Is There Evidence for Embodied Simulation During Language Translation?

Yuhan Fu

Department of Cognitive Science, UC San Diego

Honors Program 2024

Advisor: Dr. Sean Trott

Abstract

Embodied simulation has been a subject of extensive study in efforts to understand language comprehension. This theory posits that when we process language, we mentally simulate the experiences and actions described. Remarkably, the neurons that fire during this mental simulation overlap with those activated during actual experiences or perceptions (Bergen, 2012). Several studies have investigated whether embodied simulation is engaged during language tasks, such as judging the plausibility of sentences. There is also evidence suggesting that embodied simulation occurs during second language processing. Our research aims to explore this simulation process using different sentence tasks. Specifically, we examined whether embodied simulation is engaged and facilitates the language translation process. Our results indicate a trend toward a match effect for both plausibility and translation tasks. Overall, recall accuracy and memory strength were higher for the translation task than for the plausibility task. These findings suggest that similar cognitive processes may be used for different language tasks, and that the translation group had a deeper impression of the sentences they processed. Although the match effect was not statistically significant, it implies that embodied simulation is likely engaged in the language translation process and aids comprehension.

Keywords: embodied simulation, second language processing, language comprehension, language translation, plausibility judgment, match effect, memory strength

Is There Evidence for Embodied Simulation During Language Translation?

Embodied simulation is a prominent theory in language comprehension, positing that when we process language, we mentally simulate the experiences and actions described. Bergen (2012) expands on this by introducing the concept of "simulation semantics," where understanding language involves re-enacting perceptual and motor experiences. This theory bridges the gap between abstract symbols and real-world experiences, suggesting that our brains use the same neural circuits for both actions and language comprehension. Bergen's research, along with other cognitive scientists, indicates that these simulations are not just epiphenomenal but play a functional role in understanding meaning, helping us to bridge the gap between linguistic input and our experiential knowledge of the world (Bergen, 2012). Supporting this idea, other studies have shown that overlapping neural tissues are activated both when we perform an action (such as walking) and when we process words or concepts related to that action (such as the word "walk") (Gallese, 2008; Tian et al., 2020). This suggests that our understanding of language is deeply rooted in our sensorimotor experiences, which are reactivated when we encounter related linguistic stimuli.

One crucial aspect of studying embodied simulation is concept representation, which refers to how our brain encodes and processes the meaning of different concepts. Pecher et al. (2009) conducted research that supports the notion of embodied simulation during language comprehension. Their study demonstrated that comprehenders activate visual features, specifically implied orientation and shape, during language comprehension. Participants were better at recognizing pictures if the object's orientation or shape matched what was implied by a preceding sentence. This effect persisted both immediately after reading and after a 45-minute delay, suggesting that these sensorimotor simulations are retained over longer periods (Pecher et al., 2009).

Some studies have also proposed that simulation is involved in second language comprehension. Wheeler and Stojanovic (2006) explored whether non-native English speakers engage in mental simulation during language processing. Their study utilized an image-verb forced-choice matching task and found that non-native speakers, like native speakers, showed longer response times when the actions depicted in images and verbs involved the same effect (e.g., hand, mouth, foot) compared to different effectors. This finding suggests that non-native speakers perform mental imagery similar to native speakers, and this capability increases with linguistic competence. They concluded that as non-native speakers become more proficient, their processing strategies shift from translation-based to simulation-based, mirroring the processing strategies of native speakers (Wheeler & Stojanovic, 2006).

Given this background, we understand that embodied simulation facilitates language comprehension, but what about other language processes, such as language translation? Many studies have used plausibility judgment to study embodied simulation, but translation has not been studied before. Therefore, we designed a study that builds upon the findings by Pecher et al. (2009). Our research aims to investigate whether comprehenders exhibit evidence of simulation when engaged in the task of translating from one language to another. Specifically, we aim to explore whether bilingual individuals mentally represent the implied visual characteristics during their translation process and, if so, to what extent they engage in simulation. Alternatively, we examine if they bypass this simulation process, perhaps relying on a word-by-word translation approach (as in machine translation), without engaging in embodied simulation.

