5-1-2016

# Language Use Statistics and Perceptual Simulation in Language Processing

Alena Hejl
*Macalester College*, alena.hejl@gmail.com

Language use statistics and perceptual simulation in language processing

Alena E. Hejl

Honors Project in Linguistics

April 26, 2016

Advisor: Brooke Lea

Psychology Department

Macalester College

Author Note

I am deeply grateful to my advisor, Professor Brooke Lea, whose wisdom, compassion, and confidence in me made this project possible. I sincerely thank Professor Christina Esposito for her support and constructive editing, and Professors John Haiman and Stephanie Farmer for challenging me with thought-provoking questions. Special thanks to Katherine Muschler for averting computational crises, and to our fellow iLab members who assisted me with data collection. Thank you all for graciously lending your time and talents to this project!

Abstract

Humans' ability to comprehend language seems to rely on both mental reconstructions of what we have experienced in the world and statistically-based expectations of how language is used. This study adapted a comparison of perceptual and statistical explanations of word comprehension in the auditory modality. Participants completed a series of trials in which they heard cue words, some of which were spatially oriented (e.g., sky, ground), and then completed a letter identification task. In this task, the letter appeared on the computer screen in either a congruent location or an incongruent location. The position of the letter at the top or bottom of the screen was defined as congruent if it matched: 1) the spatial meaning of the cue word (e.g., up for "sky"); or 2) the direction that occurs most frequently with the cue word in English corpora (e.g., down for "slow"). Response times to the letter task were expected to replicate prior findings that participants identify letters in congruent locations faster than incongruent ones. Eye-tracking was used as an additional measure of embodied perceptual processing. Where participants looked on the computer screen was predicted to correspond with the imagined spatial location of the cue word. Differences in eye movement patterns did not support the perceptual processing hypothesis. The correlation between a word's statistical co-occurrence with spatial words and response times was significant.

*Keywords*: language use statistics, perceptual simulation, symbolic processing

# Listen Up: Language use statistics and perceptual simulation in language processing

English speakers often use spatial terms to communicate abstract concepts that have no literal spatial dimensions, as in the phrases “cool down” and “cheer up,” or describing authority figures as being “above” their subordinates. But how do we understand these phrases? According to Louwerse’s (2011) Symbol Interdependency Hypothesis, word comprehension depends on both perceptual and symbolic processing.

Theories of perceptual or embodied cognition suggest that information from bottom-up perceptual experiences can later be partially re-activated in top-down processing when the perceptual stimuli are no longer present (Barsalou, 1999). Louwerse and Connell (2010) found that people were slower to process sentences about sensory information like taste and smell when they had to switch from one sense to another, requiring re-activation of different perceptual systems. These results indicate that conceptual processing is linked to the body’s sensory processing (Louwerse & Connell, 2010). Faster response times in this study were also connected with the statistical frequency of adjectives occurring with words for a certain sensory modality, though this connection was less precise (Louwerse & Connell, 2010).

Even abstract words seem to have the power to orient people’s spatial attention. Zwaan and Yaxley (2003) found that people were faster to recognize that a pair of words was semantically related when their orientation on the screen matched their referents’ spatial configuration. For example, people recognized that “attic” and “basement” were related in meaning faster when “attic” was above “basement” than in the opposite position (Zwaan & Yaxley, 2003). Reading a word also results in faster responses to subsequent material presented in the portion of the visual field related to the word's literal or metaphorical meaning (Goodhew, McGaw, & Kidd, 2014). Goodhew, et al. (2014) found that after reading an up-related word, participants were faster to respond to a probe at the top of the screen than at the bottom. This is consistent with the view that perceptual simulation—re-creating in the mind what we have encountered in physical reality—is involved in processing words.

Even abstract words seem to have the ability to orient people's spatial attention (Hutchinson & Louwerse, 2013). Conceptual cueing effects on response times have also been found for abstract words with no literal spatial dimensions (Zwaan & Yaxley, 2003; Hutchinson & Louwerse, 2013). Hutchinson and Louwerse (2013) found that people were faster to respond to abstract word pairs whose configuration on the screen matched common metaphors, like "good" being associated with up and "bad" being associated with down. In addition to word pairs with a good-bad valence relationship, pairs relating to temperature, authority, and gender were also recognized faster when their orientation on the screen matched the metaphorical relationship (Hutchinson & Louwerse, 2013). These results indicate that in addition to perceptual spatial configurations, the way we use language plays a role in explaining how we process words.

