereby the surprisal differences between the two types of classifiers become smaller under stronger locality constraints, since the representation of classifiers is more likely to get deleted as distance increases. This tendency is strengthened as the rate of deleted words gets higher (i.e., even higher memory limitations).

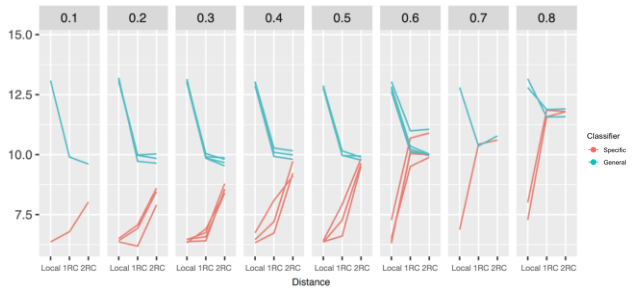

Figure 5: Model surprisal calculated from Chinese GPT2 (top panel shows different deletion rates).

Conclusion

In this study, we set out to use classifier-noun dependencies in Mandarin Chinese as an empirical testing ground for sentence processing theories. We investigated both expectation and locality effects, and more importantly, how they interact with other. Our A-Maze results showed both expectation-driven facilitation and locality-driven processing difficulty, and a previously undocumented interaction between the two factors: Expectation-driven facilitation effects from two words that highly predict each other are weakened if they are separated in linear order. Our newly implemented RR-LCS model successfully captured these results. Overall, we show that probabilistic expectations are constrained by memory limitations and that future theory building in sentence processing should take this into consideration. However, we do note that expectation and locality can interact differently in different languages, and that individual's working memory limitation may also play a role. We leave these inquiries for future work.

Acknowledgement

Y.Y. acknowledges support from a major project (22JJD740020) from the key research base of the Chinese Ministiry of Education, CLAL, GDUFS. M.H. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project-ID 232722074 — SFB/CRC 1102.

References

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In Search of On-Line Locality Effects in Sentence Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1178–1198.

Boyce, V., & Levy, R. (2023). A-maze of natural stories: Comprehension and Surprisal in the maze task. *Glossa Psycholinguistics, 2*(1).

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language, 111*, 1-13.

Chan, S. (2019). An elephant needs a head but a horse does not: An ERP study of classifier-noun agreement in Mandarin. *Journal of Neurolinguistics, 52*, 100852.

Chou, C.-J., Huang, H.-W., Lee, C.-L., & Lee, C.-Y. (2014). Effects of semantic constraint and cloze probability on Chinese classifier-noun agreement. *Journal of Neurolinguistics, 31*, 42–54.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition, 109*(2), 193–210.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods, 41* (1), 163–171.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science, 44*(3).

Gibson, E. (1998). Linguistic complexity: Locality of Syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, Bbain: Papers from the first mind articulation project symposium* (pp. 95–126). Cambridge, MA: MIT Press.

Grodner, D., & Gibson, E. (2005). Some consequences of the serial nature of linguistic input. *Cognitive Science, 29*(2), 261–290.

Grüter, T., Lau, E., & Ling, W. (2019). How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience, 35*(2), 221–234.

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *PLoS ONE, 9*(7).

Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, *119*(43).

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies 2001 - NAACL '01*.

Hsu, C.-C., Tsai, S.-H., Yang, C.-L., & Chen, J.-Y. (2014). Processing classifier–noun agreement in a long distance: An ERP study on Mandarin Chinese. *Brain and Language, 137*, 14–28.

Jäger, L., Chen, Z., Li, Q., Lin, C.-J. C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language, 79*(80), 97–120.

Just M. A., Carpenter P. A., & Woolley J. D. (1982) Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General 111*(2): 228–238.

Konieczny, L. (2000). Locality and Parsing Complexity. *Journal of Psycholinguistic Research, 29*(6), 628-645.

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences, 43*, e1.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177.

Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language, 68*(2), 199–222.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*(3), 375–419.

Lin, Y. W. (2011). Locality versus Anti-locality Effects in Mandarin Sentence Comprehension. In J.-S. Zhuo (Ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)*, Volume 1 (pp. 200-214). Eugene: University of Oregon.

Linzen, T., & Jaeger, T. F. (2015). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science, 40*(6), 1382–1411.

Lobben, M., Bochynska, A., Eifring, H., & Laeng, B. (2023). Tracking semantic relatedness: Numeral classifiers guide gaze to visual world objects. *Frontiers in Language Sciences*, 2.

Ma, A. (2015). *Hanyu geti liangci de chansheng yu fazhan [The Emergence and Development of Chinese Individual Classifiers].* China Social Sciences Press.

Nakatani, K., & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: Evidence from Japanese. *Linguistics, 46*(1), 63–87.

Nivre, J., de Marneffe, M-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., & Zeman, D. (2016). Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 1659–1666). Portoroz, Slovenia: European Language Resources Association.

Qian, Z., & Garnsey, S. (2016). An ERP study of the processing of Mandarin classifiers. Integrating Chinese *Linguistic Research and Language Teaching and Learning*, 59–80.

Safavi, M. S., Husain, S., & Vasishth, S. (2016). Dependency resolution difficulty increases with distance in Persian separable complex predicates: Evidence for expectation and memory-based accounts. *Frontiers in Psychology*, 7.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*(3), 302–319.

Shain, C., Meister, C., Pimentel, T., Cotterell, R., & Levy, R. (2022). Large-scale evidence for logarithmic effects of word predictability on reading time. *PsyArXiv*.

Shain, C., van Schijndel, M., Futrell, R., Gibson, E., & Schuler, W. (2016). Memory access during incremental sentence processing causes reading time latency. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*.

Schwab, J., Xiang, M., & Liu, M. (2022). Antilocality effect without head-final dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(3), 446–463.

Tang, X., Nelson, P., & Tollan, R. (2024). Effects of classifier (mis-)match on filler-gap dependencies in Mandarin. *Proceedings of the Linguistic Society of America*.

Vasishth, S., & Drenhaus, H. (2011). Locality in German. *Dialogue & Discourse, 2*(1), 59–82.

Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S. Springer.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*(2), 206–237.

Wilcox, E. G., Pimentel, T., Meister, C., Cotterell, R., & Levy, R. (2023). Testing the Predictions of Surprisal Theory in 11 Languages. *Transactions of the Association for Computational Linguistics*.

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research, 41* (2), 105–128.

Wu, F., Kaiser, E., & Vasishth, S. (2017). Effects of early cues on the processing of Chinese relative clauses: Evidence for experience-based theories. *Cognitive Science, 42*(S4), 1101–1133.

Xiang, M., & Wang, S. (Under Revision). Locality and expectation in Chinese wh-in-situ dependencies.

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832

Zhao, Z., Chen, H., Zhang, J., Zhao, X., Liu, T., Lu, W., Chen, X., Deng, H., Ju, Q., & Du, X. (2019). UER: An Open-Source Toolkit for Pre-training Models. *EMNLP-IJCNLP 2019*, 241.

Zhang, Y. & Zhou, P. (2019) Prediction or processing burden: Online processing of classifier and noun-noun compound in Mandarin. *The 32nd Annual CUNY Conference on Human Sentence Processing*, Boulder, CO.