Our study has two research questions. The first is whether there is a simulation effect during the translation process. Our null hypothesis is that there is no simulation effect when people translate from one language to another. If this is true, then no match effect will be observed in the translation group, and the memory strength for both match and mismatch pictures will be similar. Alternatively, participants in the translation task group may also exhibit the match effect. This effect is characterized by a tendency to choose the picture that matches during the picture judgment phase, and to demonstrate higher memory strength for pictures that align with the sentence-implied shape. The second research question builds on the first, asking: if there is a simulation effect during language translation, to what degree do they simulate? Our null hypothesis for this question is that there is no difference in the degree of simulation for translation and plausibility groups. If this holds true, then both groups should demonstrate the same amount of simulation, participants in the translation group should perform as well as the plausibility group, suggesting both tasks may employ similar cognitive processes. Alternatively, the match effect observed in the translation group could be greater or smaller compared to that of the plausibility group. This difference suggests that during translation, people may simulate to a greater or lesser extent than in a general language comprehension task. The greater or smaller match effect might be explained by the translation mechanisms they used. A smaller effect may result if excessive cognitive load is used in the translation process, therefore possibly reducing the capability for embodied simulation. A larger effect might result from a deeper processing of sentence meaning and more thorough engagement with the material. Other possible mechanisms such as language proficiency and translator's strategies, could also be factors that enhance or diminish the embodied simulation effect.

Method

Participants

A total of 149 participants who were fluent in both English and Mandarin were initially recruited for the study. However, 59 participants were excluded during the study: 33 due to the incompleteness of the study, and 26 due to low accuracy or negative memory strength (indicating an inability to distinguish between critical items and fillers). The remaining participants (N = 90) were recruited through the UCSD SONA system, and those who completed the study received SONA credit, which could be used as extra credit for certain classes.

Of the 90 participants, 69 identified as female, 16 as male, 2 as nonbinary, 1 did not specify their gender, and 2 preferred not to disclose their gender. The age range of the participants was 18 to 33 years (M = 20.9, SD = 2.34). In terms of ethnicity, 91.1% (n = 82) identified as Asian, 1.1% (n = 2) as Caucasian or White, 1.1% (n = 1) as Hispanic or Latino, and 4.4% did not specify their ethnicity. Regarding Chinese proficiency, 64.4% (n = 58) reported their proficiency as native-level, 11.1% (n = 10) as advanced, 13.3% (n = 12) as intermediate, and 11.1% (n = 10) as beginner. The average LexTALE score was 70.3% (SD = 13.4, range: 40-100), indicating a good level of English proficiency.

Study Design and Materials

This study includes two independent variables, each with two levels. The first variable is task type, with two conditions: translation and plausibility. This variable will be manipulated between subjects. The second variable is picture type comprising match and mismatch conditions, which will be manipulated within subjects. The dependent variable is recall accuracy of the pictures. We will measure this both by analyzing the task accuracy as well as *d'*, the measure used in Pecher et al. (2009).

In this study, 60 sentences implying the shapes of objects will be used. Half of these sentences (30) imply shape A, and the other half (30) imply shape B. For example:

a. There was a cucumber in the fridge.
b. There was cucumber in the salad.

These sentences imply either a whole cucumber (see the left image below) or a sliced cucumber (see the right image below).

Each sentence corresponds to two picture versions (match and mismatch). The assignment of these four conditions (Shape A match, Shape A mismatch, Shape B match, Shape B mismatch) will be counterbalanced and presented in random order. Each participant will only view one picture (match or mismatch) corresponding with one sentence (shape A or shape B). For the translation tasks, participants will be presented with sentences in English and asked to translate them into Chinese. A set of 60 nonsense sentences will serve as fillers in the plausibility task. Additionally, participants will complete the LexTALE test at the end of the study to assess their English proficiency.