Symbolic and embodied processing generally seem to occur in sequence, with linguistic frequency processes active earlier in the course of responding and with relatively large cueing effects, while perceptual simulation is active slightly later and with smaller effects (Louwerse et al., 2014). Louwerse and Hutchinson (2012) used EEG to compare participants' neural activity while they determined the semantic relatedness of word pairs like "attic" and "basement." They found greater activation in linguistic cortical areas earlier in the trials, while perceptual cortical areas showed greater activation later in the trials (Louwerse & Hutchinson, 2012). This provided neurological evidence that both symbolic and embodied processing is involved in comprehension, and that linguistic processes precede perceptual ones (Louwerse & Hutchinson, 2012).

Goodhew, McGaw, and Kidd (2014) found that statistical language usage patterns, like the co-occurrence of a word with a spatial term like "up," predict the magnitude of the spatial attention cueing phenomenon; for example, the word "dream" occurs much more frequently with the word "up" than "down" in English corpora. It also seems to cue attention upward, resulting in faster responses to probes presented at the top of the screen than at the bottom of the screen (Goodhew, McGaw, & Kidd, 2014). Their participants looked at a fixation cross in the middle of the screen, then read a word in the middle of the screen, and next were shown a letter either above or below where the word had appeared. The participants' task was to press the correct key on the keyboard to identify the letter that had appeared on the screen.

Response times revealed that participants were faster to identify the letter when its position on the screen matched the direction it occurred with most frequently (Goodhew, McGaw, & Kidd, 2014). For example, responses after the word "ceiling" were faster when the letter appeared above the middle of the screen than when it appeared below it. They found that the statistical collocation of a stimulus word with directional words in corpora predicted the magnitude of the difference in response speed for the matched and mismatched letter position (Goodhew, McGaw, & Kidd, 2014). Words occurring frequently with “up” in corpora produced faster responses to letters above the cue word than words occurring less frequently with “up.” This is consistent with the view of word comprehension relying on symbolic language processes.

The goal of the present study was to clarify the roles of statistical and perceptual simulation in the rapid processing of speech. This study adapted the methods of Goodhew, McGaw, and Kidd (2014) for speech processing and used eye-tracking as a measurement of perceptual simulation. In each trial, participants focused on a central fixation cross for 1000 milliseconds. The fixation cross then disappeared and a stimulus word was heard through the headphones followed by 800 milliseconds of silence. Next, a letter (L or S) appeared either above or below the fixation point, and the participant’s task was to identify the letter by pressing the corresponding key on the keyboard. For example, the participant would see the fixation point, then may hear the word ‘castle,’ and next see the letter ‘L’ at the bottom of the screen. The participant would then press the ‘L’ key on the keyboard as quickly as possible for a correct response.

I predicted conceptual cueing effects consistent with the results of Goodhew, McGaw, and Kidd (2014), which found that response times on the letter identification task were faster when the target letter was presented in the position indicated by statistical language use frequencies (top of the screen for words occurring more frequently with “up” and “above” or bottom of the screen for words occurring more frequently with “down” and “below”). These patterns are consistent with the involvement of symbolic language processes in word comprehension.

This study measured eye movements in an attempt to detect embodied perceptual processing of the stimulus words. Eye movements play a functional role in perceptual processing; the way we move our eyes while hearing a description mimic the eye movements made while actually looking at the same scene (Laeng & Teodorescu, 2002). If perceptual simulation were involved in rapid auditory processing, participants' gaze would be expected to move in the direction cued by the stimulus word before the target letter is presented. If only language frequencies are involved in processing, directional eye movement patterns would not be expected. The absence of eye movement patterns may indicate that perceptual simulation does not explain rapid word comprehension as well as language use patterns, or that perceptual simulation may not occur quickly enough to account for auditory language processing.

## Method

### Participants

The participants were 29 undergraduate students at Macalester College with normal or corrected-to-normal vision and normal unaided hearing. Participants had no prior knowledge of the study. Those enrolled in psychology courses received class credit for their participation.

### Materials

Eye-tracking was performed using an EyeLink1000 eye-tracking device manufactured by SR Research (Ottawa, Ontario, Canada), a PC computer running Windows 7, and equipped with a standard keyboard, mouse, and headphones connected to speakers. The computer had an 18-inch, 1024x768 resolution CRT monitor set to a refresh rate of 75 Hz. The tracker headrest was positioned 55 cm from the screen. The experiment was programmed using ExperimentBuilder software produced by SR Research, and eye movement results were analyzed using DataViewer software from SR Research. Stimuli consisted of an on-screen fixation cross, audio recordings of the words, and the capital letters S and L for the letter-identification task. These letters were selected for their distinct shapes and because their placement on the keyboard made enabled participants to comfortably rest both hands on the keyboard while using the eye-tracker.

All visual stimuli, such as the fixation cross and probe letters, were in black text on a white background. Stimulus words were selected from previous experiments and rating studies (Goodhew, McGaw, & Kidd, 2014). To operationalize perceptual processing, words were categorized as having an upward, downward, or neutral (neither strongly up or strongly down) affordance based on a conceptual or experiential association with that direction. For instance, “happy” has an upward affordance due to the metaphor “happy is up” as evidenced by idioms like “in high spirits,” while “ground” has a downward affordance due to its position below us in the perceptual world. Abstract, concrete, and literal words were used in each affordance category to adhere to the methods of Goodhew, et al. (2014). The stimulus words are listed in Table 1.

*Table 1.* Stimulus Word List

| **Word** | **Affordance** |
|---|---|
| sky | upward |
| up | upward |
| above | upward |
| high | upward |
| north | upward |
| happy | upward |
| top | upward |
| dream | upward |
| heaven | upward |
| ceiling | upward |
| castle | upward |
| head | upward |
| bottom | downward |
| down | downward |
| low | downward |
| below | downward |
| under | downward |
| sad | downward |
| drain | downward |
| sinking | downward |
| ground | downward |
| slow | downward |
| lid* | downward |
| street | downward |
| middle | neutral |
| half | neutral |
| medium | neutral |
| midway | neutral |
| core | neutral |
| center | neutral |
| belt | neutral |
| waist | neutral |
| handle | neutral |
| knob | neutral |
| equator | neutral |
| frame | neutral |

*Note: this item's affordance was misclassified; it was not used in the final analysis.

Aural stimuli consisted of recordings of English words spoken aloud in isolation in an affect-neutral tone. The speaker (the author) was a female monolingual English speaker of a general Midwestern American dialect. The stimuli were recorded in a sound-treated room with high-quality recording equipment. The order in which the stimulus words were read during recording is provided in Table 2. The speaker's pitch in items at the beginning and end of the word list did not differ significantly, and the average pitch ranged from 168.13 Hz to 206.19 Hz.

*Table 2*. Order of Audio Recordings

| **Word** |
|---|
| 1. Up |
| 2. Bottom |
| 3. Middle |
| 4. Above |
| 5. Down |
| 6. Half |
| 7. High |
| 8. Low |
| 9. Medium |
| 10. North |
| 11. Below |
| 12. Belt |
| 13. Sky |
| 14. Head |
| 15. Waist |
| 16. Ceiling |
| 17. Drain |
| 18. Handle |
| 19. Happy |
| 20. Under |
| 21. Midway |
| 22. Top |
| 23. Sad |
| 24. Core |
| 25. Dream |
| 26. Sinking |
| 27. Center |
| 28. Heaven |
| 29. Slow |
| 30. Knob |
| 31. Castle |
| 32. Lid |
| 33. Equator |
| 34. Ground |
| 35. Street |
| 36. Frame |

**Procedure**

The experiment was performed in a research lab equipped with an eye tracker and computers as previously described. After giving informed consent, participants completed four practice trials before beginning the experiment. Participants were instructed to focus their gaze on the fixation cross in the middle of the screen for 1000 milliseconds until it disappeared. At that time, they heard the stimulus word through the headphones. Sixteen hundred milliseconds after the onset of the stimulus word, the letter “S” or “L” appeared 8° (approximately 4.25 inches) above or below the center of the screen. Each letter appeared with equal frequency overall and in both positions on the screen. Trials were pseudo-randomized and counterbalanced so that no word could be presented twice in a row and all words occurred the same number of times. Each participant heard each word in four separate trials: once followed by the appearance of “L” above the fixation cross, once with “L” below, once with “S” above, and once with “S” below.