In our experiment, we utilized materials similar to those in the study by Pecher, D. et al. (2009). Following their design, each of our quadruplets consisted of two sentences and two black-and-white pictures. In this setup, one picture matched the implied shape of the object in one sentence, while the other picture corresponded to the implied shape in the alternate sentence. We excluded the measurement of implied object orientation to focus specifically on task differences. Additionally, we retained the design of Pecher, D. et al.'s (2009) plausibility experiment, and created our filled sentence sets. This methodology allows for a direct comparison of the degree of simulation in translation versus plausibility tasks, while maintaining the fundamental design elements of the original study. The LexTALE test will be administered via the LexTALE website.

Procedure

Each participant will complete the study in one sitting, which will be divided into four phases: the sentence task (either translation or plausibility tasks), the recall task (involving the judgment of matching and mismatching pictures), the LexTALE test, and the demographic phase. Upon reading and agreeing to the informed consent, each participant will begin with the sentence task. Depending on their assigned group, participants will be tasked with either translating sentences or judging their plausibility. For the translation condition, 60 sentences will be presented, while the plausibility condition will involve 120 sentences. Immediately after, participants will be informed of a surprise memory task. In this task, they will judge whether the shape of an object matches the sentence they saw in the previous task. A total of 120 pictures (60 critical and 60 fillers) will be used for this judgment. Subsequently, participants will be asked to complete the LexTALE test and report their scores. Finally, they will be requested to answer some demographic questions.

Results

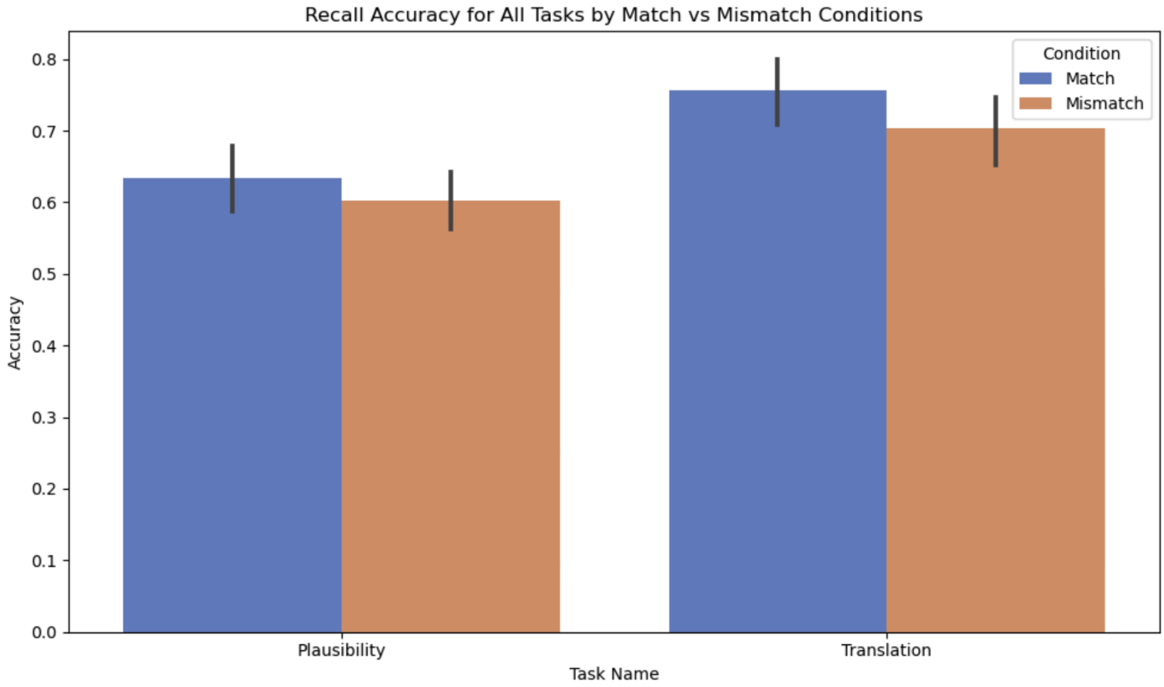
In this study, we aimed to determine whether participants demonstrated higher accuracy or memory strength for items in a matching condition, separately for the Plausibility group and the Translation group. Out of the 90 participants, 44 were assigned to the Plausibility group, and 46 were assigned to the Translation group. Participants whose task accuracy was lower than 50% in the sentence tasks were excluded from the analysis.