The participants’ task was to identify the letter they saw by pressing the corresponding key on the keyboard as quickly and accurately as possible. Once the participants responded by pressing a key, the target letter vanished and the screen remained blank for 1000 milliseconds before the next trial began. The experiment lasted approximately 20 minutes.

The dependent variables measured were the reaction time (RT) and the direction of saccades launched after the disappearance of the fixation cross but before the appearance of the probe letter. Only RTs for correct trials were analyzed.

Statistical linguistic frequencies for analysis were represented by a collocation value for each word that represented how frequently it co-occurred with directional words. Previous studies have calculated collocation scores by adding the bigram frequencies of the target word with "up" and "above," and subtracting from that the sum of the bigram frequencies with "down" and "below," then log transforming this difference (Goodhew, et al., 2014). Collocation was calculated for each word by averaging the rate of its occurrence in a bigram with the word "up" or "above" in the Google N-Gram corpus over a ten-year period and subtracting the rate of occurrence with the word "down" or "below." Due to the time-consuming nature of calculating the collocation values, not all words were analyzed.

**Response Time Results**

Aggregated response time data for words with upward and downward affordances were analyzed. Data from three participants were excluded from analysis because some results were not recorded properly. RTs exceeding three seconds were considered outliers and were excluded. Per the methods of Goodhew, et al. (2014), I calculated the difference in RTs between trials with above-fixation probes and below-fixation probes for each word for each participant (i.e., mean RT[probe-up trials] – mean RT[probe-down trials]). Mean differences for each target word appear in Table 3. A more negative RT difference for an upward-cueing word would indicate that participants took longer to respond to probes in the inconsistent downward position, demonstrating conceptual cueing. For a downward-cueing word, conceptual cueing would be manifested in a more positive score would indicate that participants took longer to respond to probes in the inconsistent upward position.

*Table 3*. Mean Response Time Differences

| Word | RT Difference (ms) | Affordance | Collocation |
|---|---|---|---|
| ABOVE | 14.65 | Upward | 3.52 |
| BELOW | 37.76 | Downward | -2.67 |
| BOTTOM | -19.48 | Downward | 3.63 |
| CASTLE | -39.63 | Upward | 1.58 |
| CEILING | 23.15 | Upward | 5.82 |
| DOWN | 13.31 | Downward | 1.93 |
| DRAIN | 15.24 | Downward | -1.98 |
| DREAM | -26.81 | Upward | 5.61 |
| GROUND | 19.94 | Downward | 2.23 |
| HAPPY | -71.11 | Upward | 3.78 |
| HEAD | -37.48 | Upward | 1.81 |
| HEAVEN | 22.33 | Upward | -.50 |
| HIGH | 9.87 | Upward | 8.57 |
| NORTH | 14.20 | Upward | 1.85 |
| SAD | -2.26 | Downward | 2.70 |
| SINKING | 65.13 | Downward | -5.21 |
| SKY | -32.59 | Upward | 4.81 |
| SLOW | 28.35 | Downward | -4.41 |
| TOP | -10.07 | Upward | -.44 |
| UNDER | -46.13 | Downward | 2.94 |
| UP | 26.89 | Upward | 1.87 |

Variance in RT difference scores with respect to collocation was analyzed. A scatterplot of the relationship between response time difference and collocation is provided in Figure 4. Differences between RTs for up-probe and down-probe trials seem to decrease as collocation scores increase. Difference score and collocation value had a strongly significant negative correlation, $r(20)=-.49$, $p<.05$.