Recall Accuracy

We used descriptive statistical analysis to calculate participants' recall accuracy. The overall mean recall accuracy for the Plausibility group was 67.2% (SD = .090), with a mean accuracy of 64.3% (SD = .188) for the match condition and 60.3% (SD = .168) for the mismatch condition. In the Translation group, the overall mean accuracy was 79.2% (SD = 0.90), with a mean accuracy of 75.6% (SD = .170) for the match condition and 70.2% (SD =.177) for the mismatch condition. For a visualization of the results, please see Figure 1.

Figure 1

Recall Accuracy for All Tasks by Match vs Mismatch Conditions



Note. The bar graph shows the mean recall accuracy for the Plausibility and Translation groups under match and mismatch conditions. Error bars represent standard deviations.

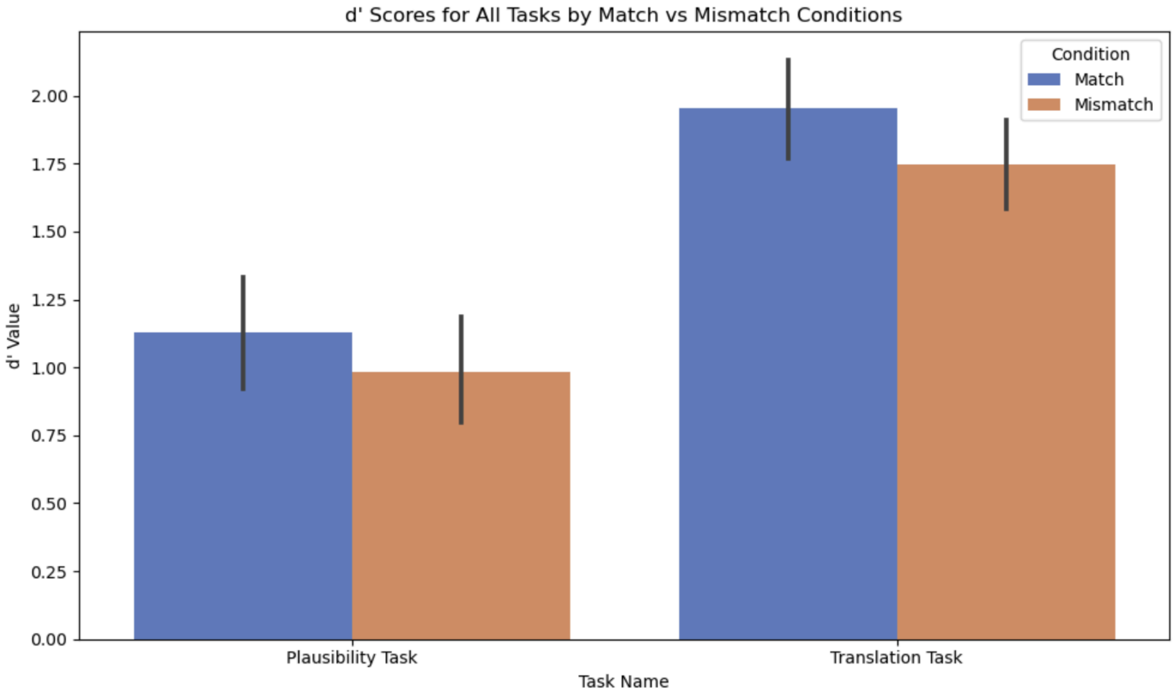
Memory strength (d')

To avoid high accuracy caused by random guessing or answer strategies, we calculated d-prime (d'), which identifies participants' ability to distinguish between critical items (match/mismatch) and fillers. D-prime was calculated using the hit rate and false alarm rate according to the following function: *d' = z(Hit rate) - z(False Alarm rate)*. A larger d' indicates a higher ability to distinguish between critical items and fillers.