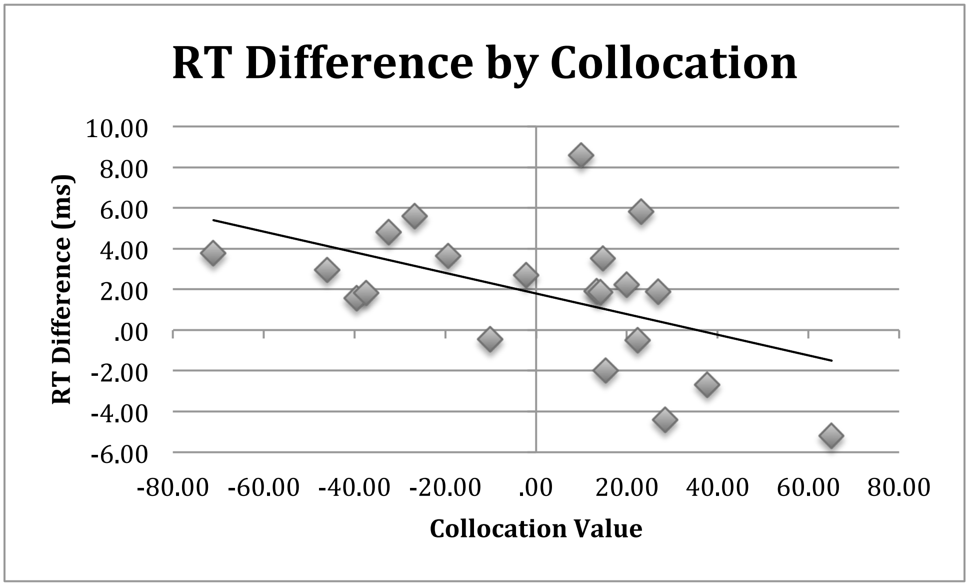


*Figure 4.* Difference (ms) in response times for trials with down-position probes subtracted from response times for up-position probes. Collocation value reported as the log mean of the bigram occurrence frequency with the words “down” and “below” subtracted from the log mean of the bigram occurrence frequency with the words “up” and “above.” Shown with line of best fit.

Variance in response time difference with respect to affordance was also analyzed. A bar graph of the response time differences is provided in Figure 5. Response time differences for words with a downward affordance were higher than those for words with an upward affordance. A t-test revealed that the response time differences for words with a downward affordance ($M$=6.05, $SD$=36.18) were not significantly different from that of words with an upward affordance ($M$=-4.10, $SD$=31.59), $t(19)$=.50, $SEM$=14.82, $p$=.78.

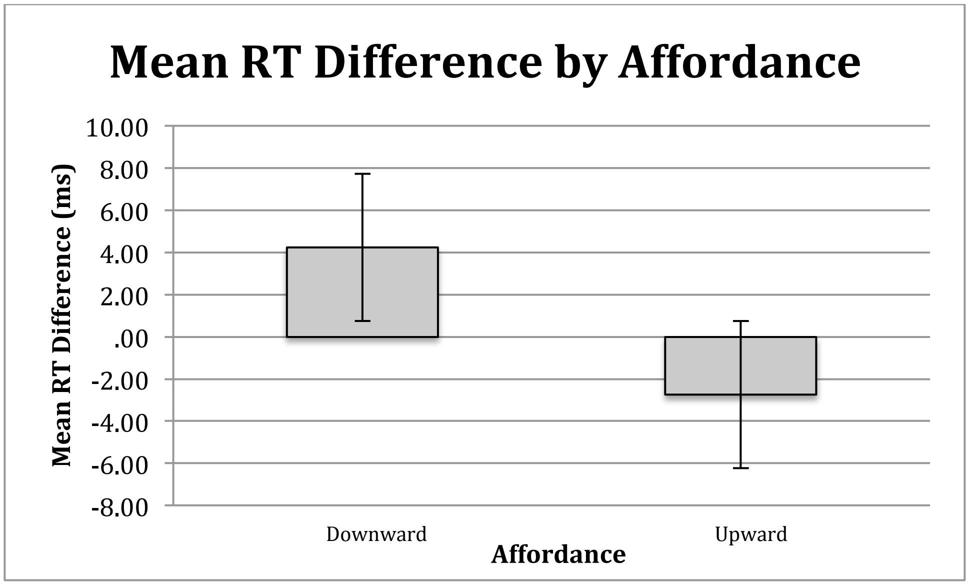


*Figure 5.* Mean difference (ms) for response times to trials with down-position probes subtracted from trials with up-position probes. The difference between the difference values for upward and downward affordance categories was not significant. Standard error bars are shown.

**Eye Movement Results**

DataViewer was used to analyze the first eye movement during the time between the disappearance of the fixation cross and the appearance of the probe letter. During some trials, participants launched no saccades during the analyzed time period; such trials were excluded from analysis. Saccades that were made during this time period of each trial were coded as upward, downward, or within the middle of the screen around the fixation cross. An ANOVA revealed an unexpected significant relationship between affordance and saccade direction. Participants made saccades to the middle of the screen more often than upward or downward saccades after hearing words with a downward affordance, $F(2,43)=3.437$, $p=.035$. Middle saccades also occurred more often than upward or downward saccades after hearing words with an upward affordance, $F(2,43)=6.535$, $p=.002$. The number of saccades upward, downward, and to the middle of the screen did not differ significantly after hearing words with a middle (neutral) affordance.

More saccades were made to the middle of the screen than upward or downward saccades for all affordance levels. The proportion of saccades in each direction (downward, to the middle, or upward) for each affordance level is presented in Figure 6.

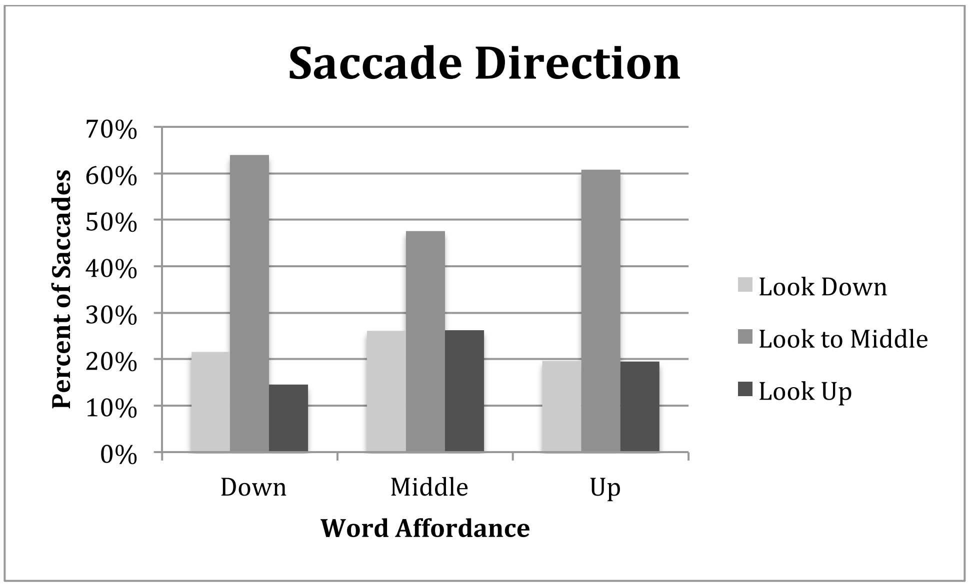


*Figure 6.* Proportion of saccades in each direction by word affordance

**Discussion**

The goal of the present study was to replicate Goodhew, et al.'s (2014) experiment substituting auditory stimuli for written stimulus words. They found that after reading a word, participants were faster to identify a letter when it was presented in either a top or bottom position on the screen and that this reaction time difference could be predicted to a similar degree by the word's statistical frequency of occurrence with a direction word or the word's directional affordance (Goodhew, et al., 2014).

The significant negative correlation between collocation and response time difference between up-probe trials and down-probe trials supports the hypothesis that collocation predicts the magnitude of conceptual cueing. Words that appear more frequently in corpora with "up" and "above," and therefore had higher collocation values, had lower response time differences. This indicated that response times of trials with the probe in the up position were faster than those of trials with the probe in the down position for words with higher collocations. This supports the assertion by Goodhew, et al. (2014) that collocation as a representation of language use statistics does predict conceptual cueing.