The average memory strength (d') for the Plausibility group was 1.05 (SD = .651), with an average of 1.13 (SD = .681) for the match condition and 0.984 (SD = .659) for the mismatch condition. The average memory strength (d') for the Translation group was 1.83 (SD = .651), with an average of 1.96 (SD = .661) for the match condition and 1.75 (SD = .585) for the mismatch condition. For a visualization of the results, please see Figure 2.

Figure 2

d' scores for All Tasks by Match vs Mismatch Conditions



Note. The bar graph shows the mean memory strength (d') for the Plausibility and Translation groups under match and mismatch conditions. Error bars represent standard deviations.

T-test

Between Groups. As shown in Figure 2, there is a significant difference between the two groups. We conducted an independent t-test of the d' values based on groups for both match and mismatch conditions. The results showed that for both match and mismatch conditions, the t-value was t = -5.83 (df = 88, p < .01), indicating that the mean memory strength for the Plausibility group was significantly lower than the Translation Group for both conditions.

We also compared the difference between match and mismatch conditions within each group.

Match/Mismatch in Plausibility Group. The t-test was not significant ($t = 1.03$, $p = .305$), suggesting no significant difference between match and mismatch conditions within the Plausibility Task.

Match/Mismatch in Translation Group.The t-test was also not significant ($t = -1.60$, $p = .114$), indicating no significant difference between match and mismatch conditions within the Translation Task.

Given the observed trend in the expected direction, we hypothesized that the study might be underpowered. Therefore, we conducted a Repeated Measures ANOVA to investigate the main effects and interactions, followed by a post hoc analysis.

Repeated Measure ANOVA

According to the results, there is a main effect of Match/Mismatch condition ($F(1, 88) = 20.854$, $p < .001$), suggesting a significant difference between the Match and Mismatch conditions overall. There is also a main effect on group conditions ($F(1, 88) = 37.000$, $p < .001$), indicating that the overall memory strength of the Plausibility group is significantly different from the overall memory strength of the Translation group. However, the change between match and mismatch conditions across task groups was not significantly different ($F(1, 88) = 0.606$, $p = .439$).

By running the Post Hoc Tests, we considered the comparisons within the whole context, benefiting from increased power and adjustments for multiple comparisons. According to Table 1, for the Plausibility group, the match condition was significantly larger than the mismatch condition ($t(88) = 2.65$, $p = .046$). There was also a significant difference between the match and mismatch conditions in the Translation group ($t(88) = 3.82$, $p = .001$), indicating a main match effect in both the Translation and Plausibility groups. Additionally, the memory strength in the match condition for the Translation group was significantly larger than for the Plausibility group, and the same was true for the mismatch condition ($t(88) = 5.83$, $p < .001$), corresponding to the between-group t-test result.

Table 1. *Post Hoc Comparison of Memory Strength (d') Across Conditions.*

Comparison	Mean Difference	SE	df	t	p
Plausibility task - Match vs Mismatch	0.147	0.0557	88	2.65	0.046
Translation task - Match vs Mismatch	0.208	0.0544	88	3.82	0.001
Match - Translation vs Plausibility	0.825	0.1415	88	5.83	<.0.001
Mismatch - Translation vs Plausibility	0.765	0.1312	88	5.83	<.0.001

Note. The table presents the comparisons between match and mismatch conditions within the Plausibility and Translation tasks, as well as between the Translation and Plausibility groups for both match and mismatch conditions. Significant differences were observed in all comparisons, with p-values less than .05 indicating statistical significance.