Eye movement data did not provide evidence consistent with perceptual processing. I predicted that participants would launch saccades upward more often after words with upward affordances than after those with downward affordances, and launch downward saccades more often after words with downward affordances than after those with upward affordances. I expected saccades to the middle of the screen most frequently after words with neutral affordances. The number of saccades to the middle after words with upward and downward affordances was expected to be lower than the number of saccades matching the affordance but higher than the number of saccades in the opposite direction of the affordance. Instead, all affordances produced more saccades to the middle than upward or downward.

The predominance of saccades to the middle of the screen may have been the consequence of participants' conscious or unconscious strategy for moving their eyes most efficiently to gather the information necessary to identify the letter. Participants could have kept their gaze near where the fixation cross had been to avoid wasting valuable time looking up or down in anticipation of the probe letter's appearance.

Similarly, the many trials in which participants made no saccades during the time between the disappearance of the fixation cross and the appearance of the probe letter may be attributable to the same gaze strategy. To address this possibility, the fixation cross could be centered horizontally on the right or left edge of the screen. If participants were indeed using an efficiency strategy for looking, then saccades away from the fixation cross to the center of the screen would be most frequent.

My results were consistent with the symbolic processing account of language comprehension. However, they did not corroborate the finding that language use frequencies and perceptual simulation predict conceptual cuing to a similar extent. Instead, the significant relationship between reaction time difference and correlation but not affordance is consistent with the idea that language use statistics explain speech processing better than perceptual simulation does.

If linguistic processing and embodied cognition are mutually exclusive phenomena, these findings may cast doubt on the importance of the latter in speech comprehension. On the other hand, if perceptual processing is a cause or consequence of systematic language use as Goodhew (2014) suggests, then affordance may not have been a sufficient way to operationalize embodied cognition.

The results of Goodhew et al. (2014) indicated that affordance and collocation explain similar amounts of variance in reaction times for trials with the probe in the matching and mismatching position. However, the present study did not find affordance to significantly predict reaction time differences. The present study therefore cannot confirm that perceptual simulation plays a significant role in language processing. It also follows that this study cannot confirm that perceptual simulation and language use statistics explain a similar amount of variance in the speed of language processing.

At this time, the results of Goodhew et al. (2014) have not yet been replicated. Confirming their results in the original modality of reading would be a valuable contribution to the psycholinguistic community by affirming that conceptual cuing is indeed a measurable effect of language processing and that both perceptual simulation and statistical language use frequencies influence how we understand words. Then, replicating these results in the aural modality—as the present study attempted to do—would confirm that these mechanisms of language processing are the same across modalities.

The next path of inquiry after replication may investigate the relationship between statistical frequency processing and perceptual processing. It is unclear whether the language we use reflects the way our minds map spatial information about the world or our perceptual mapping reflects the way we use language. Further studies may also clarify the time course of language processing to determine when and why we shift from relying on patterns of language use to depending more heavily on perceptual simulation.

**References**


Barsalou LW. (1999). Perceptual symbol systems. *The Behavioral And Brain Sciences, 22*(4), 577-609; Discussion 610-60.

Goodhew, S., McGaw, B., and Kidd, E. (2014). Why *is* the sunny side always up? Explaining the spatial mapping of concepts by language use. *Psychonomic Bulletin & Review, 21*(5), 1287-1293. doi:10.3758/s13423-014-0593-6

Hutchinson S., and Louwerse M. (2013). Language statistics and individual differences in processing primary metaphors. *Cognitive Linguistics, 24*(4), 667-687. doi:10.1515/cog-2013-0023

Laeng, B., & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science, 26*(2), 207-231. doi:10.1207/s15516709cog2602_3

Louwerse, M., & Connell, L. (2011). A Taste of Words: Linguistic Context and Perceptual Simulation Predict the Modality of Words. *Cognitive Science, 35*(2), 381-398. doi:10.1111/j.1551-6709.2010.01157.x

Louwerse, M. and Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual Simulation in conceptual processing. *Frontiers in Psychology, 3*, 385. doi:10.3389/fpsyg.2012.00385

Louwerse, M., Hutchinson, S., Tillman, R., & Recchia, G. (2014). Effect size matters: the role of language statistics and perceptual simulation in conceptual processing. *Language, Cognition and Neuroscience, 30*(4), 430-447. doi:10.1080/23273798.2014.981552

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review, 10*(4), 954-958. doi:10.3758/BF03196557