General Discussion

Our study aimed to explore the presence of the match effect in both translation and plausibility tasks, with a specific focus on embodied simulation during language processing. While our results do not provide conclusive evidence to confirm the match effect and reject the null hypothesis, we observed a notable trend suggesting active embodied simulation during translation. The potential match effect in both the translation and plausibility groups displayed only a slight difference. This suggests that the degree of embodied simulation during translation is very similar to that during language comprehension tasks. The main effect observed between the groups indicates that translating sentences may involve a deeper cognitive process, leading to better encoding of stimuli into memory compared to the plausibility task.

Several limitations of our study must be acknowledged. First, our study might have been underpowered to detect a significant effect due to the limited number of participants.

The complexity of the tasks resulted in a high dropout rate, further reducing our sample size. Additionally, conducting the experiment online posed challenges in controlling participants' attention and the experimental environment. Some participants may have used strategies to complete the study quickly, resulting in data that lacked value for our analysis.

Language proficiency, interpreting strategies, personal experience, and attention levels also significantly influenced our results. For instance, many participants mistakenly recognized "lamp" as "lamb" and "zodiac" was often interpreted as either "Birthday Zodiac" or "Chinese Zodiac" instead of an inflatable boat. These misinterpretations indicate that participants' personal experiences and familiarity with certain words affected their processing and recall, potentially skewing the outcomes of the recall task. These factors could be further explored with the current data or considered in future studies. Moreover, requiring participants to type out translations may have presented stimuli to them twice, possibly contributing to the main group effect observed.

Despite these limitations, our study provides valuable insights and several lessons for future research. Moving forward, modifying the task procedures could address some of the identified issues. For example, instead of asking participants to type out translations, researchers could ask them to verbalize the translations without repeating the original sentence, thereby avoiding potential confounding variables. Ensuring participant attention during the experiment, conducting studies in person, and removing unsuitable stimuli (such as those with which participants are unfamiliar or that have multiple meanings) could improve data quality. Additionally, recruiting a larger and more diverse sample would enhance the study's power and generalizability.

In conclusion, while our study faced several challenges and limitations, it nonetheless highlights the intricate cognitive processes involved in translation and the potential for embodied simulation during language tasks. Future research can build on these findings, refining methodologies and addressing limitations to further elucidate the cognitive mechanisms underlying translation and language comprehension.

Conclusion

This research facilitates our understanding of the cognitive mechanisms involved in translation, and also helps establish the conditions under which simulation does or does not occur. Embodied simulation was found to be involved in translation, even though we do not have sufficient evidence to conclude this, the trend we found still offers valuable insights into how humans process and comprehend language. Knowledge of the mechanisms underlying human translation can provide insights to enhance the appropriateness of machine translation. By understanding how people produce translations, we can learn more about the brain's linguistic mechanisms, the cognitive-language connection, which can facilitate the learning of new languages, the cognitive development of children, and even the development of human-computer interaction. These insights can be supportive for future research in the field of linguistics and cognitive science.

Reference

Bergen, B. (2012). Louder than Words: The New Science of How the Mind Makes Meaning. Basic Books.

Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. Social Neuroscience, 3(3-4), 317-333.

Pecher, D. et al. (2009). Short article: Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology, 62*(6), 1108-1114.

Tian, Y., Zarcone, A., & Moens, M. (2020). The role of perceptual simulation in the mental representation of meaning: New insights from a comprehension study. Cognition, 196, 104124.

Wheeler, K., & Stojnovic, D. (2006). Non-native language processing engages mental imagery. *Proceeding of an annual meeting of the cognitive science society, 28*, 882-887.

Acknowledgements

Thanks to Dr. Sean Trott for his advice throughout the project and the paper.

Thanks to Dr. Benjamin Bergen, Ph.D. Student Catherine Arnett for their advice on the proposed study.

Thanks to Dr. Bradley Voytek for his advice on the project and presentation throughout this academic year.